HarmonySeg: Tubular Structure Segmentation with Deep-Shallow Feature Fusion and Growth-Suppression Balanced Loss

Supplementary Material

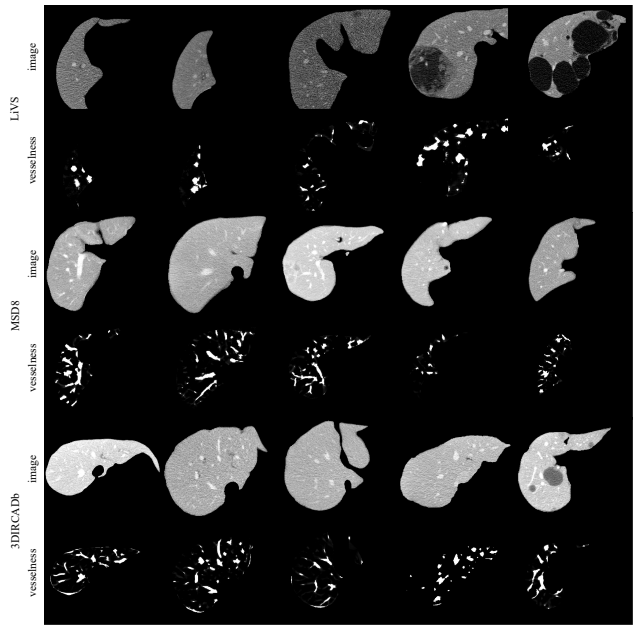


Figure 1. Visualization of images and their corresponding vesselness filtering results. In the CT images, regions with high intensities represent hepatic vessels, while dark regions indicate tumors. In the paired vesselness filtering results, high-intensity patches correspond to vessel candidates, with noise visibly present along the liver border.

Table 1. Enhanced ratio of labeled slices on the HVS task brought by refined labels (%).

Hepatic vessel segmentation (HVS)			
Label	LiVS	MSD8	3DIRCADb (Testing)
Original	$20.26_{\pm 12.77}$	$71.96_{\pm 12.29}$	$78.87_{\pm 7.27}$
Refined	$70.26_{\pm 11.75}$	$72.54_{\pm 11.90}$	

In the supplementary material, we provide detailed explanations of the vesselness filter (Appendix A), the flexible convolution block (Appendix B), the segmentation fusion in D2SD (Appendix C), along with additional experimental and visualization results, including refined hepatic vessel labels (Appendix D), ablation studies (Appendix E), our curated HVS-External dataset (Appendix F), the robustness of loss functions (Appendix G), and the trade-off between precision and recall of loss functions (Appendix H).

Table 2. Quantitative improvement on the HVS task brought by refined labels. The segmentation model is nnU-Net.

Dataset	Label	Dice(%,↑)	HD(↓)
HVS	Original	$56.69_{\pm 8.42}$	$10.83_{\pm 4.28}$
	Refined	$\mathbf{60.15}_{\pm 9.49}$	$\mathbf{10.00}_{\pm 4.22}$
HVS-External	Original	$63.09_{\pm 10.90}$	$4.02_{\pm 1.43}$
	Refined	$\mathbf{63.78}_{\pm 9.27}$	$\mathbf{3.69}_{\pm 1.17}$

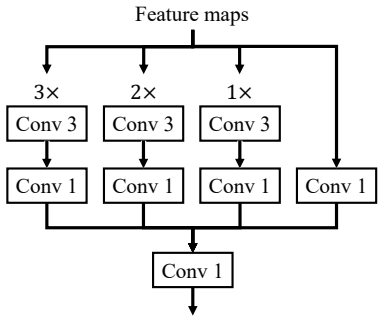


Figure 2. Details of flexible convolution block with diverse receptive fields.

A. Vesselness Filter

In this section, we elaborate on the details of vesselness filters and present several vesselness maps for visualization. Image derivatives, including first-order derivatives for border detection and second-order derivatives for shape extraction, are commonly used to highlight vascular structures in images [1]. Hessian matrix analysis is a representative method based on second-order derivates that can distinguish rounded, tubular, and planar structures [3]. Vesselness filters also employ eigen-decomposition of the Hessian matrix to measure tubularity and enhance vessel regions [8]. Let H be the hessian matrix of a voxel in CT volume, and $\mathbf{e_1}$, $\mathbf{e_2}$ and $\mathbf{e_3}$ be the three eigenvectors of H with corresponding eigenvalues of λ_1, λ_2 and λ_3 ($|\lambda_1| \leq |\lambda_2| \leq |\lambda_3|$). The tubularity is defined as [9]:

$$|\lambda_1| \approx 0, \lambda_2 \approx \lambda_3 \ll 0. \tag{1}$$

Based on this, the Jerman [6] vesselness filter used in our framework further regularizes λ_3 to reduce the sensitivity

Table 3. Quantitative comparison on the HVS-External test set stratified by diseases of the subjects.

Model	HVS-External							
	Fatty liver		Cirrhosis		Tumor		Healthy	
	Dice(%,↑)	HD(↓)	Dice(%,↑)	HD(↓)	Dice(%,↑)	HD(↓)	Dice(%,↑)	HD(↓)
nnU-Net [5]	$53.97_{\pm 1.97}$	$3.90_{\pm 0.09}$	$69.83_{\pm 9.15}$	$2.78_{\pm 0.64}$	$62.88_{\pm 10.14}$	$3.91_{\pm 1.29}$	$62.09_{\pm 2.75}$	$3.94_{\pm 1.08}$
Shit et al. [11]	$50.80_{\pm 3.10}$	$4.16_{\pm 0.11}$	$63.48_{\pm 2.72}$	$3.51_{\pm 0.18}$	$57.18_{\pm 7.33}$	$4.40_{\pm 1.31}$	$49.61_{\pm 8.00}$	$4.96_{\pm 0.72}$
Kirchhoff et al. [7]	$58.88_{\pm 1.69}$	$4.08_{\pm 1.23}$	$\mathbf{81.05}_{\pm 2.22}$	$3.21_{\pm 0.92}$	$70.28_{\pm 9.60}$	$\underline{3.35}_{\pm 1.16}$	$68.73_{\pm 4.72}$	$\underline{3.56}_{\pm 1.16}$
TransU-Net3D [2]	$51.68_{\pm 7.78}$	$6.03_{\pm 2.99}$	$38.18_{\pm 31.76}$	$4.43_{\pm 1.73}$	$39.93_{\pm 21.86}$	$9.76_{\pm 9.17}$	$56.60_{\pm 12.92}$	$5.23_{\pm 1.67}$
DSC-Net [10]	$\underline{71.44}_{\pm 4.51}$	$\mathbf{3.10}_{\pm 0.39}$	$80.50_{\pm 3.68}$	$\underline{2.73}_{\pm 0.66}$	$\underline{74.86}_{\pm 7.04}$	$\mathbf{3.18}_{\pm 1.34}$	$\underline{71.26}_{\pm 7.90}$	$3.69_{\pm 1.10}$
HarmonySeg	$\mathbf{73.15}_{\pm 1.83}$	$\underline{3.74}_{\pm 0.11}$	$\underline{80.59}_{\pm 4.61}$	$\mathbf{2.20}_{\pm 0.55}$	$\mathbf{75.67}_{\pm 7.36}$	$3.61_{\pm 1.31}$	$\mathbf{74.75}_{\pm 6.05}$	$\mathbf{3.37}_{\pm 1.18}$

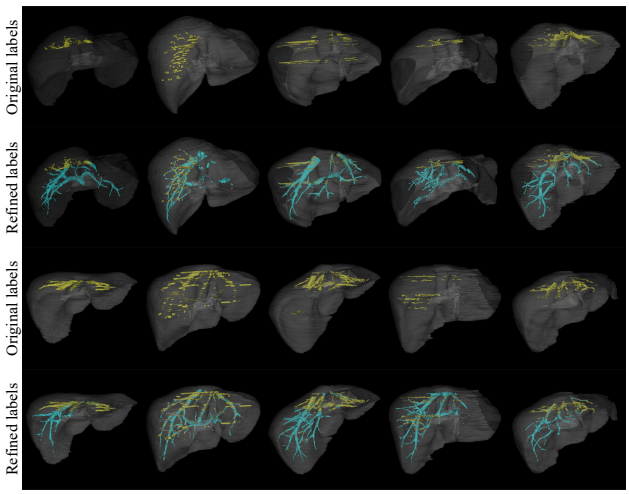


Figure 3. Visualization of original and refined hepatic vessel labels in LiVS. The liver is rendered in gray, while the original and refined vessel labels are denoted in yellow and cyan, respectively.

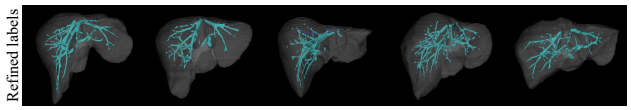


Figure 4. Visualization of predicted hepatic vessel labels in the test set of MSD8. The gray color shows the liver and the cyan color denotes the labels. Note that the original label of MSD8's test set is unavailable.

for those low-contrast regions:

$$F = \begin{cases} 0, & \lambda_2 \leq 0 \text{ or } \lambda_p \leq 0, \\ 1, & \lambda_2 \geq \frac{\lambda_p}{2} > 0, \\ \lambda_2^2(\lambda_p - \lambda_2)(\frac{3}{\lambda_p + \lambda_2})^3, & \text{otherwise}, \end{cases} \tag{2}$$

in which:

$$\lambda_p = \begin{cases} \lambda_3, & \lambda_3 > \tau \max_x \lambda_3(x), \\ \lambda_3 > \tau \max_x \lambda_3(x), & 0 < \lambda_3 \leq \tau \max_x \lambda_3(x), \\ 0, & \text{otherwise}, \end{cases} \tag{3}$$

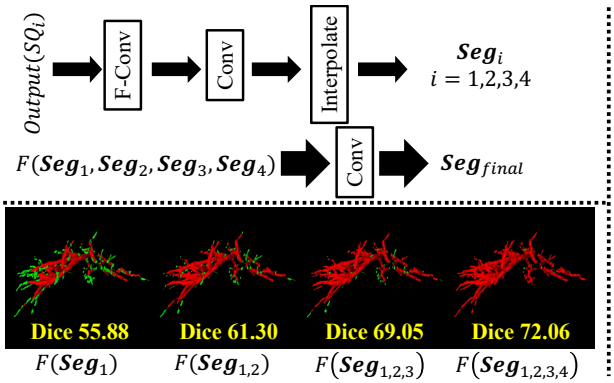
where $\tau \in [0, 1]$. Benefiting from this regularization, the Jerman vesselness filter becomes robust even when facing non-homogeneous vessel intensity. To reveal the effectiveness of the vesselness filter, we give examples of paired vesselness filtering results in Figure 1. As shown in the figure, the vesselness filter can highlight liver vessel candidates of different sizes, even for cases in which tumors exist.

B. Flexible Convolution Block

Diversifying the receptive fields of convolutions is an effective way to adapt models to targets of different sizes [15]. In our study, the sizes of liver vessels are also various, so diverse receptive fields are beneficial in enhancing the model capability. The flexible convolution block we designed is shown in Figure 2. To avoid the gridding effect of dilated convolution for extracting local details of vessels, our flexible convolution block provides different receptive fields by stacking the convolutions in parallel rather than using dilated convolution. After the feature maps are fed into this block, they are further encoded by parallel stacked convolutions with different receptive fields [12]. Then a $1 \times 1 \times 1$ convolution integrates all features and compresses the channel for output. Flexible convolution blocks are used at the encoder and the shallow query module (F-Conv in (c) of Figure 2 in the manuscript).

C. Segmentation Fusion in D2SD

Vessels of varying sizes exhibit distinct feature representations at different scales. Larger vessels can be effectively reflected in multi-scale feature maps. Yet, for smaller vessels, the information loss caused by successive convolutions and pooling tends to impair their feature representation, which is also one of the motivations why skip connections have been introduced. To mitigate this, the D2SD strategy uses low-cost pre-decoders at multiple scales to capture scale-specific information and aggregate multi-scale outputs for final segmentation, as shown in Figure 5. It is important to clarify that the D2SD is distinct from the deep supervision, which does not compute loss for each decoded result. Concretely, our model's encoder-decoder architecture employs F-Conv blocks for feature extraction, where feature maps are flattened directly as tokens (rearranging 'b c z h w' to 'b c (z h w)') and then fed into the DMQ and SQ modules. During decoding, the DMQ module outputs undergo



Dice(%) for D2SD ablation study on the HVS task. $F(*)$ means the D2SD fuses * to obtain the final segmentation.

$F(Seg_1)$	$49.58_{\pm 6.32}$
$F(Seg_{1,2})$	$51.22_{\pm 8.12}$
$F(Seg_{1,2,3})$	$55.94_{\pm 8.31}$
$F(Seg_{1,2,3,4})$	$59.19_{\pm 9.72}$

Figure 5. Qualitative and quantitative analysis of segmentation fusion in D2SD: red/green indicating the segmentation and the corresponding labels.

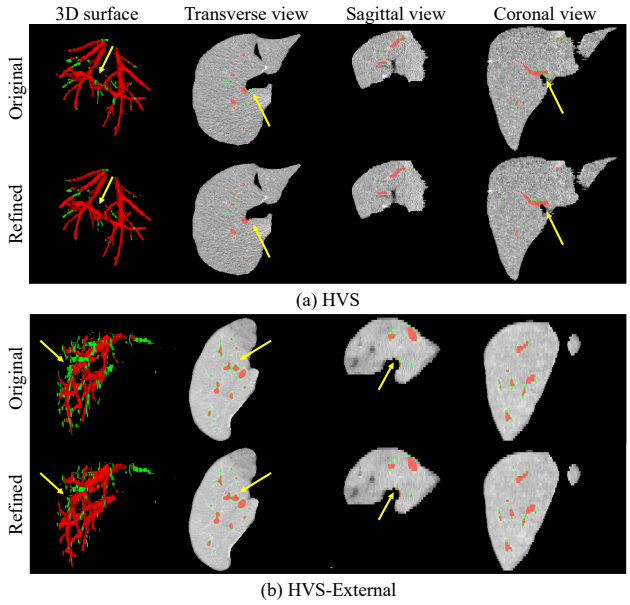


Figure 6. Visualization of hepatic vessel segmentation results using nnU-Net trained on both original and refined labels, in which the red indicates the segmentation and the green indicates the corresponding labels. Improvements are highlighted with yellow arrows.

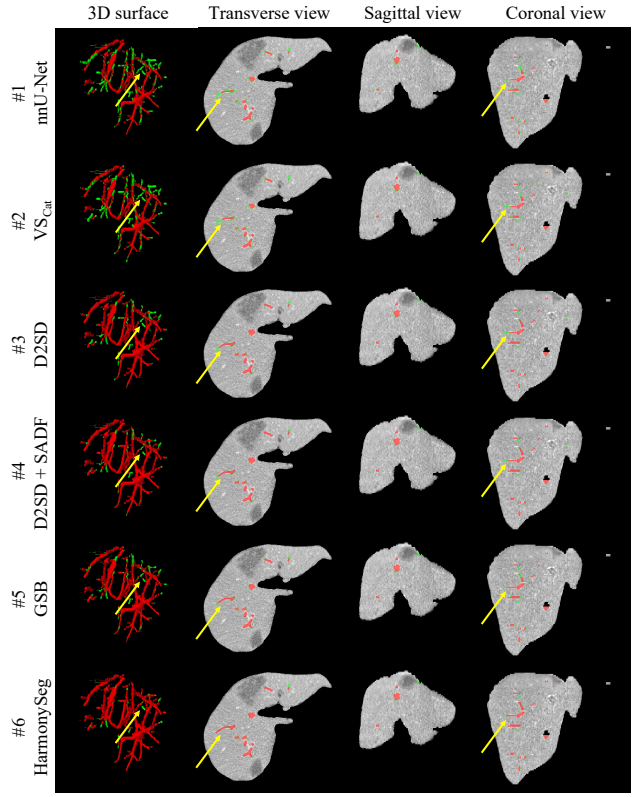


Figure 7. Visualization of hepatic vessel segmentation results in ablation study, with red indicating the segmentation and green representing the corresponding labels. Improvements are highlighted with yellow arrows.

learnable upsampling, whereas SQ module features are processed via cubic spline interpolation upsampling. Each segmentation layer, including the DMQ and SQ modules, produces an independent pre-softmax output. These outputs are then merged and refined through a convolutional layer to generate the final segmentation

D. Refined Hepatic Vessel Labels

In this paper, we use a combined liver vessel segmentation dataset called the HVS dataset. It is based on three publicly available datasets, including LiVS [4], MSD8 [13], and 3DIRCADb [14]. 532, 440, and 20 cases are available for the three datasets, respectively. The three publicly available datasets have made an impressive contribution to developing hepatic vessel segmentation models. However, some slices of the LiVS dataset are insufficiently labeled, and the

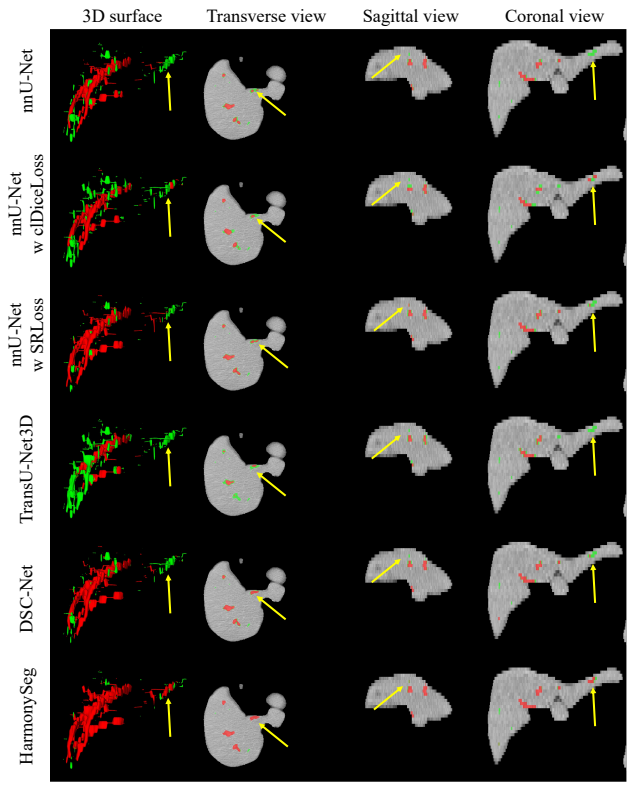


Figure 8. Visualization of hepatic vessel segmentation results in the HVS-External, with red indicating the segmentation and green representing the corresponding labels. Improvements are highlighted with yellow arrows.

labels of the test set of MSD8 are unavailable. The mentioned situations are reflected by labeled ratios (defined by the labeled slices divided by the total slice number of a 3D volume) in Table 1. Thus, to develop our model, we aim to make the best use of the data and refine these hepatic vessel labels. Fortunately, our clinical cooperator, after carefully checking the labels of the MSD8 training set, considered them to be relatively well labeled. Inspired by this, we first trained our model based on the training set of MSD8 and then used it to infer hepatic vessels of the LiVS dataset and the test set of MSD8. Subsequently, the pseudo labels were fused with the raw labels. Fused labels were checked again and manually corrected by a clinician, to serve as the final hepatic vessel labels in the HVS task. From Table 1, it can be found that the ratio of labeled slices has been significantly improved after our optimization, especially the LiVS dataset. Moreover, more visualization examples are given in Figures 3 and 4. Due to the cropping of the CT volume by the organizers of the dataset, the presence of some lesions, such as tumors, and the slice thickness, the continuity of the refined vessel labels is not fully ensured. Still, they are significantly improved compared to the original ones. We also compare the baseline performance trained by the original labels and the refined ones. As indicated by the evaluation metrics in Table 2 and the visualization examples in Figure 6, the baseline trained by the refined labels performs better in the HVS task.

E. Ablation studies

Some visualization examples in ablation studies are shown in Figure 7, it can be seen that our D2SD strategy can extract vessels with diverse sizes more effectively compared to the baseline. Moreover, liver vessel segmentation can not benefit from the simple concatenation fusion between the images and corresponding vesselness filtering results. Instead, our SADF fusion module can better utilize the vesselness filtering result to improve the segmentation accuracy. Besides, it can be observed that the GSB further preserves a reasonable continuity of the vessel tree.

F. Analysis on HVS-External

In the HVS-External, we included cases with various liver diseases, including two cases of fatty liver, four cases of cirrhosis, twelve cases of liver tumors, and three healthy livers. We analyze the results of HVS-External based on the disease stratification, as shown in Table 3. It can be found that the HarmonySeg achieves the highest mean Dice for patients with fatty liver, tumors, and healthy individuals, and mean HDs are competitive compared with other methods. Furthermore, visualization examples are demonstrated in Figure 8. The results indicate the robustness of HarmonySeg to various liver diseases and the potential to be applied in clinical practices.

G. Robustness discussion

We recognize that the reconnection loss may introduce noise. To address this, we observe that the performance gains of our loss functions following this order: $\mathcal{L}_{\text{sup-r}}(2.94\%) > \mathcal{L}_{\text{mix}}(2.01\%) > \mathcal{L}_{\text{spatial}}(1.67\%) > \mathcal{L}_{\text{con}}(1.01\%)$. The first three losses ($\mathcal{L}_{\text{sup-r}}$, $\mathcal{L}_{\text{mix}}$, $\mathcal{L}_{\text{spatial}}$) are robust and applicable to various scenarios. In contrast, the reconnection loss $\mathcal{L}_{\text{con}}$ is specifically designed to address missing vessel segments. To enhance its robustness, we employ two strategies: (a) We perform skeletonization on the defined reconnect branches, reducing their pixel width to 1, as shown in Eq.(6) of manuscript. Consequently, the loss applied to these pixels remains slight on average. (b) If incorrect pixels are mistakenly defined for reconnection, they can be effectively suppressed by the strong regularization from spatial relationships and mix augmentation invariance. Thus, we incorporate the reconnection loss as an additional strategy tailored for vessel segmentation tasks.

Table 4. **Ablations on recall and precision trade-off**: $\mathcal{L}^+$ for growth, $\mathcal{L}^-$ for suppression.

$\mathcal{L}^+_{\text{r-sup}}$	$\mathcal{L}^+_{\text{con}}$	$\mathcal{L}^-_{\text{spatial}}$	$\mathcal{L}^-_{\text{mix}}$	Recall (%)	Precision (%)	F1-score (%)
-	-	-	-	49.14	**84.12**	62.04
✓				55.57	79.12	63.09
✓	✓			**64.35**	71.20	<u>65.01</u>
✓		✓		50.33	79.92	59.68
✓			✓	53.58	<u>80.15</u>	62.15
✓	✓	✓	✓	<u>59.93</u>	73.89	**66.18**

H. Trade-off between precision and recall

We also analyze the recall-precision trade-off. The recall rates for different loss combinations on HVS are presented in Table 4. Recall is enhanced through relaxed supervision ($\mathcal{L}_{\text{r-sup}}$) and branch reconnection ($\mathcal{L}_{\text{con}}$), while noise is reduced via spatial consistency ($\mathcal{L}_{\text{spatial}}$) and mix equivalence ($\mathcal{L}_{\text{mix}}$). We achieved the best trade-off and the highest F1-score when combining all loss functions.

References

[1] Gady Agam, Samuel G Armato, and Changhua Wu. Vessel tree reconstruction in thoracic ct scans with application to nodule detection. *IEEE transactions on medical imaging*, 24 (4):486–499, 2005. 1

[2] Jieneng Chen, Jieru Mei, Xianhang Li, Yongyi Lu, Qihang Yu, Qingyue Wei, Xiangde Luo, Yutong Xie, Ehsan Adeli, Yan Wang, et al. Transunet: Rethinking the u-net architecture design for medical image segmentation through the lens of transformers. *Medical Image Analysis*, page 103280, 2024. 2

[3] Alejandro F Frangi, Wiro J Niessen, Koen L Vincken, and Max A Viergever. Multiscale vessel enhancement filtering. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI'98: First International Conference Cambridge, MA, USA, October 11–13, 1998 Proceedings 1*, pages 130–137. Springer, 1998. 1

[4] Zhan Gao, Qiuhao Zong, Yiqi Wang, Yan Yan, Yuqing Wang, Ning Zhu, Jin Zhang, Yunfu Wang, and Liang Zhao. Laplacian salience-gated feature pyramid network for accurate liver vessel segmentation. *IEEE Transactions on Medical Imaging*, 42(10):3059–3068, 2023. 3

[5] Fabian Isensee, Paul F Jaeger, Simon AA Kohl, Jens Petersen, and Klaus H Maier-Hein. nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods*, 18(2):203–211, 2021. 2

[6] Tim Jerman, Franjo Pernuš, Boštjan Likar, and Žiga Špiclin. Enhancement of vascular structures in 3d and 2d angiographic images. *IEEE transactions on medical imaging*, 35 (9):2107–2118, 2016. 1

[7] Yannick Kirchhoff, Maximilian R Rokuss, Saikat Roy, Balint Kovacs, Constantin Ulrich, Tassilo Wald, Maximilian Zenk, Philipp Vollmuth, Jens Kleesiek, Fabian Isensee, et al. Skeleton recall loss for connectivity conserving and resource efficient segmentation of thin tubular structures. *arXiv preprint arXiv:2404.03010*, 2024. 2

[8] Jonas Lamy, Odyssée Merveille, Bertrand Kerautret, and Nicolas Passat. A benchmark framework for multiregion analysis of vesselness filters. *IEEE Transactions on Medical Imaging*, 41(12):3649–3662, 2022. 1

[9] Cristian Lorenz, I-C Carlsen, Thorsten M Buzug, Carola Fassnacht, and Jürgen Weese. Multi-scale line segmentation with automatic estimation of width, contrast and tangential direction in 2d and 3d medical images. In *International Conference on Computer Vision, Virtual Reality, and Robotics in Medicine*, pages 233–242. Springer, 1997. 1

[10] Yaolei Qi, Yuting He, Xiaoming Qi, Yuan Zhang, and Guanyu Yang. Dynamic snake convolution based on topological geometric constraints for tubular structure segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6070–6079, 2023. 2

[11] Suprosanna Shit, Johannes C Paetzold, Anjany Sekuboyina, Ivan Ezhov, Alexander Unger, Andrey Zhylka, Josien PW Pluim, Ulrich Bauer, and Bjoern H Menze. cldice-a novel topology-preserving loss function for tubular structure segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16560–16569, 2021. 2

[12] Karen Simonyan. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 2

[13] Amber L Simpson, Michela Antonelli, Spyridon Bakas, Michel Bilello, Keyvan Farahani, Bram Van Ginneken, Annette Kopp-Schneider, Bennett A Landman, Geert Litjens, Bjoern Menze, et al. A large annotated medical image dataset for the development and evaluation of segmentation algorithms. *arXiv preprint arXiv:1902.09063*, 2019. 3

[14] Luc Soler, Alexandre Hostettler, Vincent Agnus, Arnaud Charnoz, Jean-Baptiste Fasquel, Johan Moreau, Anne-Blandine Osswald, Mourad Bouhadjar, and Jacques Marescaux. 3d image reconstruction for comparison of algorithm database. *URL: https://www. ircad. fr/research/datasets/liver-segmentation-3d-ircadb-01*, 2010. 3

[15] Xiang Zhong, Hongbin Zhang, Guangli Li, and Donghong Ji. Do you need sharpened details? asking mmdc-net: multilayer multi-scale dilated convolution network for retinal vessel segmentation. *Computers in Biology and Medicine*, 150: 106198, 2022. 2