Article:
Zhang, X., Ali, S. orcid.org/0000-0003-1313-3542, Liu, T. et al. (9 more authors) (2024) Robust and smooth Couinaud segmentation via anatomical structure-guided point-voxel network. Computers in Biology and Medicine, 182. 109202. ISSN: 0010-4825

https://doi.org/10.1016/j.compbiomed.2024.109202

Highlights

Robust and Smooth Couinaud Segmentation via Anatomical Structure-Guided Point-Voxel Network

Xukun Zhang,Sharib Ali,Tao Liu,Xiao Zhao,Zhiming Cui,Minghao Han,Shuwei Ma,Jingyi Zhu,Yanlan Kang,Le Wang,Xiaoying Wang,Lihua Zhang

- Introduced a novel automatic Couinaud liver segmentation framework, leveraging a dual-branch point-voxel fusion for enhanced spatial and semantic modeling.
- Implemented dense sampling with vessel structure priors, significantly enhancing segmentation accuracy, especially on thick-layer CT.
- Developed a local attention module and a novel feature-level distance loss, ensuring smooth and cohesive segmentation boundaries.
- Demonstrated superior segmentation performance and robustness across multiple public liver datasets, including out-of-distribution (OOD) validation.
- Released two public datasets for Couinaud segmentation, contributing to advancements in liver surgery planning and research.

Robust and Smooth Couinaud Segmentation via Anatomical Structure-Guided Point-Voxel Network

Xukun Zhang[a], Sharib Ali[b], Tao Liu[a], Xiao Zhao[a], Zhiming Cui[c], Minghao Han[a], Shuwei Ma[a], Jingyi Zhu[a], Yanlan Kang[a], Le Wang[a], Xiaoying Wang[d] and Lihua Zhang[a,*]

[a]*Academy for Engineering and Technology, Fudan University, Shanghai 200082, China*
[b]*The School of Computing, University of Leeds, Leeds, UK*
[c]*The School of Biomedical Engineering, ShanghaiTech University, Shanghai 200082, China*
[d]*Liver Cancer Institute, Zhongshan Hospital, Fudan University, Shanghai, 200082, China*

ARTICLE INFO

Keywords:
Couinaud Segmentation
Point-Voxel Network
Liver CT

ABSTRACT

Precise Couinaud segmentation from preoperative liver computed tomography (CT) is crucial for surgical planning and lesion examination. However, this task is challenging as it is defined based on vessel structures, and there is no intensity contrast between adjacent Couinaud segments in CT images. To solve this challenge, we design a multi-scale point-voxel fusion framework, which can more effectively model the spatial relationship of points and the semantic information of the image, producing robust and smooth Couinaud segmentations. Specifically, we first segment the liver and vessels from the CT image and generate 3D liver point clouds and voxel grids embedded with the vessel structure. Then, our method with two input-specific branches extracts complementary feature representations from points and voxels, respectively. The local attention module adaptively fuses features from the two branches at different scales to balance the contribution of different branches in learning more discriminative features. Furthermore, we propose a novel distance loss at the feature level to make the features in the segment more compact, thereby improving the certainty of segmentation between segments. Our experimental results on three public liver datasets demonstrate that our proposed method outperforms several state-of-the-art methods by large margins, including for point-based (nearly 5%-15% compared to the best method) and voxel-based methods (nearly 3%-25% compared to the best method) . Our code and manual annotations of the public datasets presented in this paper are available online: https://github.com/xukun-zhang/Couinaud-Segmentation.

1. Introduction

Primary liver cancer (PLC) is one of the most common and deadly cancer diseases in the world, and its incidence is steadily increasing (2; 3). Since the liver in the human body can regenerate after partial resection, hepatectomy is an effective method for treating PLC (4; 5). Couinaud segmentation (6) divides the liver into eight functional independent regions, namely Couinaud segments, according to the vessel supply of the liver. Each Couinaud region has its inflow and outflow of vessels and biliary system (Fig. 1 a). Based on this basis permits, segmentation of the liver only for the involved segment(s) to be ablated or resected without injuring adjacent normal parenchyma (6; 7). Couinaud segmentation based on preoperative CT intuitively displays the positional relationship between Couinaud segments and intrahepatic lesions that helps surgeons make surgical planning (8; 9).

In clinical settings, obtaining Couinaud segments using manual annotation is tedious, time-consuming, and requires significant expertise as these are based on the vasculature layout in the liver as a rough guide (10). As shown in Fig. 1, experienced medical experts usually perform Couinaud segmentation based on the context information of several adjacent slices along the z-axis direction (*i.e*, axial plane), where segments I, II, IV, VII and VIII are separated by the hepatic vein (Fig. 1(b-c)), and the other III, V, and VI are surrounded by the portal vein (Fig. 1(d-e)). The process mentioned above depends on the experience of medical experts and is highly subjective between experts. Thus, developing an automatic method to segment Couinaud segments from CT images precisely is essential for surgical planning and has attracted tremendous research in this direction.

However, robust and smooth segmentation of Couinaud segments from CT images is challenging since it is defined according to the anatomical structure of liver vessels. There are at least two challenges that often adversely affect the segmentation performance: (1) The liver parenchyma inside different Couinaud segments shows similar signal intensity due to which there is no intensity contrast between adjacent segments (Fig. 1(c)(e)), i.e., **limitation-1**, and (2) The vessels embedded in the 3D liver space are thin tubular structures, so there can be high uncertainty in the intersegmental separation planes defined by the vessels (Fig. 1(b)(d)), i.e., **limitation-2**.

Previous works (11)–(13) mainly relied on hand-designed features or atlas-based models by detecting some anatomical landmarks in the liver vessel system to derive 3D intersegmental planes. Although manual annotations are reduced, these methods often fail to robustly handle areas with limited

*Corresponding author
zhangxk21@m.fudan.edu.cn (X. Zhang); S.S.Ali@leeds.ac.uk (S. Ali); taoliu22@m.fudan.edu.cn (T. Liu); zhaox21@m.fudan.edu.cn (X. Zhao); cuizhm@shanghaitech.edu.cn (Z. Cui); mhhan22@m.fudan.edu.cn (M. Han); 23210860015@m.fudan.edu.cn (S. Ma); zhujingyi21@m.fudan.edu.cn (J. Zhu); ylkang21@m.fudan.edu.cn (Y. Kang); wangle@fudan.edu.cn (L. Wang); xiaoyingwang@fudan.edu.cn (X. Wang); lihuazhang@fudan.edu.cn (L. Zhang)
ORCID(s):

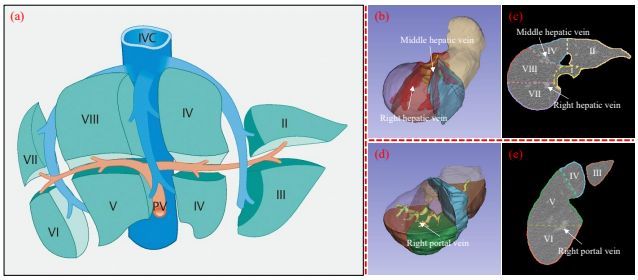

Figure 1: Couinaud segments (denoted as Roman numbers) in relation to the liver vessel structure. (a) A visual representation of the liver's division into distinct Couinaud segments by the courses of the hepatic veins (blue vascular branches) and the portal vein (orange vascular branches). This diagram is reproduced unaltered under the CC BY-NC-ND license. Source: Campos *et al.* (1). (b) and (c) originate from CT scans of the same subject, demonstrating several Couinaud segments delineated by the hepatic veins, presented in both 2D and 3D views. (d) and (e) originate from CT scans of another subject, illustrating several segments surrounded by the portal vein and segregated by the trajectories of the hepatic veins. Note that in (b), (c), (d), and (e), as well as in other images later in this paper, Couinaud segments are further identified by distinct colors.

features, such as boundaries between adjacent Couinaud segments, and have limitations in accuracy. Recently, with the advancement of deep learning, many CNN-based algorithms perform supervised training through pixel-level Couinaud annotations to automatically obtain segmentation results (10)(14; 15). Unfortunately, CNN-based methods often require downsampling operations to expand the receptive field and reduce memory consumption, inevitably losing spatial information and making it difficult to effectively capture the useful key anatomical regions for Couinaud segmentation. In addition, all of these methods directly learn the intensity of 3D voxels of the liver, which makes it easy to produce inaccurate segmentation results in the absence of intensity contrast (due to **limitation-1**). Improving the learning of spatial relationships between different segments is the key to solving this limitation, which can supplement CNN-based methods and enhance the segmentation performance in regions without intensity contrast. However, due to the high uncertainty (due to **limitation-2**) in adjacent intersections based on the Couinaud definition due to vessel course, it is challenging to produce a smooth intersegmental plane, further hindering clinical interpretation and understanding, hindering surgical procedures.

In this paper, to tackle the aforementioned challenges, we propose a point-voxel fusion framework that represents the liver in continuous points that enables us to learn the spatial structure better while performing the convolutions in voxels to obtain the complementary semantic information of the Couinaud segments. 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 comprises a series of convolutions to learn semantic features, followed by de-voxelization to transfer them back to points. The voxel-based branch gradually expands the receptive field of convolution operations by performing voxelization and de-voxelization operations at different resolutions. In addition, to balance the contribution of the two branches to learning the Couinaud segmentation, we design a local attention module to adaptively fuse the features at each scale. Finally, the multi-scale fusion features are fed to the decoder to predict the Couinaud category at any sampling point in the 3D liver space. To further enhance the discriminability of the model in the uncertain regions between segments, we design a novel distance loss at the feature level to make the features within the segment more compact, thus producing a smooth segmented interface. Extensive experiments on three publicly available datasets named 3Dircadb (16), LiTS (17) and MSD8 (14; 18) demonstrate that our proposed framework achieves state-of-the-art (SOTA) performance, outperforming cutting-edge methods quantitatively and qualitatively.

This work is a comprehensive extension of a preliminary conference paper (19). Compared with the preliminary version, the major extensions are five-fold. 1) We provide a more comprehensive review of the work related to Couinaud segmentation. 2) We introduce a semi-supervised segmentation network to generate more robust vessel attention maps for liver CT with missing or low-quality labels. 3) We design a local attention module that adaptively fuses features to balance the contributions of the two branches and learns more discriminative feature representations. 4) We design a novel feature distance loss, further enhancing the discrimination ability of the pixel-voxel model in the uncertainty region between segments and producing smooth segmentation results. 5) We perform more comprehensive experiments on three public datasets to demonstrate the superiority of our proposed method over several SOTA methods in segmenting Couinaud segments on the intra-distribution and out-of-distribution (OOD) datasets.

The rest of the paper is organized as follows. The most related works, including Couinaud segmentation and point-voxel fusion learning, are more comprehensively reviewed in Section 2. The studied data and our full framework are described in Section 3. Section 4 presents experimental results and comparisons of our methods with other SOTA methods. Section 5 discusses the hyperparameters and other settings in our method in detail. Finally, the paper is concluded in Section 6.

2. Related Work

In this section, we review existing methods in the literature that are closely related to our study, including Couinaud segmentation and point-voxel fusion learning.

2.1. Couinaud Segmentation

Various methods have been proposed for Couinaud segmentation, which can be roughly grouped as 1) traditional manual or semi-automatic methods and 2) voxel-based deep learning methods, as briefed below.

2.1.1. Traditional Methods

Traditional methods for Couinaud segmentation tasks typically rely on manual or semi-automatic image processing, including level set, region growing, and atlas-based methods. Selle *et al.* (20) presented a technique for semi-automatic contour selection for Couinaud segmentation, utilizing a region-growing approach for the vessels and graph analysis to identify the eight segments. Huang *et al.* (12) proposed a complete framework, 3D thinning, vascular tree extraction and pruning, vascular classification and projection, for identifying Couinaud categories in 3D liver space. Oliveira *et al.* (21) combined level set and region growing methods to deal with 3D liver and vessels, thus constructing four planes for Couinaud segmentation. Pla-Alemany *et al.* (13) proposed a Couinaud classification method based on multi-atlas registration and implemented it on the 3Dircadb database (16). However, the methods mentioned above usually rely on the experience of clinicians, and the complete process is time-consuming and labour-intensive. In contrast, deep learning-based methods can perform end-to-end processing and are widely used in 3D medical image segmentation tasks, which can produce performance close to clinical annotations.

2.1.2. Voxel-based Deep Learning Methods

The Couinaud segmentation methods are conventionally based on CNN models to generate or derive the corresponding three-dimensional Couinaud masks from liver CT scans. Specifically, Arya *et al.* (10) constructed a landmark localization model using deep learning-based heatmap regression to predict landmarks in the liver. They also derived the Couinaud segments from the predicted landmarks and compared the performance with that obtained by the network-based direct segmentation model, showing close accuracy. Similarly, Wang *et al.* (22) proposed an attentive residual hourglass-based cascaded network to identify six key bifurcation points of the liver vascular system. Subsequently, the plane dividing the liver into different Couinaud segments was derived using the detection points. The method of deriving intersegmental planes directly from landmarks does not consider the subtle course of critical vessels, potentially impacting surgical pathway planning. Furthermore, Tian *et al.* (14) and Jia *et al.* (15) explored direct pixel-level Couinaud segmentation from CT/MR images using CNNs. Tian *et al.* proposed the GLC-UNet (14), which processes pre-detected key vascular slices and the target slice for segmentation, leveraging attention mechanisms to learn inter-slice relationships, thereby achieving Couinaud segmentation across 193 cases of thick-layer CT scans. Recent work by Jia *et al.* (15) focused on employing a dual attention module that enhances critical information between adjacent segments and experiments with 59 cases of clinically acquired 3D MRI data. However, the methods proposed by Tian *et al.* (14) and Jia *et al.* (15), still based on 2D U-Net architectures, do not directly capture spatial relationships within sequences. Moreover, the CNN model directly for voxel intensity makes it easy to treat all voxel-wise features equally in training, and the commonly used downsampling operation will lose important spatial information, which still poses a challenge to accurate Couinaud segmentation.

2.2. Point-Voxel Fusion Learning

The point-voxel fusion method aims to learn information about the object at different levels or from various perspectives, thereby improving the performance of the task. Zhou *et al.* (23) aggregated point data into the 3D voxel grid and then combined point-wise features with locally aggregated features through the voxel feature coding layer, aiming at accurate 3D detection. Liu *et al.* (24) proposed the PVCNN model, which converts the input 3D data into a high-resolution point cloud and low-resolution voxel grid. PVCNN consists of two different branches, which are used to extract fine-grained features of points and local features of voxel grids, respectively. The two features are fused, and the decoder outputs the segmentation or classification results. Similarly, Meng *et al.* (25) proposed a new point cloud segmentation method, which transforms unstructured point clouds into regular voxel grids and further uses a kernel-based interpolated variational autoencoder (26) architecture to encode the local geometry within each voxel. In addition, many works (27; 28) have explored the deep integration of voxel-based and point-based networks and achieved advanced performance in 3D object detection and semantic segmentation tasks in natural scenes. Unlike those methods, our proposed framework is carefully designed and targets 3D medical imaging problems, aiming to precisely parse Couinaud segments using spatial priors and standing robustness against corner cases.

3. Materials and Method

Our framework for robust and smooth segmentation of Couinaud fragments from 3D CT volume is shown in Fig. 2, including liver segmentation, vessel attention map generation, point data sampling and multi-scale point-voxel fusion network.

3.1. Data and Pre-Processing

This studied data from three publicly available datasets, 3Dircadb (16), LiTS (17) amd MSD8 (18). The 3Dircadb dataset (16) contains 20 CT volumes with an average thickness of 1.78 *mm*, and an average image size is 512×512×141 voxels with liver and vessel segmentation labels. The LiTS dataset (17) consists of 200 CT volumes, with an average thickness of 1.51 *mm* and an average image size of 512 × 512×448 voxels, and has liver and tumour labels. The MSD8 dataset (18) includes 443 cases, the average thickness is 4.6 *mm*, and the average image size is 512 × 512 × 60 voxels, with low-quality vessel labels. In our study, we annotated all

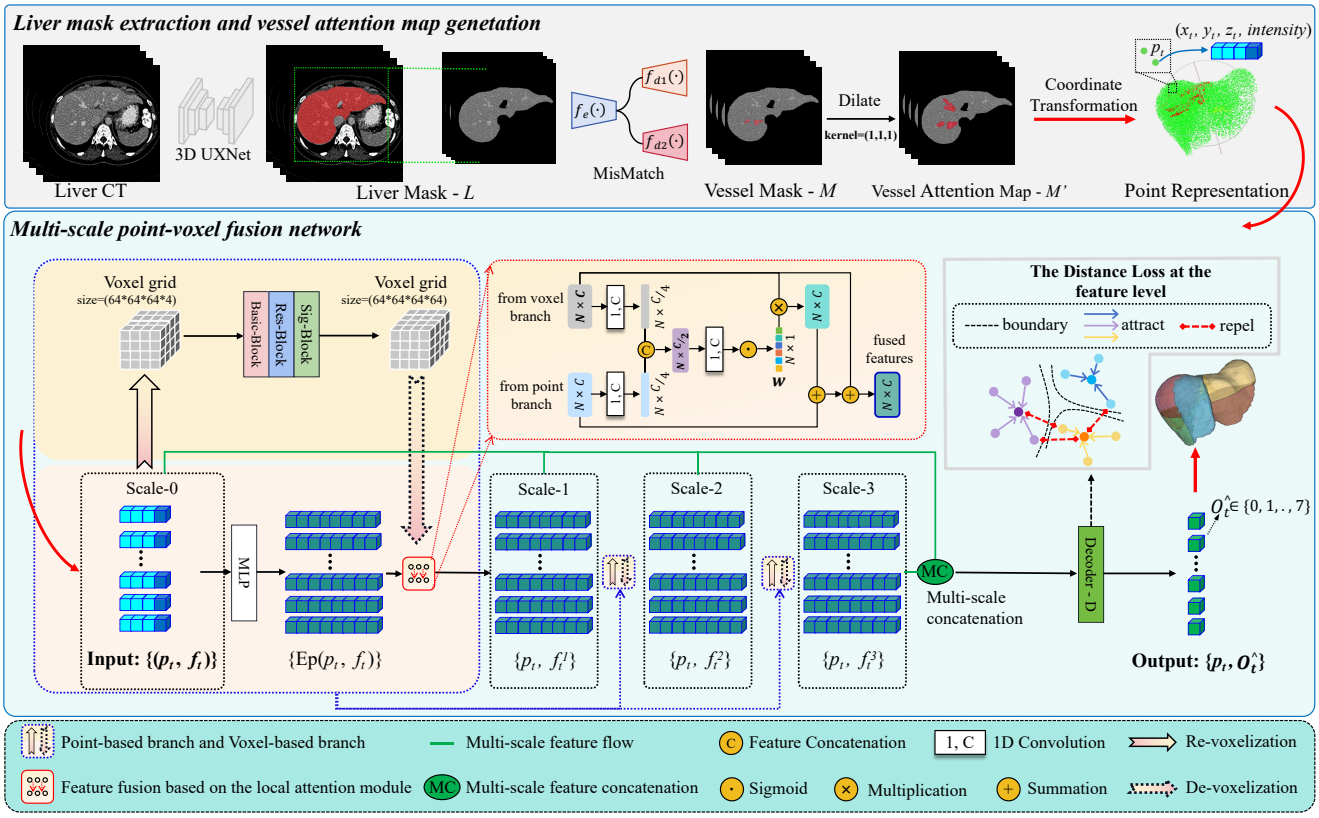

Figure 2: Overall framework of our proposed method for Couinaud segmentation. The 3D UXNet is used to obtain the liver mask, the semi-supervised MixMatch is used to extract the vessel structure in the image, and then the multi-scale point-voxel fusion network performs Couinaud classification on the sampling points in the liver space to reconstruct the complete 3D Couinaud mask.

20 subjects of the 3Dircadb dataset (16) and 131 subjects of the LiTS dataset (17) with Couinaud segmentation. For the MSD8 dataset (18), we used 193 CT volumes annotated by Tian *et al.* (14) to carry out experiments. Note that the 3Dircadb (16) and LiTS datasets (17) have an overlap and are derived from the same institutions, but the MSD8 dataset (18) is derived from different institutions. Our study uses the same window width and window level setting for all 3D CT volumes, maintaining the image intensity between -100 and 300 Hounsfield units (HU). In addition, for each stage of our proposed method, the input images were normalized using the $z-$score method.

3.2. Liver Mask and Vessel Attention Map Generation

Liver segmentation is a fundamental step in the Couinaud segmentation task. Considering that the liver is large and easily identified in the abdominal organs, we extracted the liver mask through a trained 3D UXNet (33). Different from liver segmentation, delicate vessel structure extraction is challenging, so we aim to extract rough vessel regions, namely vessel attention maps, as a guide for Couinaud segmentation. Specifically, given a 3D CT image containing only the area covered by the liver mask (L), the semi-supervised method (MixMatch) (30) can generate the corroded and dilated vessel attention regions from the image, respectively. Then, we combine the two regional maps and binarize them with a threshold of 0.5 to generate the vessel mask (M). Finally, a morphological dilation encloses more vessel pixels in the covered area represented by M to create the vessel attention map (M'). Based on this, we can extract reliable vessel structures without additional training from both the LiTS (17) and MSD8 (18) datasets with missing or low-quality labels.

3.3. Couinaud Segmentation

Based on the above work, we can sample the point data with the vessel from the 3D CT volume, which can be transformed into voxel grids by re-voxelization (as shown in Fig. 2). Inspired by (24), a novel multi-scale point-voxel fusion network is proposed to simultaneously process point and voxel data through a point-based branch and voxel-based branch, aiming to perform Couinaud segmentation accurately. The details of this part of our method are described below.

3.3.1. Point Sampling Embedded with Vessel Stucture

To learn the topological relationship between Couinaud segments, a direct strategy is to obtain the coordinate point data with 3D spatial information from liver CT and perform point-wise classification. To this end, we first unified convert the image coordinate points $I = \{i_1, i_2, ..., i_t, i_t \in \mathbb{R}^3\}$ in liver CT into the world coordinate points $P = \{p_1, p_2, ..., p_t, p_t \in \mathbb{R}^3\}$:

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

where $Spacing$, $Direction$, and $Origin$ represent the voxel spacing, scanning direction, and the world coordinates of the origin of the CT image, respectively. 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 ignores the vessel structure, which is crucial for Couinaud segmentation. We propose a dense sampling strategy embedded with vessel prior based on the M' to solve this issue. Specifically, T points are randomly sampled as input in each training epoch, where $T/2$ points fall in a smaller space covered by M', which allows the point-voxel network to increase access to essential 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, the label of the point $p_t = (x_t + \Delta x, y_t + \Delta y, z_t + \Delta z)$ is generated by:

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

where R denotes the rounding integer function, and O_t indicates that the point p_t belongs to the class of Couinaud segmentation. Note that in our experiment, $\{0, 1, ..., 7\}$ is used to indicate {I, II, ..., VIII}. Based on this, we achieve arbitrary resolution sampling in the continuous space covered by the M' and use it to input multi-scale point-voxel fusion networks.

3.3.2. Re-voxelization

It is insufficient to extract the topological 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):

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

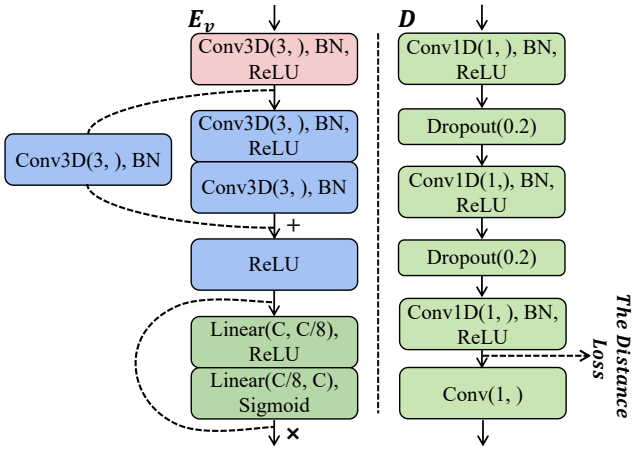

Figure 3: Detailed structure of voxel branch (E_v) and decoder (D).

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 network 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 proposed point sampling strategy, the converted voxel grid inherits the vessel structure from the point data. It dilutes the unimportant information in CT images and extracts effective semantic features for Couinaud segmentation through the voxel branch.

3.3.3. Multi-scale Point-voxel Fusion Network

Based on the above-mentioned basis, we propose a multi-scale point-voxel fusion network for accurate Couinaud segmentation, which exploits the topological relationship of coordinate points in 3D space and leverages the semantic information of voxel grids. As shown in Fig. 2, our method has two branches: 1) point-based and 2) voxel-based. 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. The structure of E_v is shown in Fig. 3, where the residual convolution stabilises the features and the sigmoid layer balances the weights between the feature channels. Subsequently, the features extracted from E_v are re-transform into point representations through trilinear interpolation (tri) and combined with the fine-grained features extracted from the point-based branch, which provide complementary information:

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

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, and $LAM(\cdot)$ is a local attention module for feature fusion, as shown in Fig. 2, which can adaptively balance the contributions of different branches and generate more discriminative features. 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 halved compared to the previous round. After three rounds of point-voxel operations, the original point feature f_t and the features $\{f_t^1, f_t^2, f_t^3\}$ from multiple voxel scales are concatenated and then sent to a point-wise decoder D (Fig. 3), parameterized by a fully connected network, to predict the corresponding Couinaud segment category:

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

where $\{0, 1, ..., 7\}$ denotes the Couinaud segment category predicted by our model for the point p_t. We employ cross-entropy (CE) and Dice losses to supervise the learning process. Furthermore, we design a novel distance loss at the feature level, aiming to enhance the discriminative ability of the network in the uncertain regions between Couinaud segments. We describe the local attention module (LAM) and the distance loss in detail below.

Local attention module for adaptively feature fusion. As shown in Fig. 2, given the features from the two branches, the channel is first reduced to 1/4 of the original by 1D convolution, then concatenated based on the channel. Followed by convolution and activation operations, the feature weight vector $W \in \mathbb{R}^{N\times 1}$ of all points is obtained. The features from the voxel branch are fused based on the weight W with the point features and then added with the features from the voxel branch to obtain new point features. Note that the model stores the location of each point, so the features represented by the points from the two branches can be effectively fused by locally weighted addition, and the fused features will be used for the next round of point-voxel operation.

The distance loss enhances model discrimination. Inspired by (31; 32), we design a distance loss that includes 1) R_{intra}, which is the centre loss (31), used to narrow the internal feature distance of the Couinaud segment; 2) R_{inter}, which is designed to enhance further the discrimination ability of the pixel-voxel model to identify uncertain regions between segments. R_{intra} is described as the following equation,

$$R_{\text{intra}} = \frac{1}{N_{all}} \sum_{i}^{8} \sum_{j}^{N_i} \left\| x_{i,j} - c_i \right\|_2^2, \quad (6)$$

where $c = \{c_0, c_1, ..., c_7\} \in \mathbb{R}^{8\times 128}$ is the learnable features of eight Couinaud segments centre, $x_{i,j} \in \mathbb{R}^{1\times 128}$ is the feature vector of the $j-$th point in segment$-i$ (note that we use the features connected to the classifier in the experiment, and the channel dimension is 128), and N_{all} is the total number of sampling points for all segments in the training. According to Equation (6), in the model's training process, the point features $x_{i,j}$ in each Couinaud segment will aggregate toward c_i. However, the points between Couinaud segments are highly uncertain, and the distance between classes may be less than the distance within the class, so R_{intra} cannot effectively enhance the model's ability to classify these hard points. To this end, we further design R_{inter}:

$$R_{\text{inter}} = CE((x' * c^T), O). \quad (7)$$

The objective of our R_{inter} can be understood as minimizing the cross entropy between the features (x') of points in uncertain regions between segments based on centroid aggregation and the corresponding label (O) of the segment. Specifically, in the network's learning process, the points in the uncertain region between segments are often difficult to classify, which has a greater loss value. Therefore, we rank the CE loss of all points in a batch, adaptively select the $top10\%$ as the points of the uncertain region between segments, and then obtain the features (x') of the corresponding points. The feature (x') is multiplied by the transposition centroid feature (c^T) of different segments to recalculate loss and update the model parameters. On this basis, the hard points between segments will be further attracted by the centroids of the same class and repelled by the centroids of other classes, as shown in Fig. 2, thus promoting the production of smooth segmentation planes. The complete loss of our point-voxel fusion network is:

$$\mathcal{L} = \mathcal{L}_{CE} + \mathcal{L}_{Dice} + \lambda(\alpha \cdot R_{\text{intra}} + (1-\alpha) \cdot R_{\text{inter}}), \quad (8)$$

where λ is the weight to control the distance loss, and α is the weight to adjust the internal term.

3.4. Implementation Details

3.4.1. Network Details

As shown in Fig. 2, our point-voxel network consists of a voxel-based branch, a point-based branch, the feature fusion module, and the decoder. Three point-voxel operations output fusion features of 64 channels, 128 channels, and 256 channels, respectively. The resolution of the re-voxelized grid in the first point-voxel processing is 64×64×64. All MLPs in our point-voxel model are implemented by 1D convolution and followed by the BatchNorm and ReLU layers, except for the final classification header.

3.4.2. Traning Details

The proposed framework was implemented on an RTX A6000 GPU using PyTorch. For the point-voxel network, we sample $T = 50,000$ points in each epoch and perform scaling within the range of 0.9 to 1.1, arbitrary axis flipping, and rotation of 0 to 5 degrees as the augmentation strategy. Note that $T/2 = 25,000$ points in each input epoch are from the area covered by the vessel attention map. Besides, we use the stochastic gradient descent optimizer with a learning rate of 0.01. All our experiments were trained for 600 epochs, and then we used the model with the best performance on the validation 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 performance of all segments in all testing subjects.

Method	3Dircadb (16)			LiTS (17)			MSD8 (18)		
	Acc (% ↑)	Dice (% ↑)	ASD (mm ↓)	Acc (% ↑)	Dice (% ↑)	ASD (mm ↓)	Acc (% ↑)	Dice (% ↑)	ASD (mm ↓)
3D UNet (29)	$59.29_{\pm 18.}$	$72.97_{\pm 15.}$	$8.64_{\pm 4.5}$	$68.73_{\pm 10.}$	$80.95_{\pm 8.4}$	$6.50_{\pm 3.6}$	$77.04_{\pm 7.3}$	$86.99_{\pm 4.9}$	$3.75_{\pm 1.8}$
3D UXNet (33)	$37.41_{\pm 15.}$	$52.97_{\pm 16.}$	$28.78_{\pm 3.2}$	$61.71_{\pm 11.}$	$77.89_{\pm 9.2}$	$7.44_{\pm 3.8}$	$77.74_{\pm 9.0}$	$87.17_{\pm 6.1}$	$4.18_{\pm 2.2}$
Swin UNetR (34)	$29.35_{\pm 15.}$	$43.68_{\pm 17.}$	$21.49_{\pm 9.7}$	$60.53_{\pm 13.}$	$76.74_{\pm 11.}$	$6.86_{\pm 3.7}$	$76.92_{\pm 9.6}$	$86.60_{\pm 6.7}$	$3.98_{\pm 1.9}$
PointNet2Plus (35)	$69.48_{\pm 13.}$	$81.36_{\pm 9.7}$	$5.66_{\pm 3.7}$	$69.79_{\pm 8.2}$	$84.15_{\pm 5.8}$	$5.39_{\pm 2.6}$	$71.93_{\pm 11.}$	$83.14_{\pm 8.2}$	$5.30_{\pm 2.8}$
Jia *et al.*'s (15)	$19.02_{\pm 6.1}$	$24.19_{\pm 9.0}$	—	$48.83_{\pm 11.}$	$64.86_{\pm 11.}$	—	$73.47_{\pm 12.}$	$84.09_{\pm 9.1}$	$4.71_{\pm 2.4}$
Ours	$\mathbf{83.70}_{\pm 4.2}$	$\mathbf{91.05}_{\pm 2.7}$	$\mathbf{3.19}_{\pm 1.1}$	$\mathbf{84.44}_{\pm 5.9}$	$\mathbf{91.45}_{\pm 3.6}$	$\mathbf{2.95}_{\pm 1.3}$	$\mathbf{81.41}_{\pm 9.1}$	$\mathbf{89.46}_{\pm 6.0}$	$\mathbf{3.44}_{\pm 2.5}$

— indicates that there is an unrecognized Couinaud segment category, resulting in ASD cannot be calculated.

4. Experimental Results

4.1. Experimental Setup

In the experiments, we randomly divided each dataset into the training, validation, and test sets in a 10:3:7 ratio. For vessel region extraction, we use all training samples (168 cases) from three training sets for model training in a semi-supervised manner and save the trained model with the least validation loss (52 cases). We can obtain the vessel attention maps on three test sets based on this trained vessel model. Note that although we also obtained the liver segmentation mask through a trained model, we used the liver label for point sampling to train the point-voxel network, so the classifier's output category is 8 Couinaud segments.

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

4.2. Comparison With Competing Methods

We compare our framework with several SOTA approaches, including 3D UNet (29), 3D UXNet (33), SwinUNetR (34), the methods of Jia *et al.* (15), and point-based PointNet2Plus (35). The method of Jia *et al.* uses 2D UNet (36) with dual attention to focus on the object boundary and is specifically for the Couinaud segmentation. We use PyTorch to implement the above methods and maintain the same learning rate and total training rounds as our point-voxel network. However, since the CNN methods require the input as a voxel grid of cubes, we resize the original CT image to 256×256×64 and make the outputs 9 categories (including background classes). In addition, we introduce random rotation, translation, and crop as data augmentation to train these CNN models. PointNet2Plus (35) does not have the above limitations, so it maintains the same experimental settings as our point-voxel model, except that the input points are all randomly sampled from the liver region. Due to the input differences between voxel-based and point-based methods, we evaluate the Couinaud segmentation performance of the liver region in this study.

4.2.1. Testing in-distribution Dataset

Table 1 summarizes the comparison results of three metrics on three datasets. Compared with the CNN methods, we can see that the 3D UNet(29), 3D UXNet (33), and SwinUNetR (34) are superior to Jia *et al.*'s (15) 2D method on three datasets, which shows that learning spatial relationships is the key to dealing with the Couinaud segmentation task. In addition, the performances of the CNN methods on the MSD8 (18) dataset are better than on the 3Dircadb (16) and LiTS (17) datasets in three metrics. The reason is that the CT images from the 3Dircadb (16) and LiTS (17) datasets have a lot of slices outside the liver, which interferes with the training of the model. However, these CNN models will also learn effective information about the Couinaud segment from more training samples, so the model's performance on LiTS (17) also outperforms that of the 3Dircadb dataset (16). By comparing the point-based methods, we can see that PointNet2Plus (35) has achieved good and close performance on all three datasets, which shows the potential of point-based methods in Couinaud segmentation tasks while further highlighting the importance of spatial relationships. Finally, our proposed point-voxel fusion framework achieves the best performance. Especially on the 3Dircadb dataset (16) with only 10 training subjects, the ACC and Dice achieved by our method exceeds other methods by a large margin, and the ASD is also greatly reduced, which demonstrates the effectiveness of our method.

To further evaluate the effectiveness of our method, we also provide qualitative results on three different datasets, as shown in Fig. 4. For the first two rows, we can see that our method can learn the structural guidance of the right hepatic vein and the middle hepatic vein, resulting in accurate segmentation between Couinaud segments. For the middle two rows, our method also captured the key structures from the middle hepatic vein and the left hepatic vein and accurately segmented the Couinaud segments IV and II. However, CNN methods fail to accurately capture this key structural relationship, generating inaccurate boundary segmentation. Note that the last two rows show that the portal vein branches divide the liver roughly into different Couinaud segments. It can be seen that our method can also learn the boundary guidance provided by the portal vein to deal with the uncertain boundary robustly. Besides, compared with the 3D view, it is obvious that the CNN methods are easy to pay attention to the local region and produce a large area of error segmentation, so the reconstructed surface is uneven (e.g., the middle two rows). Although SwinUNetR (34) with transformer can learn the region dependence in 3D space, it is difficult to utilise effective spatial information in this task and prone to error segmentation. In addition,

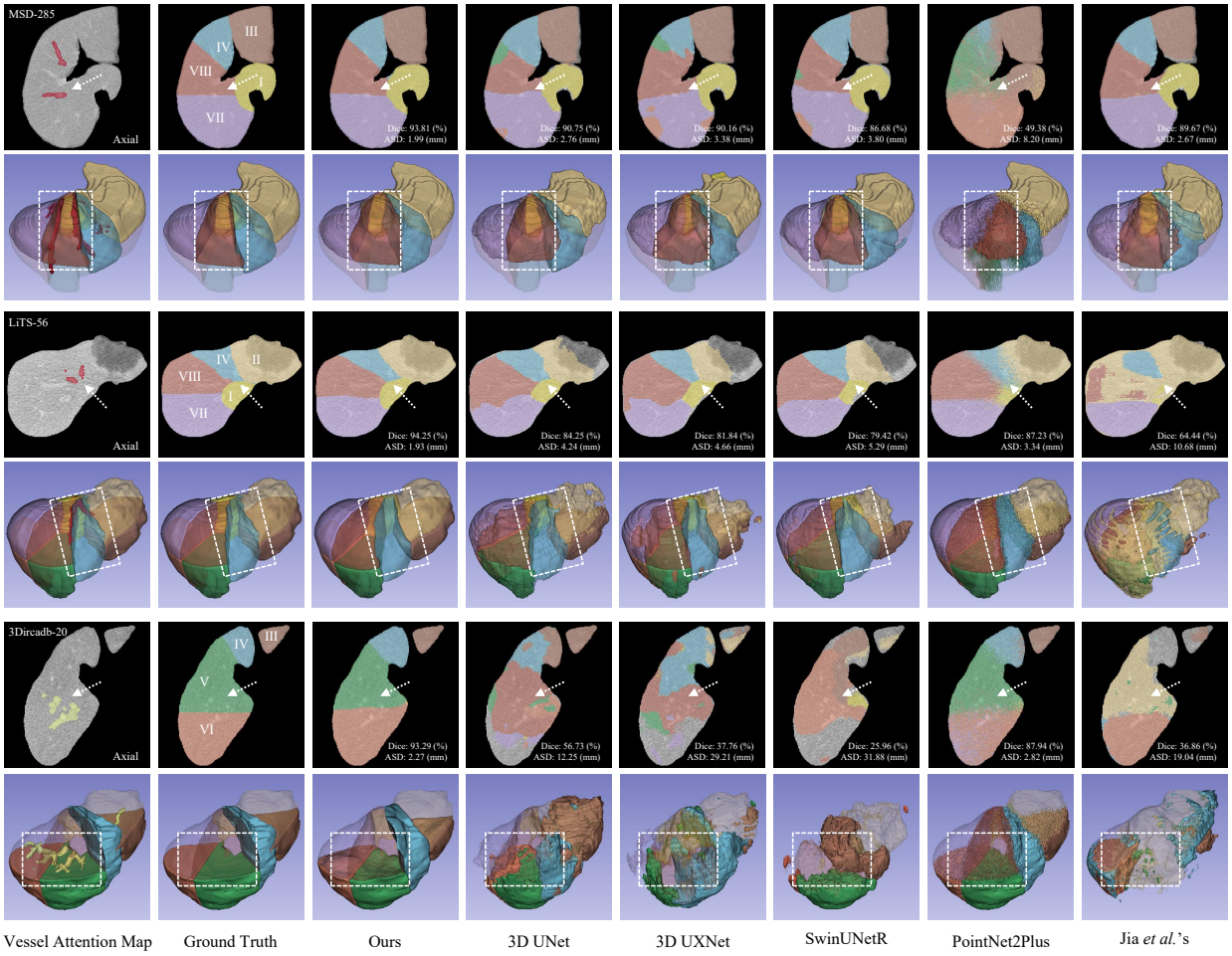

Figure 4: Comparison of segmentation results of different methods. Different colours represent different Couinaud segments. In the first column, we show the key vessels in 2D and 3D views as an additional reference for the segmentation results. Note that although the entire CT image was used as input for the CNN methods, only the liver region is displayed for visualization (Best viewed electronically, zoom in).

PointNet2Plus (35) directly learns the spatial location and performs classification on each point in the liver, obtaining the higher Dice and smooth 3D visualization. However, it is noticeable that the PointNet2Plus (35) is more likely to cause segmentation blur in boundary regions with high uncertainty. In contrast, our method combines the advantages of point-based and voxel-based methods and remedies their respective defects, yielding smooth and accurate Couinaud segmentation in three datasets.

4.2.2. Testing out-of-distribution Dataset

To further discuss the advantages of the proposed methods, we evaluate the generalization of all methods. However, due to the differences in input, the comparison methods may have two weaknesses compared to our complete framework: a) Since the input of the CNN methods is the entire CT image, the model's training is susceptible to interference from other organs and tissues. b) Compared with our method, other methods do not directly utilize the extracted vessel region as structural guidance. To this end, before evaluating generalization, we enhanced all of the methods in the following ways: **a)** The input CT is cropped to obtain the liver region and resized, and then the pixel intensity outside the liver is set to 0 to eliminate irrelevant interference. **b)** Combine the vessel attention map with the cropped CT image as the model input to inject the anatomical prior. Note that for the point-based method, we add a channel to the input point feature, with a value of 0 or 1, to label whether the point is

Table 2
Quantitative comparison with different enhanced methods on the 3Dircadb dataset.

Method	3Dircadb (16)		
	Acc (% ↑)	Dice (% ↑)	ASD (mm ↓)
3D UNet [a][b]	$80.59_{\pm 5.2}$ $\underline{82.34}_{\pm 4.2}$	$89.17_{\pm 3.1}$ $\underline{90.27}_{\pm 2.5}$	$3.57_{\pm 1.2}$ $\underline{3.22}_{\pm 0.8}$
3D UXNet [a][b]	$73.34_{\pm 4.0}$ $\underline{81.17}_{\pm 4.5}$	$84.57_{\pm 2.7}$ $\underline{89.55}_{\pm 2.7}$	$14.56_{\pm 4.3}$ $\underline{5.03}_{\pm 1.0}$
SwinUNetR [a][b]	$77.94_{\pm 4.5}$ $\underline{81.35}_{\pm 4.1}$	$87.54_{\pm 2.9}$ $\underline{89.67}_{\pm 2.5}$	$9.73_{\pm 1.5}$ $\underline{3.68}_{\pm 1.1}$
PointNet2Plus [b]	$69.48_{\pm 13.}$ $\underline{65.61}_{\pm 13.}$	$81.36_{\pm 9.7}$ $\underline{78.55}_{\pm 1.0}$	$5.66_{\pm 3.7}$ $\underline{5.97}_{\pm 3.0}$
Jia *et al.*'s [a][b]	$71.89_{\pm 11.}$ $\underline{76.90}_{\pm 7.5}$	$83.22_{\pm 7.8}$ $\underline{86.77}_{\pm 4.9}$	$4.97_{\pm 2.1}$ $\underline{4.36}_{\pm 1.9}$
Ours	$\mathbf{83.70}_{\pm 4.2}$	$\mathbf{91.05}_{\pm 2.7}$	$\mathbf{3.19}_{\pm 1.1}$

[a] [b] The model is enhanced by the proposed enhancement strategy.

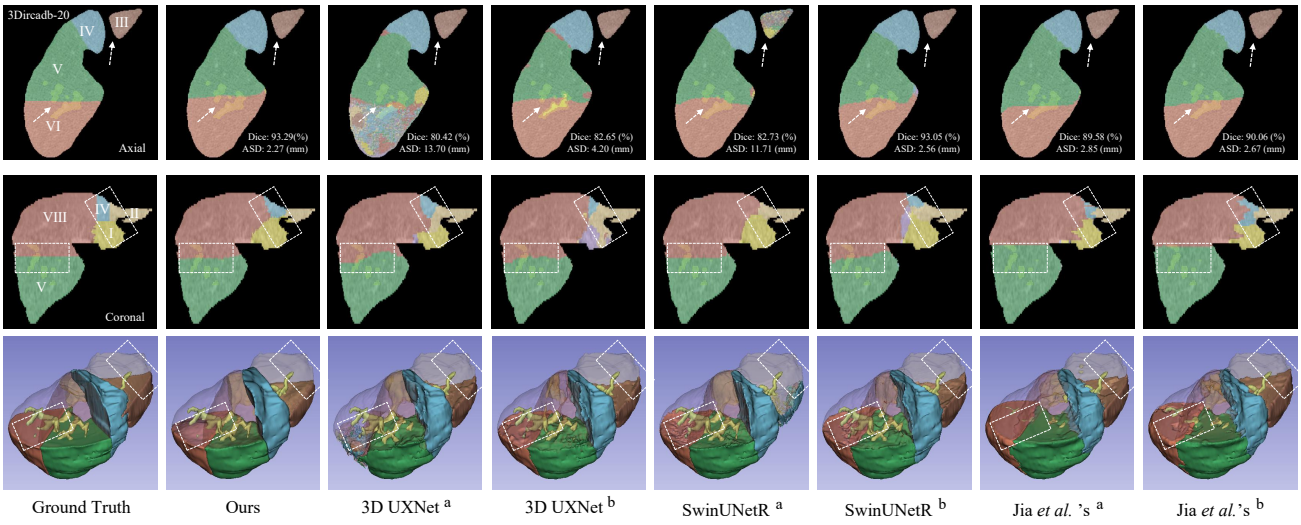

Figure 5: The visual comparison between our method and the enhanced CNN methods ([a] [b] represents the enhanced CNN model).

Table 3
Quantitative comparison with different enhanced methods on the LiTS dataset.

Method	LiTS (17)					
	Acc (% ↑)		Dice (% ↑)		ASD (mm ↓)	
3D UNet [ab]	76.94±6.4	77.47±5.7	89.07±4.0	89.44±3.5	3.95±1.9	3.89±1.8
3D UXNet [ab]	79.20±6.6	80.03±6.3	88.24±4.3	88.71±4.0	4.40±2.1	4.21±1.7
SwinUNetR [ab]	82.28±4.7	82.00±4.8	90.21±2.8	90.04±2.9	3.10±1.4	3.10±1.0
PointNet2Plus [b]	69.79±8.2	73.29±7.8	84.15±5.8	84.35±5.3	5.39±2.6	5.20±2.7
Jia *et al.*'s [ab]	62.55±15.	62.30±17.	78.07±12.	77.60±14.	8.37±4.3	8.72±5.4
Ours	**84.44**±5.9		**91.45**±3.6		**2.95**±1.3	

[a] [b] The model is enhanced by the proposed enhancement strategy.

Table 4
Quantitative comparison with different enhanced methods on the MSD dataset.

Method	MSD (18)					
	Acc (% ↑)		Dice (% ↑)		ASD (mm ↓)	
3D UNet [ab]	78.29±9.2	80.74±8.0	88.37±6.1	89.11±5.1	3.49±3.2	3.22±2.7
3D UXNet [ab]	78.01±10.	80.66±8.1	88.45±6.8	89.07±6.3	3.58±3.2	3.49±3.0
SwinUNetR [ab]	79.55±10.	80.79±8.0	88.14±6.2	88.56±5.2	3.70±3.1	3.50±2.6
PointNet2Plus [b]	71.93±11.	72.81±10.	83.14±8.2	83.82±7.5	5.30±2.8	5.19±2.6
Jia *et al.*'s [ab]	74.84±11.	74.11±12.	85.27±8.1	84.68±9.1	4.13±2.8	4.37±3.2
Ours	**81.41**±9.1		**89.46**±6.0		**3.44**±2.5	

[a] [b] The model is enhanced by the proposed enhancement strategy.

from the vessel region. The enhanced methods are retrained and tested on three datasets. The results are shown in Table 2, 3 and 4. Compared with the results shown in Table 1, the enhanced CNN methods showed significant performance improvements across all three datasets, especially notable in the 3Dircadb dataset (16) using Jia *et al.*'s 2D UNet-based method (14) (e.g., 24.19 vs. 83.22 vs. 86.77 in Dice score). However, the gains from utilizing the vessel as an additional input for enhancement-(b) were not as significant. For example, the performance of the SwinUNetR (34) on the LiTS dataset (17) decreased, indicating that while coarse vessel images can provide guidance, they can also introduce additional interference, negatively impacting segmentation performance. This observation applies to point-based methods as well. In contrast, our complete method uses the dense sampling strategy embedded with the vessel prior, enabling the resistance of noise interference and learning the guidance from the key vessel structure. The visualization results of some comparison methods are shown in Fig. 5. We can see that the enhancement strategy helps the CNN models to improve the segmentation of intersegmental regions, but our method still shows great advantages, especially in generating accurate and smooth segmentation results between the upper and lower segments of the liver.

Following the aforementioned experiments, we externally tested the models trained on the 3Dircadb dataset (16) on the LiTS (17) and MSD8 (18) datasets, thereby evaluating the methods' generalization capabilities on OOD datasets. The results of these tests are presented in Table 5. Table 5 demonstrates that, despite enhancements, the performance of models based on CNN significantly deteriorated on the LiTS (17) and MSD8 (18) datasets. This degradation could be attributed to the models' excessive reliance on specific features prevalent within the training dataset's voxel distribution, adversely affecting their generalization capabilities on OOD datasets. Thus, an interesting phenomenon is observed, wherein the PointNet2Plus (35) model, which takes pixel spatial coordinates as input, surpassed the performance of the CNN-based methods. Similarly, by incorporating the complementary spatial information provided by point data, our proposed method consistently outperformed, showcasing superior generalization ability on OOD datasets.

Further, Fig. 6 displays the performance of various methods in segmenting each Couinaud segment, as measured by Dice and ASD. The external testing results on the LiTS (17) dataset reveal that our proposed method surpasses other methods in the segmentation of every Couinaud segment, with notable superiority in segments I, IV, VI, and VII. By contrast, while our approach demonstrated the most robust generalization capability on the MSD8 (18) dataset

Table 5
Evaluate generalization with different enhanced methods on out-of-distribution datasets.

Method	3Dircadb → $LiTS$			3Dircadb → $MSD8$		
	Acc (% ↑)	Dice (% ↑)	ASD (mm ↓)	Acc (% ↑)	Dice (% ↑)	ASD (mm ↓)
3D UNet [ab]	54.37±21. 54.73±22.	67.99±18. 68.16±19.	17.91±13. 18.63±14.	58.36±8.3 57.89±9.6	73.34±7.1 72.82±8.6	11.50±4.2 11.04±4.8
3D UXNet [ab]	50.78±17. 54.26±21.	65.77±15. 67.98±18.	22.97±8.6 17.80±12.	52.76±7.9 65.16±9.4	68.71±7.4 78.48±7.8	15.82±5.8 10.13±3.7
SwinUNetR [ab]	51.76±21. 55.67±22.	65.76±18. 69.35±17.	20.22±8.6 17.04±12.	56.81±10. 62.30±9.4	71.91±9.0 76.33±7.8	17.94±4.6 10.27±4.0
PointNet2Plus [b]	67.75±14. 60.38±14.	79.88±11. 74.19±13.	7.29±6.1 9.22±7.4	69.07±10. 64.89±12.	81.25±7.8 78.06±9.4	**5.74**±3.2 6.60±3.1
Jia *et al.*'s [ab]	47.66±26. 51.11±25.	60.43±24. 64.20±22.	18.29±13. 16.06±11.	37.27±9.4 27.49±6.7	53.61±10. 42.71±8.1	19.16±5.6 22.93±4.2
Ours	**78.46**±6.3	**87.79**±4.0	**4.48**±1.9	**70.04**±15.	**81.30**±12.	6.53±4.6

[a] [b] The model is enhanced by the proposed enhancement strategy.

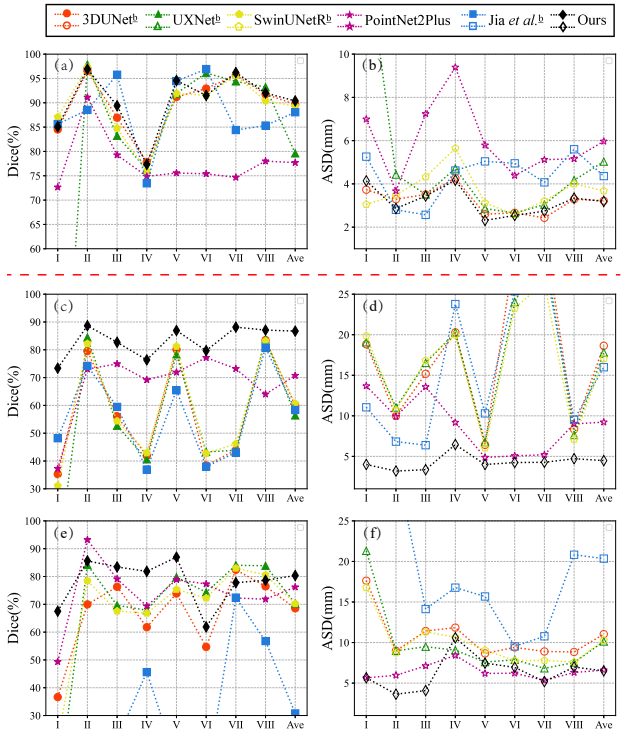

Figure 6: Evaluating the generalization effects of different enhancement methods on OOD datasets. (a) and (b) show the segmentation performance for each segment by different methods trained and tested on the 3Dircadb dataset. Based on the outcomes in Table 1, besides PointNet2Plus, we selected CNN-based methods that underwent two types of enhancement. (c) and (d) depict the results of the chosen methods when externally tested on the LiTS dataset. (e) and (f) demonstrate the external testing results of the selected methods on the MSD8 dataset.

during external testing, it also exhibited a certain degree of performance decline (Dice: 91.05 (within-distribution) → 81.30 (OOD)). We speculate that the main reason for this phenomenon is the substantial difference in average slice thickness between the MSD8 (18) and 3Dircadb (16) datasets (4.6mm vs. 1.78mm), against which our proposed dense sampling strategy effectively relieves (refer to the ablation analysis). Moreover, it is imperative to highlight that the segmentation of segment I (i.e., the caudate lobe), located at the posterior part of the liver, relies heavily on the model's comprehensive understanding of spatial structures and semantic information due to its relatively concealed anatomical position and the significant variability in shape and size among individuals. As depicted in Fig. 6, our method, integrating the complementary information of points and voxels, achieved markedly superior segmentation of segment I on both OOD datasets, exhibiting robust Dice coefficients and ASDs, which is critically important for clinical practice.

Table 6
Ablation study on different key components.

Method	3Dircadb (16)		LiTS (17)		MSD8 (18)	
	Dice (↑)	ASD (↓)	Dice (↑)	ASD (↓)	Dice (↑)	ASD (↓)
(a)	89.36±4.2	3.54±1.7	90.63±3.4	3.65±1.7	86.53±9.6	4.93±4.2
(b)	89.49±4.5	3.96±2.4	91.36±3.3	3.22±2.4	88.62±7.3	4.09±2.8
(c)	90.80±2.6	3.38±1.1	91.43±3.4	3.17±1.9	89.36±6.1	3.62±2.5
(d)	90.21±2.9	3.69±2.1	**91.45**±3.7	3.11±1.8	89.11±6.3	3.71±2.8
Ours	**91.05**±2.7	**3.19**±1.1	**91.45**±3.6	**2.95**±1.3	**89.46**±6.0	**3.44**±2.5

4.3. Ablation Study

To verify the effectiveness of crucial components in our approach, we conducted an ablation study by individually removing each component and comparing it to the complete pixel-voxel model. Table 6 shows the experimental results obtained on three datasets under the Dice and the ASD metrics. In Table 6, **(a)** indicates the removal of the dense sampling strategy embedded with vessel priors, choosing instead to sample T points solely from the liver space; **(b)** indicates that the point-based branch in the encoder is eliminated, thus the input point data is re-voxelized, aggregated through convolution to extract multi-scale information, and the Couinaud segmentation is output through an MLP-based decoder; **(c)** indicates the removal of the attention-based fusion module, where features from different branches are fused through simple addition operations; and **(d)** indicates the elimination of distance loss, relying on Dice and CE losses for model training. It's noteworthy that our previous version (19) could be considered a combination of **(c)** and **(d)**, namely removing the attention-based fusion module as well as the feature-based distance loss. Additionally, the results of our complete method are detailed in the last row. Compared with scenario **(a)**, our full method significantly improves Dice and ASD, particularly on the MSD8 dataset

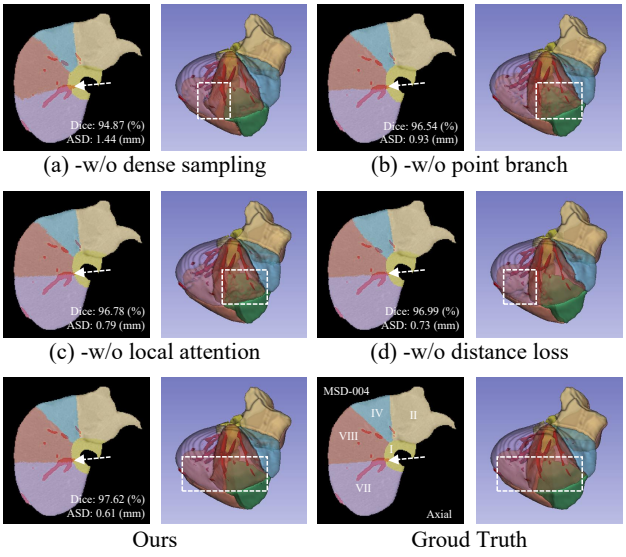

Figure 7: The visualization results of our method after removing different components (Best viewed electronically, zoom in).

(18) (Dice: 86.53 → 89.46, ASD: 4.93mm → 3.44mm). This improvement is attributed to the dense sampling strategy with vessel priors, which facilitates increased critical data access across the boundaries of the Couinaud segments. Moreover, this strategy allows for continuous, dense subpixel sampling in vital vessel regions, which is crucial for fine-grained semantic recognition in thick-slice CT images. For scenario **(b)**, since the input point features include spatial coordinates and voxel intensity, individual voxel branches can learn features from both perspectives as well. However, our method further enhances the extraction of fine-grained spatial information through the point branch in the encoder, thereby improving segmentation performance across all three datasets. Additionally, although the encoder in scenario **(b)** operates feature maps through the voxel branch, it still samples as points to connect with the decoder, potentially introducing confusion between feature concatenations. In Table 6, scenario **(c)**, which features two branches, uses simple feature fusion to mitigate the shortcomings of **(b)**. Nonetheless, compared with **(c)**, our method adaptively balances the contributions of both branches based on the attention module to achieve the best performance. Lastly, our complete approach introduces feature-based distance loss, enhancing the model's discriminative power in ambiguous inter-segment areas, thus maintaining a high Dice score while reducing ASD.

Fig. 7 presents the visualization of ablation results on a thick-layer CT data. Observing Fig. 7(a), we can see that the method without the dense sampling strategy struggles to accurately learn the course of vessels in thick-layer CT data (as seen in the 2D view), resulting in notable segmentation errors in the generated 3D Couinaud segments (as seen in the 3D view). Building upon the dense sampling, removing

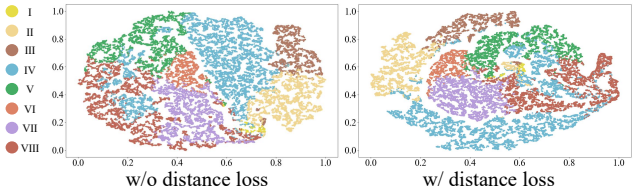

Figure 8: Visualization of the effect of distance loss on the features of the front layer of the classifier.

Table 7
The results of liver and vessel segmentation on different datasets, and the impact on Couinaud segmentation.

Dataset	Liver	Vessel	Couinaud		Couinaud (*Ref*)	
	Dice (↑)	Dice (↑)	Dice (↑)	ASD (↓)	Dice (↑)	ASD (↓)
3Dircadb	96.84	87.84	$89.15_{\pm 3.0}$	$4.25_{\pm 2.1}$	$91.05_{\pm 2.7}$	$3.19_{\pm 1.5}$
LiTS	94.05	—	$89.04_{\pm 3.9}$	$4.63_{\pm 1.7}$	$91.45_{\pm 3.6}$	$2.95_{\pm 1.3}$
MSD8	96.22	—	$86.91_{\pm 7.4}$	$4.24_{\pm 2.7}$	$89.46_{\pm 6.0}$	$3.44_{\pm 2.5}$

the point branch diminishes the capacity to extract fine-grained spatial information, hence producing uneven inter-segment segmentation planes (Fig. 7(b), 3D view). Similarly, by comparing the presentations in Fig. 7(c) and (d), the complete method, incorporating all components, accurately segments uncertain regions between Couinaud segments and yields smoother results in 3D view (Fig. 7(e)). Furthermore, to underscore the effectiveness of our proposed distance loss, we performed PCA on the features, followed by t-SNE visualization (as illustrated in Fig. 8). It is evident from Fig. 8 that employing distance loss significantly boosts the discriminative capability of the model-generated inter-segment features and contributes to smooth intersegment plane segmentation.

5. Discussion

5.1. Liver and vessel region analysis

The extraction of liver and vessel regions lays the groundwork for the task of Couinaud segmentation. In this study, our goal was to develop a fully automated framework capable of robustly extracting smooth Couinaud segmentations from CT images. To this end, we employed a trained 3D UXNet (33) model for liver extraction, alongside a semi-supervised model for extracting vessel attention maps across three datasets. Our experiments aimed to analyze the impact of extracting liver and vessel regions on Couinaud segmentation. As illustrated in Table 7, the first column on the left displays the Dice scores achieved by the trained 3D UXNet (33) model for liver segmentation across all three datasets, exceeding 94%. While we believe that training more complex networks individually on each dataset might yield higher performance, the results obtained with the pre-trained 3D UXNet (33) model demonstrate its powerful capability in segmenting the liver in CT images, fulfilling the requirements of this study. However, unlike liver segmentation, vessel segmentation from CT images poses a

Table 8
Analyze the influence of different settings of hyperparameter λ on three datasets.

λ	LiTS (17)			MSD8 (18)		
	Acc (↑)	Dice (↑)	ASD (↓)	Acc (↑)	Dice (↑)	ASD (↓)
0.1	63.70±16.	48.45±15.	15.34±5.7	55.78±19.	69.73±16.	9.74±3.1
0.01	83.93±6.3	91.02±3.9	3.29±2.2	**80.88±9.6**	**89.10±6.3**	4.08±3.1
0.001	**84.31±5.9**	**91.39±3.4**	**3.09±1.4**	80.80±9.5	89.06±6.3	**3.93±2.9**

Table 9
Analyze the influence of different settings of hyperparameter α. on three datasets.

α	LiTS (17)			MSD8 (18)		
	Acc (↑)	Dice (↑)	ASD (↓)	Acc (↑)	Dice (↑)	ASD (↓)
0.1	84.44±5.9	91.45±3.6	**2.95±1.3**	81.41±9.1	**89.46±6.0**	**3.44±2.5**
0.3	**84.61±5.8**	**91.56±3.5**	3.04±1.5	**81.43±9.4**	89.37±6.4	3.59±2.7
0.7	84.31±6.0	91.37±3.6	3.04±1.8	81.06±9.4	89.22±6.4	3.85±3.0
0.9	84.14±6.0	91.27±3.6	3.11±2.1	81.09±9.7	89.24±6.2	3.70±2.9

challenge due to the fine topology of vessels, compounded by the diverse imaging parameters, such as slice thickness, commonly encountered in clinical CT scans. Our proposed method leverages a dense sampling strategy to learn key vessel priors from coarse vessel attention maps for Couinaud segmentation across varied data distributions. To achieve this objective, we selected a semi-supervised vessel segmentation method, MixMatch (30), which required only one training session to achieve vessel segmentation across all three datasets. The second column in Table 7 showcases the testing results of the trained MixMatch (30) model. It is noteworthy that, due to the absence or low quality of vessel labels in the LiTS (17) and MSD8 (18) datasets, we only display vessel segmentation performance on the 3Dircadb dataset (16). After obtaining liver and vessel segmentation masks, we created three new point cloud test sets. Subsequently, testing these new sets with our trained models to perform Couinaud segmentation, the performances shown in the middle two columns of Table 7 slightly decrease compared to the results obtained on liver and vessel labels (denoted as *Ref* in the last column). This slight decline demonstrates that the liver and vessel segmentation methods chosen meet the initial intentions of our study, further illustrating our method's robustness to the input liver contours and vessel segmentation results, facilitating clinical application.

5.2. Hyperparameter Analysis in Distance Loss

In Couinaud segmentation tasks, smooth segmentation surfaces are easier to interpret and understand, providing clearer excision guidance for surgeons and facilitating the planning of surgical pathways. Therefore, we introduced a feature-based distance loss to enhance segmentation precision in the uncertain areas between Couinaud segments, resulting in smoother segmentations. Our proposed distance loss includes two key components and is based on the mutual adjustment of hyperparameters λ and α to exert a positive effect. We detailed the influence of hyperparameters λ and α on distance loss, as demonstrated in Table 8 and 9. Initially

Table 10
The average time cost (s) of processing a sample at different stages in our complete method.

Dataset	Liver	Vessel	Couinaud		
			Sam	Seg	Rec
3Dircadb (16)	35.6 (s)	6.4 (s)	57.3 (s)	26.4 (s)	7.6 (s)
LiTS (17)	91.2 (s)	6.9 (s)	81.5 (s)	36.0 (s)	14.8 (s)
MSD (18)	19.1 (s)	5.5 (s)	17.5 (s)	9.1 (s)	2.7 (s)

"Sam" denotes to sample point cloud data embeded vascular-priori from the liver space, "Seg" denotes to perform Couinaud segmentation on the input point cloud data, and "Rec" denotes to reconstruct the output point cloud segmentation results into 3D Couinaud mask.

setting α to 1, we adjusted λ to determine the weight of R_{intra}, the basic component of distance loss. We can observe that using smaller λ values leads to better segmentation performances while reducing the constraints on features imposed by R_{intra}. However, to enhance the discriminability of the model, and also consider that α will further reduce the weight of R_{intra}, we selected 0.01 as the setting of λ in this paper. Based on this setting, we further conduct several experiments on the LiTS (17) and MSD8 (18) datasets to evaluate the role of α. As shown in Table 9, our methods achieve the best performances on both datasets when α is set to 0.1 or 0.3, with 0.1 selected as the α setting for this comparative study.

5.3. Limitations and Future Trends

Table 10 shows the time consumption of our complete framework in inference, including liver segmentation, vessel region extraction, and Couinaud segmentation. It can be seen that Couinaud segmentation takes the longest time in the whole process, mainly because our method involves processing all individual points in the liver space. Therefore, to improve the efficiency of our method in clinical practice, we plan to design a fast point cloud segmentation and mask reconstruction based on regional broadcasting in the future.

6. Conclusion

In this study, we introduced a novel multi-scale point-voxel fusion framework for fully automated Couinaud segmentation, a critical task for liver surgery planning. By innovatively leveraging the topological relationships of coordinate points in 3D space and the rich semantic information encoded in voxel grids, our method not only recognizes but also intricately understands the spatial hierarchies and relationships crucial for precise Couinaud segmentation. Moreover, our approach is the integration of the dense point sampling strategy, enriched with vessel priors, which significantly enhances our model's focus on critical areas. This strategy facilitates a detailed understanding of the trajectories of key vascular structures, thus paving the way for safer surgical pathways that substantially minimize the risk of damaging major blood vessels. Furthermore, our framework introduces a feature fusion module utilizing local attention mechanisms, coupled with a distance loss function at the feature level. These innovations collectively enhance

the model's discriminative power, enabling the generation of accurate and smooth boundary segmentations. Such smooth segmentation surfaces are not only clinically interpretable but also provide surgeons with clear and precise excision guidelines, facilitating the meticulous planning of surgical approaches. Extensive experiments conducted across three publicly available datasets have demonstrated the superiority of our proposed method, especially its strong generalization capability on out-of-distribution (OOD) datasets, which surpasses that of current state-of-the-art (SOTA) methods. This significant generalization highlights the tremendous potential of our framework for clinical application, particularly in the preoperative phase of liver surgery, where it can contribute to optimizing surgical outcomes and enhancing patient safety.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

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] S. Campos, J.-W. Poley, L. van Driel, and M. J. Bruno, "The role of EUS in diagnosis and treatment of liver disorders," *Endoscopy International Open*, vol. 7, no. 10, pp. E1262-E1275

[2] X. Liu et al., "Secular Trend of Cancer Death and Incidence in 29 Cancer Groups in China, 1990-2017: A Joinpoint and Age-Period-Cohort Analysis," (in eng), *Cancer Manag Res*, vol. 12, pp. 6221-6238, 2020

[3] L. Satriano, M. Lewinska, P. M. Rodrigues, J. M. Banales, and J. B. Andersen, "Metabolic rearrangements in primary liver cancers: cause and consequences," *Nature Reviews Gastroenterology & Hepatology*, vol. 16, no. 12, pp. 748-766, 2019

[4] M. Feng, Y. Pan, R. Kong, and S. Shu, "Therapy of primary liver cancer," *The Innovation*, vol. 1(2), pp. 100032, Aug. 2022

[5] S. T. Orcutt, and D. A. Anaya, "Liver resection and surgical strategies for management of primary liver cancer." *Cancer Control*, 25(1), Jan. 2018

[6] C. Couinaud, "Liver Anatomy: Portal (and Suprahepatic) or Biliary Segmentation," *Dig Surg*, vol. 16(6), pp. 459—467, July. 1999

[7] R.M. Juza, and E.M. Pauli, "Clinical and surgical anatomy of the liver: A review for clinicians," *Clin. Anat.*, vol. 27, pp. 764–769, Jan. 2014

[8] R. C. Nelson, J. L. Chezmar, P. H. Sugarbaker, D. R. Murray, and M. E. Bernardino, "Preoperative localization of focal liver lesions to specific liver segments: utility of CT during arterial portography," *Radiology*, vol. 176, no. 1, pp. 89-94, Jul. 1990

[9] H. Bismuth, "Surgical anatomy and anatomical surgery of the liver," *World J. Surg.*, vol. 6, pp. 3—9, Jan. 1982

[10] Z. Arya, G. Ridgway, A. Jandor, and P. Aljabar, "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, 2021, pp. 227-237.

[11] D. Boltcheva, N. Passat, V. Agnus, M.-A. J.-D. Col, C. Ronse, and L. Soler, "Automatic anatomical segmentation of the liver by separation planes," in *Proceedings Volume 6141, Medical Imaging 2006: Visualization, Image-Guided Procedures, and Display,* San Diego, California, United States, 2006, pp. 383–394

[12] S. Huang, B. Wang, M. Cheng, W.-l. Wu, X. Huang, and Y. Ju, "A Fast Method to Segment the Liver According to Couinaud's Classification," in *Medical Imaging and Informatics. MIMI 2007,* Berlin, Heidelberg, 2007, pp. 270–076

[13] S. Pla-Alemany, J. A. Romero, J. M. Santabárbara, R. Aliaga, A. M. Maceira, and D. Moratal, "Automatic Multi-Atlas Liver Segmentation and Couinaud Classification from CT Volumes," in *43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC),* Mexico, 2021, pp. 2826-2829

[14] J. Tian, L. Liu, Z. Shi, and F. Xu, "Automatic Couinaud Segmentation from CT Volumes on Liver Using GLC-UNet," in *Machine Learning in Medical Imaging. MLMI*, Shenzhen, China, 2019, pp. 274-282.

[15] X. Jia et al., "Boundary-Aware Dual Attention Guided Liver Segment Segmentation Model," *KSII Trans. Internet Inf. Syst.*, vol. 16, no. 1, pp. 16-37, Jan. 2022

[16] L. Soler et al., "3D image reconstruction for comparison of algorithm database: A patient specific anatomical and medical image database." IRCAD, Strasbourg, France, Tech. Rep, 2010.

[17] P. Bilic et al., "The Liver Tumor Segmentation Benchmark (LiTS)," *Medical Image Anal.*, vol. 84, Feb. 2023

[18] A. L. Simpson et al., "A large annotated medical image dataset for the development and evaluation of segmentation algorithms," arXiv preprint arXiv:1902.09063.

[19] X. Zhang et al., "Anatomical-Aware Point-Voxel Network for Couinaud Segmentation in Liver CT," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2023.*, Vancouver, BC, Canada, 2023, pp. 465–474.

[20] D. Selle, B. Preim, A. Schenk, and H.-O. Peitgen, "Analysis of Vasculature for Liver Surgical Planning," *IEEE Trans. Medical Imaging*, vol. 21, no. 11, pp. 1344-1357, 2002

[21] D. A. B. Oliveira, R. Q. Feitosa, and M. M. Correia, "Automatic Couinaud Liver and Veins Segmentation from CT Images," in *In Proceedings of the First International Conference on Bio-inspired Systems and Signal Processing*, 2008, pp. 249–252

[22] M. Wang, R. Jin, J. Lu, E. Song, and G. Ma, "Automatic CT liver Couinaud segmentation based on key bifurcation detection with attentive residual hourglass-based cascaded network," *Comput. Biol. Medicine*, vol. 144, May. 2022

[23] Y. Zhou and O. Tuzel, "VoxelNet: End-to-End Learning for Point Cloud Based 3D Object Detection," 2018.

[24] Z. Liu, H. Tang, Y. Lin, and S. Han, "Point-Voxel CNN for Efficient 3D Deep Learning," 2019.

[25] H.-Y. Meng, L. Gao, Y.-K. Lai, and D. Manocha, "VV-Net: Voxel VAE Net With Group Convolutions for Point Cloud Segmentation," 2019.

[26] Diederik P. Kingma and Max Welling. Auto-encoding variational bayes. CoRR, abs/1312.6114, 2013.

[27] Z. Fang, B. Xiong, F. Liu, "Sparse point-voxel aggregation network for efficient point cloud semantic segmentation," *IET Comput. Vis.*, vol. 16, no. 7, pp. 644—654, Aug. 2022

[28] S. Shi et al., "PV-RCNN++: Point-Voxel Feature Set Abstraction With Local Vector Representation for 3D Object Detection," *Int. J. Comput. Vis.*, vol. 131, no. 2, pp. 531–551, Feb. 2023

[29] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation," in *Medical Image Computing and Computer-Assisted Intervention. MICCAI*, Athens, Greece, 2016, pp. 424–432.

[30] M. Xu et al., "Learning morphological feature perturbations for calibrated semi-supervised segmentation," in *Proceedings of The 5th International Conference on Medical Imaging with Deep Learning.* PMLR, vol. 172, pp. 1413-1429, 2022

[31] Y. Wen, K. Zhang, Z. Li, and Y. Qiao, "A discriminative feature learning approach for deep face recognition," in *Computer Vision–ECCV 2016: 14th European Conference,* Amsterdam, The Netherlands, Oct. 2016, pp. 499–515

[32] X. Wang, Z. Liu, and S. X. Yu, "Unsupervised Feature Learning by Cross-Level Instance-Group Discrimination," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 12586–12595

[33] H. H. Lee, S. Bao, Y. Huo, and B. A. Landman, "3D UX-Net: A Large Kernel Volumetric ConvNet Modernizing Hierarchical Transformer for Medical Image Segmentation," in *The Eleventh International Conference on Learning Representations,* ICLR, 2022

[34] A. Hatamizadeh, V. Nath, Y. Tang, D. Yang, H. R. Roth, and D. Xu, "Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images," in *International MICCAI Brainlesion Workshop*, Singapore, 2021, pp. 272-284

[35] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space," in *Advances in neural information processing systems.* NIPS, 2017.

[36] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015*, Munich, Germany, Oct. 2015, pp. 234–241