



Anatomical-Aware Point-Voxel Network for Couinaud Segmentation in Liver CT

Xukun Zhang¹, Yang Liu¹, Sharib Ali², Xiao Zhao¹, Mingyang Sun¹, Minghao Han¹, Tao Liu¹, Peng Zhai¹, Zhiming Cui^{3(✉)}, Peixuan Zhang⁴, Xiaoying Wang^{5(✉)}, and Lihua Zhang^{1(✉)}

¹ Academy for Engineering and Technology, Fudan University, Shanghai, China
lihuazhang@fudan.edu.cn

² School of Computing, University of Leeds, Leeds, UK

³ School of Biomedical Engineering, ShanghaiTech University, Shanghai, China
cuizhm@shanghaitech.edu.cn

⁴ Changchun Boli Technologies Co., Ltd., Jilin, China

⁵ Department of Liver Surgery, Key Laboratory of Carcinogenesis and Cancer Invasion of Ministry of Education, Liver Cancer Institute, Zhongshan Hospital, Fudan University, Shanghai, China
xiaoyingwang@fudan.edu.cn

Abstract. Accurately segmenting the liver into anatomical segments is crucial for surgical planning and lesion monitoring in CT imaging. However, this is a challenging task as it is defined based on vessel structures, and there is no intensity contrast between adjacent segments in CT images. In this paper, we propose a novel point-voxel fusion framework to address this challenge. Specifically, we first segment the liver and vessels from the CT image, and generate 3D liver point clouds and voxel grids embedded with vessel structure prior. Then, we design a multi-scale point-voxel fusion network to capture the anatomical structure and semantic information of the liver and vessels, respectively, while also increasing important data access through vessel structure prior. Finally, the network outputs the classification of Couinaud segments in the continuous liver space, producing a more accurate and smooth 3D Couinaud segmentation mask. Our proposed method outperforms several state-of-the-art methods, both point-based and voxel-based, as demonstrated by our experimental results on two public liver datasets. Code, datasets, and models are released at <https://github.com/xukun-zhang/Couinaud-Segmentation>.

Keywords: Couinaud Segmentation · Point-Voxel Network · Liver CT

X. Zhang and Y. Liu—Contributed equally.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-43898-1_45.

1 Introduction

Primary liver cancer is one of the most common and deadly cancer diseases in the world, and liver resection is a highly effective treatment [11,14]. The Couinaud segmentation [7] based on CT images divides the liver into eight functionally independent regions, which intuitively display the positional relationship between Couinaud segments and intrahepatic lesions, and helps surgeons for make surgical planning [3,13]. In clinics, Couinaud segments obtained from manual annotation are tedious and time-consuming, based on the vasculature used as rough guide (Fig. 1). Thus, designing an automatic method to accurately segment Couinaud segments from CT images is greatly demanded and has attracted tremendous research attention.

However, automatic and accurate Couinaud segmentation from CT images is a challenging task. Since it is defined based on the anatomical structure of live vessels, even no intensity contrast (Fig. 1.(b)) can be observed between different Couinaud segments, and the uncertainty of boundary (Fig. 1.(d)) often greatly affect the segmentation performance. Previous works [4,8,15,19] mainly rely on handcrafted features or atlas-based models, and often fail to robustly handle those regions with limited features, such as the boundary between adjacent Couinaud segments. Recently, with the advancement of deep learning [5,10,18], many CNN-based algorithms perform supervised training through pixel-level Couinaud annotations to automatically obtain segmentation results [1,9,21]. Unfortunately, the CNN models treat all voxel-wise features in the CT image equally, cannot effectively capture key anatomical regions useful for Couinaud segmentation. In addition, all these methods deal with the 3D voxels of the liver directly without considering the spatial relationship of the different Couinaud segments, even if this relationship is very important in Couinaud segmentation. It can supplement the CNN-based method and improve the segmentation performance in regions without intensity contrast.

In this paper, to tackle the aforementioned challenges, we propose a point-voxel fusion framework that represents the liver CT in continuous points to better learn the spatial structure, while performing the convolutions in voxels to obtain the complementary semantic information of the Couinaud segments.

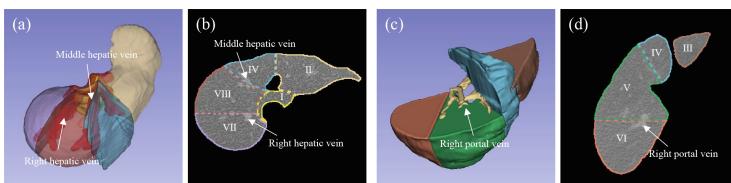


Fig. 1. Couinaud segments (denoted as Roman numbers) in relation to the liver vessel structure. (a) and (b) briefly several Couinaud segments separated by the hepatic vein. (c) and (d) show several segments surrounded by the portal vein, which are divided by the course of the hepatic vein.

Specifically, the liver mask and vessel attention maps are first extracted from the CT images, which allows us to randomly sample points embedded with vessel structure prior in the liver space and voxelize them into a voxel grid. Subsequently, points and voxels pass through two branches to extract features. The point-based branch extracts the fine-grained feature of independent points and explores spatial topological relations. The voxel-based branch is composed of a series of convolutions to learn semantic features, followed by de-voxelization to convert them back to points. Through the operation of voxelization and de-voxelization at different resolutions, the features extracted by these two branches can achieve multi-scale fusion on point-based representation, and finally output the Couinaud segment category of each point. Extensive experiments on two publicly available datasets named **3Dircadb** [20] and **LITS** [2] demonstrate that our proposed framework achieves state-of-the-art (SOTA) performance, outperforming cutting-edge methods quantitatively and qualitatively.

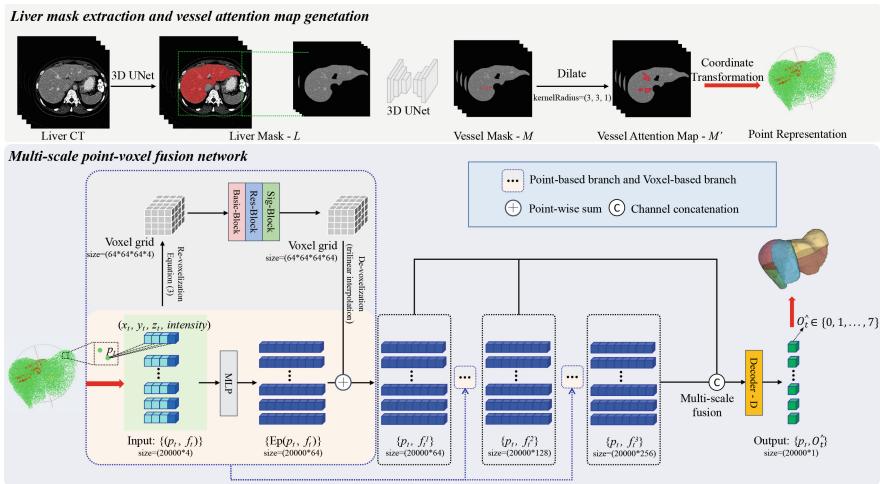


Fig. 2. Overall framework of our proposed method for Couinaud segmentation.

2 Method

The overview of our framework to segment Couinaud segments from CT images is shown in Fig. 2, including the liver segmentation, vessel attention map generation, point data sampling and multi-scale point-voxel fusion network.

2.1 Liver Mask and Vessel Attention Map Generation

Liver segmentation is a fundamental step in Couinaud segmentation task. Considering that the liver is large and easy to identify in the abdominal organs, we

extracted the liver mask through a trained 3D UNet [6]. Different from liver segmentation, since we aim to use the vessel structure as a rough guide to improving the performance of the Couinaud segmentation, we employ another 3D UNet [6] to generate the vessel attention map more easily. Specifically, given a 3D CT image containing only the area covered by the liver mask (L), the 3D UNet [6] output a binary vessel mask (M). A morphological dilation is then used to enclose more vessel pixels in the M -covered area, generating a vessel attention map (M'). We employ the BCE loss to supervise the learning process.

2.2 Couinaud Segmentation

Based on the above work, we first use the M' and the L to sample get point data, which can convert into a voxel grid through re-voxelization. The converted voxel grid embeds the vessel prior and also dilutes the liver parenchyma information. Inspired by [12], a novel multi-scale point-voxel fusion network then is proposed to simultaneously process point and voxel data through point-based branch and voxel-based branch, respectively, aiming to accurately perform Couinaud segmentation. The details of this part of our method are described below.

Continuous Spatial Point Sampling Based on the Vessel Attention Map. In order to obtain the topological relationship between Couinaud segments, a direct strategy is to sample the coordinate point data with 3D spatial information from liver CT and perform point-wise classification. Hence, we first convert the image coordinate points $I = \{i_1, i_2, \dots, i_t, i_t \in \mathbb{R}^3\}$ in liver CT into the world coordinate points $P = \{p_1, p_2, \dots, p_t, p_t \in \mathbb{R}^3\}$:

$$P = I * Spacing * Direction + Origin, \quad (1)$$

where $Spacing$ represents the voxel spacing in the CT images, $Direction$ represents the direction of the scan, and $Origin$ represents the world coordinates of the image origin. Based on equation(1), we obtain the world coordinate $p_t = (x_t, y_t, z_t)$ corresponding to each point i_t in the liver space. However, directly feeding the transformed point data as input into the point-based branch undoubtedly ignores the vessel structure, which is crucial for Couinaud segmentation. To solve this issue, we propose a strategy of continuous spatial sampling point data based on the M' . Specifically, the model randomly samples T points in each training epoch, of which $T/2$ points fall in the smaller space covered by the M' , which enables the model to increase access to important data in the region during training. In addition, we apply a random perturbation $Offset = (\Delta x, \Delta y, \Delta z)$ in the range of $[-1, 1]$ to each point $p_t = (x_t, y_t, z_t) \in M'$ in this region to obtain a new point $p_t = (x_t + \Delta x, y_t + \Delta y, z_t + \Delta z)$, and the intensity in this coordinate obtained by trilinear interpolation. In the training stage of the network, the label of the point $p_t = (x_t + \Delta x, y_t + \Delta y, z_t + \Delta z)$ is generated by:

$$O_t = O_t(R(x_t + \Delta x), R(y_t + \Delta y), R(z_t + \Delta z)) \in \{0, 1, \dots, 7\}, \quad (2)$$

where R denotes the rounding integer function. Based on this, we achieve arbitrary resolution sampling in the continuous space covered by the M' .

Re-voxelization. It is not enough to extract the topological information and fine-grained information of independent points only by point-based branch for accurate Couinaud segmentation. To this end, we transform the point data $\{(p_t, f_t)\}$ into voxel grid $\{V_{u,v,w}\}$ by re-voxelization, where $f_t \in \mathbb{R}^c$ is the feature corresponding to point p_t , aiming to voxel-based convolution to extract complementary semantic information in the grid. Specifically, we first normalize the coordinates $\{p_t\}$ to $[0, 1]$, which is denoted as $\{\hat{p}_t\}$. Note that the point features $\{f_t\}$ remain unchanged during the normalization. Then, we transform the normalized point cloud $\{(\hat{p}_t, f_t)\}$ into the voxel grids $\{V_{u,v,w}\}$ by averaging all features f_t whose coordinate $\hat{p}_t = (\hat{x}_t, \hat{y}_t, \hat{z}_t)$ falls into the voxel grid (u, v, w) :

$$\mathbf{V}_{u,v,w,c} = \frac{\sum_{t=1}^n \mathbb{I}[R(\hat{x}_t * r) = u, R(\hat{y}_t * r) = v, R(\hat{z}_t * r) = w] * f_{t,c}}{N_{u,v,w}}, \quad (3)$$

where r denotes the voxel resolution, $\mathbb{I}[\cdot]$ is the binary indicator of whether the coordinate \hat{p}_t belongs to the voxel grid (u, v, w) , $f_{t,c}$ denotes the $c - th$ channel feature corresponding to \hat{p}_t , and $N_{u,v,w}$ is the number of points that fall in that voxel grid. Note that the re-voxelization in the model is used three times (as shown in Fig. 2), and the $f_{t,c}$ in the first operation is the coordinate and intensity, with $c = 4$. Moreover, due to the previously mentioned point sampling strategy, the converted voxel grid also inherits the vessel structure from the point data and dilutes the unimportant information in the CT images.

Multi-scale Point-Voxel Fusion Network. Intuitively, due to the image intensity between different Couinaud segments being similar, the voxel-based CNN model is difficult to achieve good segmentation performance. We propose a multi-scale point-voxel fusion network for accurate Couinaud segmentation, take advantage of the topological relationship of coordinate points in 3D space, and leverage the semantic information of voxel grids. As shown in Fig. 2, our method has two branches: point-based and voxel-based. The features extracted by these two branches on multiple scales are fused to provide more accurate and robust Couinaud segmentation performance. Specifically, in the point-based branch, the input point data $\{(p_t, f_t)\}$ passes through an MLP, denoted as E_p , which aims to extract fine-grained features with topological relationships. At the same time, the voxel grid $\{V_{u,v,w}\}$ passes the voxel branch based on convolution, denoted as E_v , which can aggregate the features of surrounding points and learn the semantic information in the liver 3D space. We re-transform the features extracted from the voxel-based branch to point representation through trilinear interpolation, to combine them with fine-grained features extracted from the point-based branch, which provide complementary information:

$$(p_t, f_t^1) = E_p(P(p_t, f_t)) + tri(E_v(V))_t, \quad (4)$$

where the superscript 1 of (p_t, f_t^1) indicates that the fused point data and corresponding features f_t^1 are obtained after the first round of point-voxel operation. Then, the point data (p_t, f_t^1) is voxelized again and extracted point features and voxel features through two branches. Note that the resolution of the voxel grid in

this round is reduced to half of the previous round. After three rounds of point-voxel operations, we concatenate the original point feature f_t and the features $\{f_t^1, f_t^2, f_t^3\}$ with multiple scales, then send them into a point-wise decoder D , parameterized by a fully connected network, to predict the corresponding Couinaud segment category:

$$\hat{O}_t = D(\text{cat} \{f_t, f_t^1, f_t^2, f_t^3\}) \in \{0, 1, \dots, 7\}, \quad (5)$$

where $\{0, 1, \dots, 7\}$ denotes the Couinaud segmentation category predicted by our model for the point p_t . We employ the BCE loss and the Dice loss to supervise the learning process. More method details are shown in the supplementary materials.

3 Experiments

3.1 Datasets and Evaluation Metrics

We evaluated the proposed framework on two publicly available datasets, 3Diradb [20] and LiTS [2]. The 3Diradb dataset [20] contains 20 CT images with spacing ranging from 0.56 mm to 0.87 mm, and slice thickness ranging from 1 mm to 4 mm with liver and liver vessel segmentation labels. The LiTS dataset [2] consists of 200 CT images, with a spacing of 0.56 mm to 1.0 mm and slice thickness of 0.45 mm to 6.0 mm, and has liver and liver tumour labels, but without vessels. We annotated the 20 subjects of the 3Diradb dataset [20] with the Couinaud segments and randomly divided 10 subjects for training and another 10 subjects for testing. For LiTS dataset [2], we observed the vessel structure on CT images, annotated the Couinaud segments of 131 subjects, and randomly selected 66 subjects for training and 65 for testing.

We have used three widely used metrics, i.e., accuracy (ACC, in %), Dice similarity metric (Dice, in %), and average surface distance (ASD, in mm) to evaluate the performance of the Couinaud segmentation.

3.2 Implementation Details

The proposed framework was implemented on an RTX8000 GPU using PyTorch. Based on the liver mask has been extracted, we train a 3D UNet [6] on the 3Diradb dataset [20] to generate the vessel attention map of two datasets. Then, we sample $T = 20,000$ points in each epoch to train our proposed multi-scale point-voxel fusion network each epoch. We perform scaling within the range of 0.9 to 1.1, arbitrary axis flipping, and rotation in the range of 0 to 5°C on the input point data as an augmentation strategy. Besides, we use the stochastic gradient descent optimizer with a learning rate of 0.01, which is reduced to 0.9 times for every 50 epochs of training. All our experiments were trained 400 epochs, with a random seed was 2023, and then we used the model with the best performance on the training set to testing.

Table 1. Quantitative comparison with different segmentation methods. ACC and Dice are the averages of all testing subjects, while ASD is the average of the average performance of all segments in all testing subjects.

Method	3Dircadb [20]			LiTS [2]		
	Acc (%) ↑	Dice (%) ↑	ASD (mm ↓)	Acc (%) ↑	Dice (%) ↑	ASD (mm ↓)
3D UNet [6]	71.05	82.19	9.13	80.50	88.47	5.81
PointNet2Plus [16]	57.96	72.41	13.54	78.75	88.01	6.88
Jia <i>et al.</i> 's [9]	59.74	72.69	15.61	64.83	77.59	14.71
Ours	82.42	90.29	5.49	85.51	92.12	5.18

3.3 Comparison with State-of-the-Art Methods

We compare our framework with several SOTA approaches, including voxel-based 3D UNet [6], point-based PointNet2Plus [16], and the methods of Jia *et al.* [9]. The method of Jia *et al.* is a 2D UNet [17] with dual attention to focus on the boundary of the Couinaud segments and is specifically used for the Couinaud segmentation task. We use PyTorch to implement this model and maintain the same implementation details as other methods. The quantitative and qualitative comparisons are shown in Table 1 and Fig. 3, respectively.

Quantitative Comparison. Table 1 summarizes the overall comparison results under three metrics. By comparing the first two rows, we can see that PointNet2Plus [16] and 3D UNet [6] have achieved close performance in the LiTS dataset [2], which demonstrates the potential of the point-based methods in the Couinaud segmentation task. In addition, the third row shows that Jia *et al.*'s [9] 2D UNet [17] as the backbone method performs worst on all metrics, further demonstrating the importance of spatial relationships. Finally, our proposed point-voxel fusion segmentation framework achieves the best performance. Especially on the 3Dircadb dataset [20] with only 10 training subjects, the ACC and Dice achieved by our method exceed PointNet2Plus [16] and 3D UNet [6] by nearly 10 points, and the ASD is also greatly reduced, which demonstrates the effectiveness of the combining point-based and voxel-based methods.

Qualitative Comparison. To further evaluate the effectiveness of our method, we also provide qualitative results, as shown in Fig. 3. The first two rows show that the vessel structure is used as the boundary guidance for Couinaud segmentation, but voxel-based 3D UNet [6] fails to accurately capture this key structural relationship, resulting in inaccurate boundary segmentation. Note that, it can be seen that our method can learn the boundary guidance provided by the portal vein (the last two rows), to deal with the uncertain boundary more robustly. Besides, compared with the 3D view, it is obvious that the voxel-based CNN methods are easy to pay attention to the local area and produce a large area of error segmentation, so the reconstructed surface is uneven. The point-based method obtains smooth 3D visualization results, but it is more likely to cause

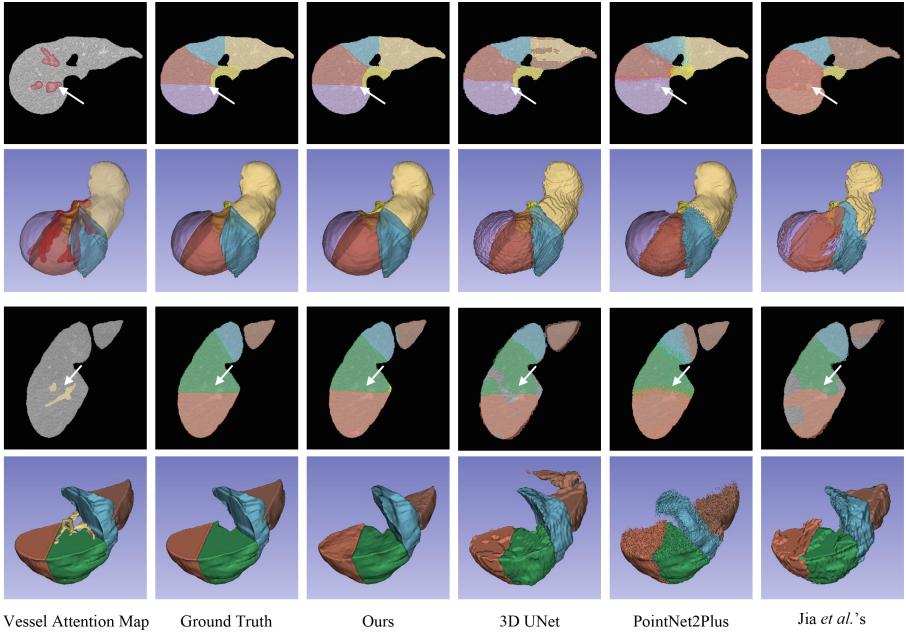


Fig. 3. Comparison of segmentation results of different methods. Different colours represent different Couinaud segments. The first two rows show a subject from the LiTS dataset [2], and the last two rows show another subject from the 3Diradb dataset [20]. In the first column, we show the vessel attention map in 2D and 3D views as an additional reference for the segmentation results.

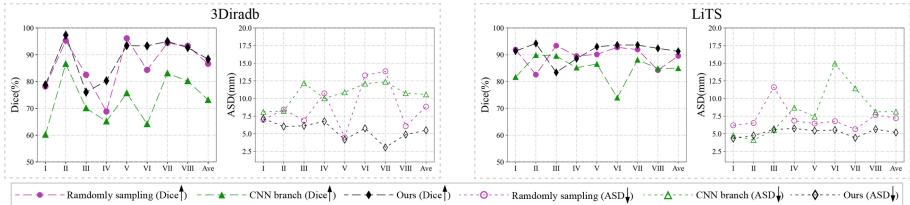


Fig. 4. Ablation Study on Two Key Components. Different Table 1, here ACC and Dice are the average performance of each Couinaud segment.

segmentation blur in boundary areas with high uncertainty. Our method combines the advantages of point-based and voxel-based methods, and remedies their respective defects, resulting in smooth and accurate Couinaud segmentation.

3.4 Ablation Study

To further study the effectiveness of our proposed framework, we compared two ablation experiments: 1) random sampling of T points in the liver space, with-

out considering the guidance of vascular structure, and 2) considering only the voxel-based branch, where the Couinaud segments mask is output by a CNN decoder. Figure 4 shows the ablation experimental results obtained on all the Couinaud segments of two datasets, under the Dice and the ASD metrics. It can be seen that our full method is significantly better than the CNN branch joint decoder method on both metrics of two datasets, which demonstrates the performance gain by the combined point-based branch. In addition, compared with the strategy of random sampling, our full-method reduces the average ASD by more than $2mm$ on eight Couinaud segments. This is because to the vessel structure-guided sampling strategy can increase the important data access between the boundaries of the Couinaud segments. Besides, perturbations are applied to the points in the coverage area of the vessel attention map, so that our full method performs arbitrary point sampling in the continuous space near the vessel, and is encouraged to implicitly learn the Couinaud boundary in countless points.

4 Conclusion

We propose a multi-scale point-voxel fusion framework for accurate Couinaud segmentation that takes advantage of the topological relationship of coordinate points in 3D space, and leverages the semantic information of voxel grids. Besides, the point sampling strategy embedded with vascular prior increases the access of our method to important regions, and also improves the segmentation accuracy and robustness in uncertain boundaries. Experimental results demonstrate the effectiveness of our proposed method against other cutting-edge methods, showing its potential to be applied in the preoperative application of liver surgery.

Acknowledgement. This project was funded by the National Natural Science Foundation of China (82090052, 82090054, 82001917 and 81930053), Clinical Research Plan of Shanghai Hospital Development Center (No. 2020CR3004A), and National Key Research and Development Program of China under Grant (2021YFC2500402).

References

1. Arya, Z., Ridgway, G., Jandor, A., Aljabar, P.: Deep learning-based landmark localisation in the liver for Couinaud segmentation. In: Papiež, B.W., Yaqub, M., Jiao, J., Namburete, A.I.L., Noble, J.A. (eds.) MIUA 2021. LNCS, vol. 12722, pp. 227–237. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-80432-9_18
2. Bilic, P., et al.: The liver tumor segmentation benchmark (LiTS). Med. Image Anal. **84**, 102680 (2023)
3. Bismuth, H.: Surgical anatomy and anatomical surgery of the liver. World J. Surg. **6**, 3–9 (1982)
4. Boltcheva, D., Passat, N., Agnus, V., Jacob-Da Col, M.A., Ronse, C., Soler, L.: Automatic anatomical segmentation of the liver by separation planes. In: Medical Imaging 2006: Visualization, Image-Guided Procedures, and Display, vol. 6141, pp. 383–394. SPIE (2006)

5. Ching, T., et al.: Opportunities and obstacles for deep learning in biology and medicine. *J. R. Soc. Interface* **15**(141), 20170387 (2018)
6. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016, Part II. LNCS, vol. 9901, pp. 424–432. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46723-8_49
7. Couinaud, C.: Liver anatomy: portal (and suprahepatic) or biliary segmentation. *Dig. Surg.* **16**(6), 459–467 (1999)
8. Huang, S., Wang, B., Cheng, M., Wu, W., Huang, X., Ju, Y.: A fast method to segment the liver according to Couinaud's classification. In: Gao, X., Müller, H., Loomes, M.J., Comley, R., Luo, S. (eds.) MIMI 2007. LNCS, vol. 4987, pp. 270–276. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-79490-5_33
9. Jia, X., et al.: Boundary-aware dual attention guided liver segment segmentation model. *KSII Trans. Internet Inf. Syst. (TIIS)* **16**(1), 16–37 (2022)
10. Litjens, G., et al.: A survey on deep learning in medical image analysis. *Med. Image Anal.* **42**, 60–88 (2017)
11. Liu, X., et al.: Secular trend of cancer death and incidence in 29 cancer groups in china, 1990–2017: a joinpoint and age-period-cohort analysis. *Cancer Manage. Res.* **12**, 6221 (2020)
12. Liu, Z., Tang, H., Lin, Y., Han, S.: Point-voxel CNN for efficient 3d deep learning. In: Advances in Neural Information Processing Systems, vol. 32 (2019)
13. Nelson, R., Chezmar, J., Sugabaker, P., Murray, D., Bernardino, M.: Preoperative localization of focal liver lesions to specific liver segments: utility of CT during arterial portography. *Radiology* **176**(1), 89–94 (1990)
14. Orcutt, S.T., Anaya, D.A.: Liver resection and surgical strategies for management of primary liver cancer. *Cancer Control* **25**(1), 1073274817744621 (2018)
15. Pla-Alemany, S., Romero, J.A., Santabarbara, J.M., Aliaga, R., Maceira, A.M., Moratal, D.: Automatic multi-atlas liver segmentation and Couinaud classification from CT volumes. In: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), pp. 2826–2829. IEEE (2021)
16. Qi, C.R., Yi, L., Su, H., Guibas, L.J.: PointNet++: deep hierarchical feature learning on point sets in a metric space. In: Advances in Neural Information Processing Systems, vol. 30 (2017)
17. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015, Part III. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
18. Shen, D., Wu, G., Suk, H.I.: Deep learning in medical image analysis. *Ann. Rev. Biomed. Eng.* **19**, 221–248 (2017)
19. Soler, L., et al.: Fully automatic anatomical, pathological, and functional segmentation from CT scans for hepatic surgery. *Comput. Aided Surg.* **6**(3), 131–142 (2001)
20. Soler, L., et al.: 3D image reconstruction for comparison of algorithm database: a patient specific anatomical and medical image database. IRCAD, Strasbourg, France, Technical report, 1(1) (2010)
21. Tian, J., Liu, L., Shi, Z., Xu, F.: Automatic Couinaud segmentation from CT volumes on liver using GLC-UNet. In: Suk, H.-I., Liu, M., Yan, P., Lian, C. (eds.) MLMI 2019. LNCS, vol. 11861, pp. 274–282. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32692-0_32