Anatomy-Aware Frequency-Attention Transformer Networks for Liver Couinaud CT/MR Segmentation

Wenkang Fan[1], Hao Fang[1,2], Rui Li[3], Yanduan Lin[3], Chao An[4], and Xiongbiao Luo[1,2,5,*]

[1] Department of Computer Science and Engineering, Xiamen University, Xiamen 361102, China
[2] National Institute for Data Science in Health and Medicine, School of Medicine, Xiamen University, Xiamen 361102, China
[3] Zhongshan Hospital, Xiamen University, Xiamen 361004, China
[4] Department of Interventional ultrasound, Chinese PLA General Hospital, Beijing 100853, China
[5] Discipline of Intelligent Instrument and Equipment, Xiamen University, Xiamen 361102, China. xiongbiao.luo@gmail.com, Asterisk indicates the corresponding author

Abstract. Accurate Couinaud segmentation of liver CT/MR is essential in helping surgeons perceive the positional relationship between liver anatomy and intrahepatic lesions to make surgical planning. Unfortunately, current conventional and deep-learning based methods remain challenges in accurate Couinaud segmentation since the segmentation boundaries of different categories depending on hepatic vascular information are hard to predict. This work proposes a new deeply learned framework called anatomy-aware frequency-attention transformer networks (AFATN) for Couinaud segmentation of liver anatomy which contains the hybrid anatomy-aware preprocessing and frequency-attention transformer networks (FATN). Specifically, our framework first uses hybrid anatomy-aware preprocessing to integrate the hybrid cues of liver contour and hepatic venous centerline, then effectively utilizes hybrid cues for accurate Couinaud segmentation through the frequency-attention transformer networks with omission re-detected loss function. Our segmentation model FATN uses transformers to extract local structure and global semantic features and further focus on the hybrid cues with frequency-attention mechanisms. The proposed method was evaluated on clinical CT data and compared with currently available deep learning approaches, with the experimental results demonstrating that our method outperforms other approaches especially in accurately segmenting the Couinaud boundaries.

Keywords: Couinaud segmentation · Liver segmentation · Hepatic vessels · Transformer · Frequency attention.

1 Introduction

The Couinaud segmentation standard is widely used to describe the functional anatomy of the liver, which divides the liver into eight distinct units based on its umbilical fissure and vascular supply, with each unit having its own inflow and outflow hepatic vessels [6]. Generating an accurate three-dimensional visualization of the Couinaud segmentation from Computed Tomography (CT) and Magnetic Resonance (MR) can help

surgeons intuitively perceive the positional relationship between hepatic segments and intrahepatic lesions, and determine suitable treatment between radiofrequency ablation and hepatectomy procedures [19]. Since manual annotation in clinics is time-consuming and labor-intensive, automatic and accurate Couinaud segmentation for surgical planning is certainly beneficial to surgeons and patients.

Conventional Couinaud segmentation contains two categories: voxel-based and plane-based. Voxel-based methods allocate liver voxels to these segments by considering the distances between their locations and a specific branch of the hepatic vein [23], which does not follow the standard of Couinaud segmentation. Plane-based approaches first detect anatomical geometric elements like landmarks and lines from the liver surface and vessels, then reconstruct hepatic segments by the planes made up of these elements. For instance, Pamulapati et al. [1] built skeletonization veins and depend on the orientation between points and skeleton to construct the functional segment, while Wang and Arya et al. [2, 20] employed key point detection for Couinaud classification.

Deep-learning based methods are more popular for automatic medical image segmentation. Convolutional neural networks (CNN) such as U-Net [5, 11, 12] and ResUnet [24] have achieved remarkable results for many segmentation tasks. CNN works well in extracting the local texture, while they have the limited ability to extract the global features because convolution only focuses on the surrounding regions instead of capturing long-distance relationships. To solve this problem, some works propose attention mechanisms to focus on crucial information of images [15, 21, 25, 26]. Recently, transformer-based models or integrated CNN with transformer models have become more popular and achieved great performance on many liver-related segmentation tasks [4, 9, 10, 16, 22].

However, existing methods still remain challenging for accurate Couinaud segmentation because the boundary of the hepatic segments depends on the location of the hepatic veins and portal veins and there is no significant contrast between the boundaries of the different categories. This work considers the importance of utilizing the hepatic vessels for Couinaud segmentation and proposes a new deeply learned anatomy-aware frequency-attention transformer framework (AFATN) to accurately obtain hepatic segments. The highlights of our work are clarified as follows. First, we propose a hybrid anatomy-aware preprocessing to extract hybrid cues of liver contour and hepatic vessels and highlight these cues on CT/MR images to help networks perform feature extraction, which applies the advantages of plane-based approaches to deep learning methods. Second, we propose the frequency-attention transformer networks (FATN) for anatomy-aware Couinaud segmentation which uses a transformer-based encoder to extract sufficient local texture and global semantic features and employs global frequency-attention to enhance the structural boundary, especially the hepatic vascular information.

2 METHODS

The proposed anatomy-aware frequency-attention transformer framework (Fig. 1) generally consists of two parts: (1) hybrid anatomy-aware preprocessing, and (2) frequency-attention transformer networks (Fig. 2) with omission re-detected loss function.

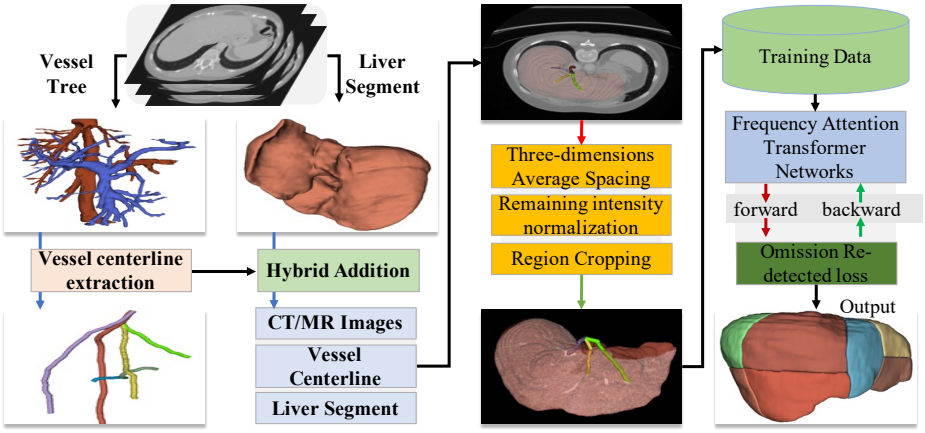

Fig. 1: Our new framework consists of hybrid anatomy-aware preprocessing and frequency-attention transformer networks (FATN) with omission re-detected loss.

2.1 Hybrid Anatomy-aware Preprocessing

The Couinaud segmentation boundary is based on the five planes consisting of the left umbilical fissure and four venous trunks (middle and right of the hepatic vein, left and right of the portal vein). Therefore, our framework aims to extract the vascular trunks and highlight this information in training CT/MR data, so that the segmentation network can better focus on this information to extract accurate hepatic segment boundaries.

For the input CT/MR images $\mathrm{I}(x,y,z)$, we first segment the liver mask $\mathrm{L}(x,y,z)$ through nnUNet [11] and segment hepatic and portal veins $\widehat{\mathrm{V}}(x,y,z)$ through [22]. Then, we employ the vessel centerline extraction approach from VTK [13] to extract hepatic venous trunks $\mathrm{V}(x,y,z)$ for segmented veins $\widehat{\mathrm{V}}(x,y,z)$. Note that all points on the extracted centerline will be selected according to their obtained vessel diameters, and points with relatively small vessel diameters are removed.

Now we have three intermediate outputs: the CT/MR input $\mathrm{I}(x,y,z)$, liver mask $\mathrm{L}(x,y,z)$ and intra-hepatic veins centerline $\mathrm{V}(x,y,z)$. The new CT/MR data $\mathrm{H}(x,y,z)$ can be obtained by a hybrid addition operation which integrates the original input with the extracted hybrid cues of liver mask and hepatic venous trunks

$$\mathrm{H}(x,y,z) = \begin{cases} 0, & \mathrm{L}(x,y,z) = False, \\ \text{maximum HU}, & \mathrm{L}(x,y,z) = True, \mathrm{V}(x,y,z) = True, \\ \mathrm{I}(x,y,z), & \mathrm{L}(x,y,z) = True, \mathrm{V}(x,y,z) = False, \end{cases} \tag{1}$$

where HU represents the intensity value range of $\mathrm{I}(x,y,z)$. After the hybrid addition operation, we averagely space the three-dimensions voxel (x,y,z) to (0.78, 0.78, 4.0) for all cases and use Z-score normalization to normalize the intensity value of $\mathrm{H}(x,y,z)$ to narrow the variability from different patients according to [11]

$$\Phi(x,y,z) = \psi_{(0.78,0.78,4.0)}(\mathrm{Norm}_z^{x,y}(\mathrm{H}(x,y,z))), \tag{2}$$

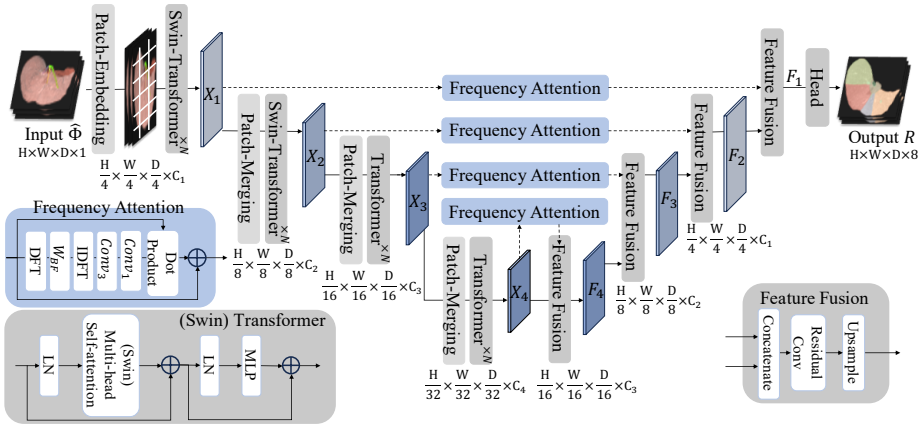

Fig. 2: Frequency-attention transformer networks (FATN) for Couinaud segmentation.

where ψ is uniform spacing operation and $\text{Norm}_z^{x,y}$ describes Z-score normalization.

Eventually, to reduce the size of input images and decrease unnecessary distractions of irrelevant regions, we perform region cropping operation $\varphi(\cdot)$ for all cases to exclude regions without liver to obtain the final training data $\hat{\Phi}$

$$\hat{\Phi} = \varphi(\Phi(x, y, z)) \, , if \; \mathrm{L}(x, y, z) = False. \tag{3}$$

2.2 Frequency-Attention Transformer Networks

We explore and analyze the performance of CNN and the transformer for medical image segmentation. The CNN modules such as residual block [24] can extract sufficient local texture and structure features while they have the limited ability to extract the global semantic features. On the contrary, the transformer [7] using multi-head self-attention can capture long-distance relationships to extract more global semantic features, and the sliding window-based attention [14] can also capture sufficient local features.

Therefore, we design a transformer-based feature extraction model that integrates the swin-transformer [14] and standard transformer [7] to extract both local and global features. Additionally, we propose a frequency-attention module to enhance the edge structural information of feature maps, which makes networks pay more attention to vessel characteristics of anatomy-aware CT/MR data to obtain more accurate boundaries of hepatic segments. The architecture of our FATN is shown in Fig. 2.

Encoder. The transformer-based encoder contains four stages. The first stage will first divide the input training data $\hat{\Phi} \in \mathbb{R}^{H \times W \times D \times 1}$ into 3-D patches $\in \mathbb{R}^{M \times (P^3 \times 1)}$, where P is the patch size we set to 4 and $M = (H \times W \times D)/P^3$ represents the number of patches. Then, each 3-D patch will be flattened and projected into C_1 dimension token through a trainable linear projection. The first stage finally employs N swin-transformer layers [14] with shifted-window multi-head self-attention (MSA) and multi-layer perceptron (MLP) to extract local texture features $X_1 \in \mathbb{R}^{M \times C_1}$ from these

tokens. After feature extraction, local features X_1 will be assembled into image-like feature maps $\in \mathbb{R}^{H/4\times W/4\times D/4\times C_1}$ for the next feature-extraction stage and feature-fusion decoder. The operation of the second to the fourth feature-extraction stage is almost similar to the first stage. All the patch-merging modules set the patch size to 2 for downsampling the feature map of the previous stage. Note that the second stage also uses swin-transformer layers [14] for local feature extraction, while the third and fourth use standard transformer layers [7] to extract multi-scale global features. After that, we obtain two local feature maps and two global feature maps with different sizes $\{X_1, X_2, X_3, X_4\}$ from the transformer-based encoder.

Decoder. The feature-fusion decoder can be divided into two steps. It first uses the global frequency-attention module to further highlight the noteworthy or essential boundary structural information on each feature map, then fuses these feature maps through the feature-fusion module from coarse to fine and predicts the final segmentation result by a head module. The operation of the frequency attention module on feature maps $X_i \in \mathbb{R}^{H_i\times W_i\times D_i\times C_i}$ can be formulated as

$$\hat{X}_i = X_i + W^i_{FA} \circ X_i, i = \{1, 2, 3, 4\}, \tag{4}$$

$$W^i_{FA} = Conv_1(Conv_3(f_{IDFT}(W_{BF} \circ (f_{DFT}(X_i))))), \tag{5}$$

$$W_{BF}(x, y, z) = 1 - exp(-\alpha(\frac{\sqrt{(x - x_c)^2 + (y - y_c)^2 + (z - z_c)^2}}{\sqrt{(H_i/2)^2 + (W_i/2)^2 + (D_i/2)^2}})^\beta) \tag{6}$$

where f_{DFT} and f_{IDFT} are 3-D discrete Fourier transform (DFT) and 3-D inverse discrete Fourier transform (IDFT), which are applied to each channel of the feature maps, W_{BF} is our designed bandpass filter weighted map which is used to filter out low-frequency information in the spectrum and preserve high-frequency edge structural information, x_c, y_c, and z_c represents the center position of map, α and β are hyper-parameters used to control the degree of low-frequency information filtering, "$\circ$" represents dot product operation, and $Conv_3$ and $Conv_1$ are $3 \times 3 \times 3$ and $1 \times 1 \times 1$ convolution operations to obtain frequency attentive map W_{FA}.

After feature maps $\{\hat{X}_1, \hat{X}_2, \hat{X}_3, \hat{X}_4\, X_4\}$ are obtained through the transformer-based encoder and the frequency attention module, four feature fusion modules corresponding to four stages of the encoder are used to fuse these features

$$F_4 = \mathcal{U}(\Theta(\otimes(X_4, \hat{X}_4))), \tag{7}$$

$$F_i = \mathcal{U}(\Theta(\otimes(F_{i+1}, \hat{X}_i))), i = 3, 2, 1, \tag{8}$$

where $\otimes$, Θ, and $\mathcal{U}$ represent channel-based concatenation, residual block [24], and upsample, respectively. Eventually, the head module with a residual block, upsample operation, and a $1 \times 1 \times 1$ convolution is to obtain the segmentation results

$$R = Conv_1(\mathcal{U}(\Theta(F_1))) \tag{9}$$

2.3 Omission Re-detected Loss

In the multi-class segmentation task, topological interactions between different foreground classes generally include two types of Containment and Exclusion according

to [8]. Because the relationship between different hepatic segments is excluded in the Couinaud segmentation task, we employ the convolution-based topological interaction module [8] to go through all pairs of adjacent voxels of the result R and identify the positions that violate the desired constraint

$$\mathrm{E}(x,y,z) = \begin{cases} 0, & \Pi_{R,G}(x,y,z) = False, \\ 1, & \Pi_{R,G}(x,y,z) = True, \end{cases} \tag{10}$$

where R and G describe the segmentation result and ground truth, respectively. $\Pi_{R,G} = True$ denotes the prediction violate the desired constraint, while $\Pi_{R,G} = False$ conversely. We re-detect the omission region $\mathrm{E}(x,y,z)$ and calculate the loss as follows.

We adopt a hybrid loss containing cross-entropy loss and dice coefficient loss

$$\mathcal{S}_L = \omega_{Dice}(R,G) + \omega_{BCE}(R,G), \tag{11}$$

$$\omega_{Dice}(R,G) = 1 - \frac{2\sum x_i \hat{x}_k}{\sum(x_k) + \sum(\hat{x}_k)}, \tag{12}$$

$$\omega_{BCE}(R,G) = -\sum(x_k \log(\hat{x}_k) + (1-x_k)\log(1-\hat{x}_k)), \tag{13}$$

where x_k and $\hat{x}_k$ are each pixel in G and R, respectively. Finally, the total omission re-detected Loss $\mathcal{OR}_L$ can be describe as:

$$\mathcal{OR}_L = (1 + \lambda_{\mathcal{E}}\mathrm{E}(x,y,z))\mathcal{S}_L(R,G) \tag{14}$$

where $\lambda_{\mathcal{E}}$ is the coefficient to measure the penalty level for the omission area.

3 Experimental settings

We collected 330 CT cases from Task08_HepaticVessel [17] in the Medical segmentation Decathlon study(180 cases with hepatic vessel labels and Couinaud labels), 3D-IRCADb-01 [18] dataset(20 cases with liver and hepatic vessel labels) and Liver Tumor Segmentation Benchmark(LiTS) [3](130 cases with liver labels). Couinaud labels of all 150 cases from 3D-IRCADb-01 and LiTS were annotated manually by three physicians.

We used Task08_HepaticVessel and 3D-IRCADb-01 to train our FATN to segment hepatic and portal veins of LiTS, and used 3D-IRCADb-01 and LiTS to train the networks [11] to obtain liver labels of Task08_HepaticVessel. The average spacing of all cases is around (0.78,0.78,4.0)mm^3 and we uniformed the size of 320 $\times$ 256 $\times$ 96 for all cases after region-cropping. For pre-processing and data augmentation of training data, we refer to nnU-Net [11]. The transformer layers N was 4 and the α, β, and $\lambda_{\mathcal{E}}$ were 5, 2, and 0.4, respectively. We set the learning rate from 10^{-4} to 10^{-3} and used a stochastic gradient descent algorithm as an optimizer with a momentum of 0.9 during training. We divided the collected data into training and testing data according to 7:3. The batch size, epoch, and iterations were set to 1, 500, and 1000, respectively.

We employ four popular metrics to evaluate the Couinaud segmentation of different methods: Dice similarity coefficient (DSC), intersection over union (IoU), precision, and recall. We compare our method anatomy-aware attention transformer networks (AFATN) with several 3-D segmentation methods: (1) Anatomical-aware point-voxel network (APVN) [25] which also highlights intra-hepatic veins for Couinaud

Table 1: Quantitative comparison of the hepatic vessel and liver Couinaud segmentation.

	Liver Couinaud				Hepatic Vessels			
Methods	DSC	IoU	Precision	Recall	DSC	IoU	Precision	Recall
APVN [25]	0.858	0.752	0.855	0.862	-	-	-	-
3D UNet [5]	0.887	0.799	0.895	0.882	0.814	0.687	0.818	0.811
UnetR++ [16]	0.910	0.836	0.928	0.894	0.819	0.694	0.822	0.817
TransUNnet [4]	0.935	0.879	0.942	0.930	0.837	0.720	**0.839**	0.836
FATN-(ours)	0.927	0.864	0.931	0.924	0.829	0.709	0.832	0.828
FATN(ours)	0.949	0.888	0.941	0.940	**0.840**	**0.724**	**0.839**	**0.842**
AFATN(ours)	**0.952**	**0.909**	**0.956**	**0.949**	-	-	-	-

segmentation, (2) 3-D UNet [5] which is a convolution-based model, (3) UnetR++ [16] which uses U-shape transformers to extract global features, (4) TransUNnet [4] which integrates convolution and transformer to extract local and global features with attention. Then, we conduct ablation studies to evaluate the effectiveness of our proposed anatomy-aware processing and frequency attention module. Therefore, we compare (5) FATN and (6) FATN-: without frequency attention. Additionally, we compare the effects of different methods on both hepatic vessels (hepatic veins and portal veins) and liver Couinaud segmentation.

4 Results and Discussion

Results. Fig. 3 visually compares the 3-D Couinaud segmentation results of different methods and Fig. 4 illustrates the selected 2-D slices of results. Although UnetR++ [16] and TransUNnet [4] predict little misclassification within hepatic segments, the boundaries of different categories are not accurate, especially near the middle and right hepatic veins (according to *Row* 2 in Fig. 3 and *Columns* 1$\sim$2 Fig. 4) and right portal veins (according to *Rows* 4$\sim$5 in Fig. 3). We can see our FATN performs better than UnetR++ [16] and TransUNnet [4] in boundaries prediction and our AFATN can predict more accurate hepatic segment boundaries that better fit the location of the left umbilical fissure and the hepatic and portal veins. Table 1 summarizes the quantitative results of different models. For both hepatic vessels and liver Couinaud segmentation, transformer-based models UnetR++ [16], TransUNnet [4], and our models perform better than convolution-based or local feature extraction models APVN [25] and 3D UNet [5]. Our FATN predicts higher DSC, IoU, and Recall than UnetR++ [16], TransUNnet [4], and FATN-. Meanwhile, our AFATN achieves better performance across various metrics compared to other methods, especially in terms of precision and recall.

Discussion. This work proposes a new deep-learning framework for Couinaud segmentation. The effectiveness of our method is as follows. First, our FATN extracts sufficient local texture and global structure features through transformers and employs global frequency attention to enhance the structural boundary, which is well-suited for segmenting objects with complex boundaries. Moreover, our AFATN uses hybrid anatomy-aware processing to extract and highlight the hybrid cues of liver contour and intra-hepatic veins, which can be effectively utilized for more accurate boundary seg-

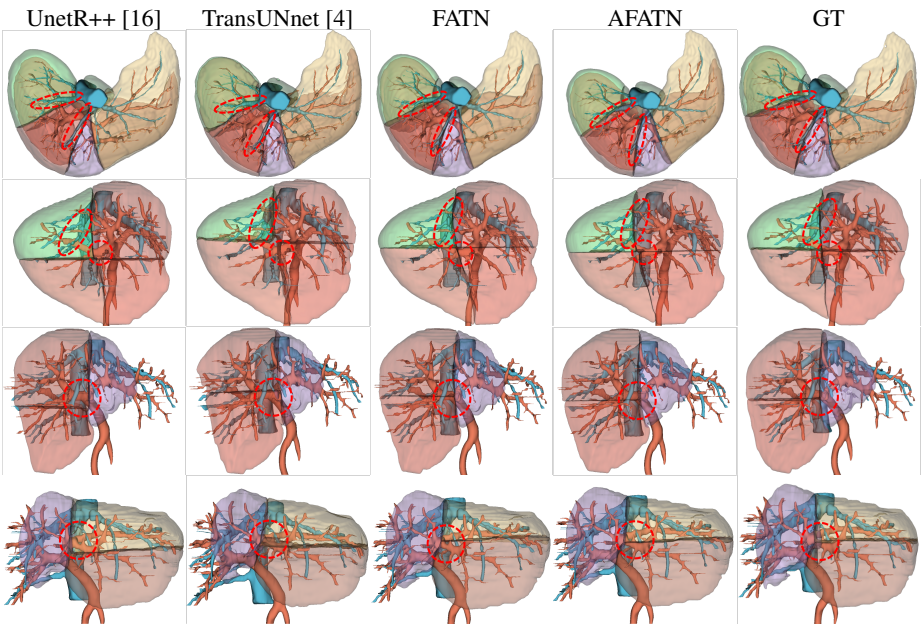

Fig. 3: Visual comparison of the Couinaud segmentation. *Rows* 2∼5 shows a vertical view of the result with hepatic veins, the results near the right hepatic veins, near the middle hepatic veins and left portal veins, and near the right portal veins, respectively.

mentation. Transformer-based methods perform much better than APVN [25] and 3D UNet [5] in our experiments since APVN uses small volumes for prediction and 3-D UNet only extracts local features. FATN- performs better than UnetR++ [16] because it integrates local and global features. Since both FATN and TransUNnet [4] employ attention modules to enhance extracted features, they achieve better results than FATN-. Our FATN achieves better results than TransUNnet [4] due to the effective frequency-attention mechanism. Our approach has several limitations. First, it is difficult to accurately segment some cases with irregular lesions. Moreover, the effectiveness of our method depends on the accurate hepatic vessel segmentation. However, vessel segmentation remains challenging due to the irregular distribution. Additionally, the proposed segmentation model has a large number of parameters and a high computational cost. We will optimize it using lightweight convolutional and self-attention modules.

Conclusion. In summary, this work proposes anatomy-aware frequency-attention transformer networks (AFATN) using the information of liver contour and hepatic vessels to achieve accurate Couinaud segmentation.

Acknowledgments. This work was supported in part by the National Natural Science Foundation of China under Grant 82272133, in part by the High-Quality Development Science and Technology Major Project of Xiamen Health Commission under Grant 2024GZL-ZD03, and in part by Ningbo 2035 Key Research and Development Program under Grant 2024Z127.

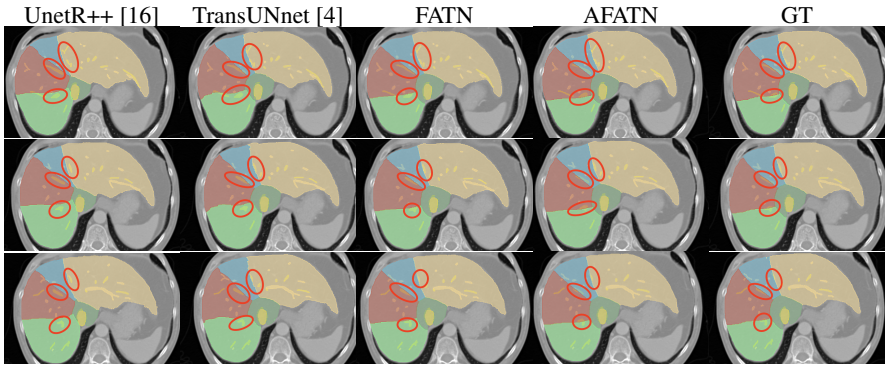

Fig. 4: Visual comparison of the Couinaud segmentation results of 2D slices using the four methods. Note that the *red circles* on the figures from top to down represent the position of the left umbilical fissure, and the middle and right hepatic veins, respectively.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Alirr, O.I., Abd. Rahni, A.A.: Automatic atlas-based liver segmental anatomy identification for hepatic surgical planning. International Journal of Computer Assisted Radiology and Surgery **15**, 239–248 (2020)
2. Arya, Z., Ridgway, G., Jandor, A., Aljabar, P.: Deep learning-based landmark localisation in the liver for couinaud segmentation. In: Medical Image Understanding and Analysis: 25th Annual Conference, MIUA 2021, Oxford, United Kingdom, July 12–14, 2021, Proceedings 25. pp. 227–237. Springer (2021)
3. Bilic, P., Christ, P., Vorontsov, E., Chlebus, G., Chen, H., Dou, Q., Fu, C., Han, X., Heng, P., Hesser, J., et al.: The liver tumor segmentation benchmark (lits). 1–43. arXiv preprint arXiv:1901.04056 (2019)
4. Chen, J., Mei, J., Li, X., Lu, Y., Yu, Q., Wei, Q., Luo, X., Xie, Y., Adeli, E., Wang, Y., et al.: Transunet: Rethinking the u-net architecture design for medical image segmentation through the lens of transformers. Medical Image Analysis **97**, 103280 (2024)
5. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3d u-net: learning dense volumetric segmentation from sparse annotation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19. pp. 424–432. Springer (2016)
6. Couinaud, C.: Liver anatomy: portal (and suprahepatic) or biliary segmentation. Digestive Surgery **16**(6), 459–467 (1999)
7. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020)
8. Gupta, S., Hu, X., Kaan, J., Jin, M., Mpoy, M., Chung, K., Singh, G., Saltz, M., Kurc, T., Saltz, J., et al.: Learning topological interactions for multi-class medical image segmentation. In: European Conference on Computer Vision. pp. 701–718. Springer (2022)

9. He, Y., Nath, V., Yang, D., Tang, Y., Myronenko, A., Xu, D.: Swinunetr-v2: Stronger swin transformers with stagewise convolutions for 3d medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 416–426. Springer (2023)
10. Heidari, M., Kazerouni, A., Soltany, M., Azad, R., Aghdam, E.K., Cohen-Adad, J., Merhof, D.: Hiformer: Hierarchical multi-scale representations using transformers for medical image segmentation. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 6202–6212 (2023)
11. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. Nature methods **18**(2), 203–211 (2021)
12. Isensee, F., Wald, T., Ulrich, C., Baumgartner, M., Roy, S., Maier-Hein, K., Jaeger, P.F.: nnu-net revisited: A call for rigorous validation in 3d medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 488–498. Springer (2024)
13. Izzo, R., Steinman, D., Manini, S., Antiga, L.: The vascular modeling toolkit: a python library for the analysis of tubular structures in medical images. Journal of Open Source Software **3**(25), 745 (2018)
14. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B.: Swin transformer: Hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 10012–10022 (2021)
15. Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., et al.: Attention u-net: Learning where to look for the pancreas. arxiv 2018. arXiv preprint arXiv:1804.03999 (1804)
16. Shaker, A.M., Maaz, M., Rasheed, H., Khan, S., Yang, M.H., Khan, F.S.: Unetr++: delving into efficient and accurate 3d medical image segmentation. IEEE Transactions on Medical Imaging (2024)
17. Simpson, A.L., Antonelli, M., Bakas, S., Bilello, M., Farahani, K., Van Ginneken, B., Kopp-Schneider, A., Landman, B.A., Litjens, G., Menze, B., et al.: A large annotated medical image dataset for the development and evaluation of segmentation algorithms. arXiv preprint arXiv:1902.09063 (2019)
18. Soler, L., Hostettler, A., Agnus, V., Charnoz, A., Fasquel, J.B., Moreau, J., Osswald, A.B., Bouhadjar, M., Marescaux, J.: 3d image reconstruction for comparison of algorithm database. URL: https://www. ircad. fr/research/data-sets/liver-segmentation-3d-ircadb-01 (2010)
19. Wakabayashi, G., Cherqui, D., Geller, D.A., Abu Hilal, M., Berardi, G., Ciria, R., Abe, Y., Aoki, T., Asbun, H.J., Chan, A.C., et al.: The tokyo 2020 terminology of liver anatomy and resections: Updates of the brisbane 2000 system. Journal of Hepato-Biliary-Pancreatic Sciences **29**(1), 6–15 (2022)
20. Wang, M., Jin, R., Lu, J., Song, E., Ma, G.: Automatic ct liver couinaud segmentation based on key bifurcation detection with attentive residual hourglass-based cascaded network. Computers in Biology and Medicine **144**, 105363 (2022)
21. Woo, S., Park, J., Lee, J.Y., Kweon, I.S.: Cbam: Convolutional block attention module. In: Proceedings of the European conference on computer vision (ECCV). pp. 3–19 (2018)
22. Wu, Y., Shen, D., Jin, J., Xu, G., Chen, Y., Luo, X.: Local-global progressive u-transformers for accurate hepatic and portal veins segmentation in abdominal mr images. In: ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 1–5. IEEE (2023)
23. Yang, X., Do Yang, J., Hwang, H.P., Yu, H.C., Ahn, S., Kim, B.W., You, H.: Segmentation of liver and vessels from ct images and classification of liver segments for preoperative liver

surgical planning in living donor liver transplantation. Computer Methods and Programs in Biomedicine **158**, 41–52 (2018)
24. Yu, W., Fang, B., Liu, Y., Gao, M., Zheng, S., Wang, Y.: Liver vessels segmentation based on 3d residual u-net. In: 2019 IEEE International Conference on Image Processing (ICIP). pp. 250–254. IEEE (2019)
25. Zhang, X., Liu, Y., Ali, S., Zhao, X., Sun, M., Han, M., Liu, T., Zhai, P., Cui, Z., Zhang, P., et al.: Anatomical-aware point-voxel network for couinaud segmentation in liver ct. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 465–474. Springer (2023)
26. Zhou, W., Zhang, X., Gu, D., Wang, S., Huo, J., Zhang, R., Jiang, Z., Shi, F., Xue, Z., Zhan, Y., et al.: Henet: Hierarchical enhancement network for pulmonary vessel segmentation in non-contrast ct images. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 551–560. Springer (2023)