IEEE Journal of Translational Engineering in Health and Medicine

Received 15 January 2024; revised 26 June 2024 and 1 September 2024; accepted 28 October 2024.
Date of publication 4 November 2024; date of current version 15 November 2024.

Digital Object Identifier 10.1109/JTEHM.2024.3491612

A Multi-Task Based Deep Learning Framework With Landmark Detection for MRI Couinaud Segmentation

DONG MIAO[1,2], (Graduate Student Member, IEEE), YING ZHAO[3,4], XUE REN[3,4], MENG DOU[1,2], YU YAO[1,2], YIRAN XU[5], YINGCHAO CUI[5], AND AILIAN LIU[3,4]

[1]Chengdu Institute of Computer Application, Chinese Academy of Sciences, Beijing 100045, China
[2]School of Computer Science and Technology, University of Chinese Academy of Sciences, Beijing 101408, China
[3]Department of Radiology, The First Affiliated Hospital of Dalian Medical University, Dalian 116014, China
[4]Dalian Engineering Research Center for Artificial Intelligence in Medical Imaging, Dalian 116011, China
[5]School of Medical Imaging, Dalian Medical University, Dalian 116041, China

(Dong Miao and Ying Zhao contributed equally to this work.) CORRESPONDING AUTHORS: Y. YAO (casitmed2022@163.com) AND A. LIU (liuailian@dmu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61971091, in part by Dalian Young Stars of Science and Technology Project Support Program under Grant 2022RQ074, and in part by Dalian Medical Science Research Program under Grant 2212011.

ABSTRACT To achieve precise Couinaud liver segmentation in preoperative planning for hepatic surgery, accommodating the complex anatomy and significant variations, optimizing surgical approaches, reducing postoperative complications, and preserving liver function. This research presents a novel approach to automating liver segmentation by identifying seven key anatomical landmarks using portal venous phase images from contrast-enhanced magnetic resonance imaging (CE-MRI). By employing a multi-task learning framework, we synchronized the detection of these landmarks with the segmentation process, resulting in accurate and robust delineation of the Couinaud segments. To comprehensively validate our model, we included multiple patient types in our test set—those with normal livers, diffuse liver diseases, and localized liver lesions—under varied imaging conditions, including two field strengths, two devices, and two contrast agents. Our model achieved an average Dice Similarity Coefficient (DSC) of 85.29%, surpassing the next best-performing models by 3.12%. Our research presents a pioneering automated approach for segmenting Couinaud segments using CE-MRI. By correlating landmark detection with segmentation, we enhance surgical planning precision. This method promises improved clinical outcomes by accurately adapting to anatomical variability and reducing potential postoperative complications. Clinical impact: The application of this technique in clinical settings is poised to enhance the precision of liver surgical planning. This could lead to more tailored surgical interventions, minimization of operative risks, and preservation of healthy liver tissue, culminating in improved patient outcomes and potentially lowering the incidence of postoperative complications. Clinical and Translational Impact Statement: This research offers a novel automated liver segmentation technique, enhancing preoperative planning and potentially reducing complications, which may translate into better postoperative outcomes in hepatic surgery.

INDEX TERMS Couinaud segments, MRI, multi-task learning, landmark detection, segmentation.

I. INTRODUCTION

THE liver, a tremendously versatile organ, plays indispensable roles in metabolism, detoxification, and digestion within the human body [1], [2]. One of the critical aspects of understanding liver anatomy comes from a system proposed by Dr. Claude Couinaud, based on the intrinsic functional and vascular anatomy of the liver [3]. The Couinaud classification of liver anatomy divides the liver into eight functionally independent segments (I-VIII), ranging from the left lateral segment to the right posterior segment.

Accurate liver segmentation into these Couinaud sections is crucial in clinical practice, particularly for surgical

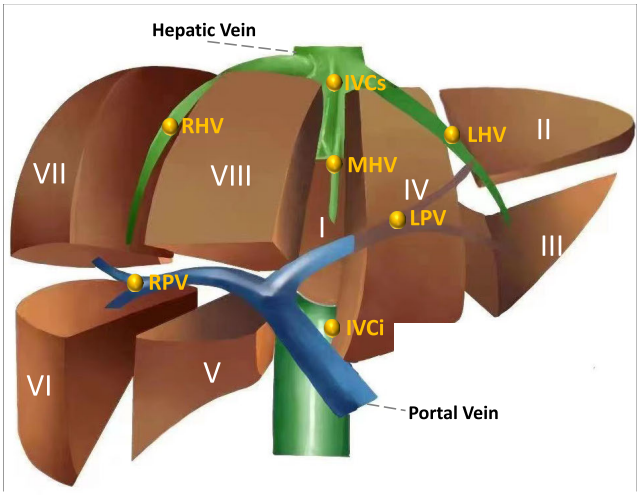


FIGURE 1. Schematic illustration of eight Couinaud liver segments (I-VIII) [16], where blue represents the portal vein and green the hepatic vein; we define seven landmarks on the vessel system, marked by the yellow origin.

planning and disease staging. As one of the main treatments for hepatocellular carcinoma, the type of hepatectomy mainly depends on the segmental positioning of the liver [4]. In addition, accurate identification of liver segments can reduce the risk of surgery and is useful in surgical planning and lesion monitoring [5]. As shown in Fig. 1, these segments are not arbitrary sections; they correspond intricately to the liver's internal vasculature [5]. Each Couinaud liver segment consists of distinct liver tissue that drains into a specific hepatic vein. Meanwhile, each segment is supplied by separate branches of the portal vein and hepatic artery. This classification method effectively aids in identifying the precise location of liver lesions and provides a systematic approach to hepatic resection. More specifically, the Couinaud system is structured around the concept of 'portal triads', which refer to the combination of a branch of the hepatic artery, hepatic portal vein, and a bile duct. The arrangement of these components dictates the division of the liver into its eight segments [6]. Essentially, the portal vein carries oxygen-poor but nutrient-rich blood from the gut to the liver, enabling the liver to metabolize absorbed nutrients and convert them into forms more accessible for the body, thereby profoundly influencing the segmental structure of the liver [7], [8].

Manual delineation of Couinaud segments on liver radiological images is labor-intensive and subject to considerable interobserver variability [9]. Automating this process can significantly reduce time, cost, and variability. With the recent advancements in deep learning, automated segmentation of medical images has gained significant advance. Deep learning methods, especially convolutional neural networks (CNNs) and Vision Transformers (ViTs), have demonstrated promising capabilities in segmenting liver structures accurately [10], [11], [12], [13]. Leveraging the inherent capacity of CNNs to learn complex spatial features, these models can effectively capture the subtle variations in liver anatomy and provide consistent and reproducible segmentations: Tian et al. [14] proposed a UNet-based segmentation network called GLCUNet, which segmented the liver first and then segmented the liver segments for Couinaud segmentation using computed tomography (CT) images. Han et al. [15] introduced a boundary layer and a pixel-wise layer in 3D UNet to simultaneously identify liver contours and predict Couinaud liver segments. However, segmentation-based networks have a significant drawback—for scans with hepatic focal lesions or diffuse diseases, their predictions produce many small irregular regions that deviate significantly deviate significantly from the standard Couinaud Segments, which results in unusability in the clinical process [15].

Rather than merely establishing a multi-target segmentation network, Wang et al. [17] defined six points on the liver to generate Couinaud segments. Consequently, the automated segmentation of eight Couinaud segments transformed into detecting six landmarks. They developed a network based on Attentive Residual principles to detect these six points and generate all liver segments. However, due to the non-regular shape of the first segment (also known as the caudate lobe) compared to other sections, the detection results were not highly satisfactory. Furthermore, the framework based solely on landmark detection presents potential implicit flaws: if the detection of a single point deviates even slightly, a significant portion of the resulting Couinaud segment may be misaligned.

In this paper, we propose a multi-task learning framework for the automatic segmentation of Couinaud segments in the portal venous phase contrast-enhanced magnetic resonance imaging (CE-MRI) scans.

The main contributions of our work can be summarized as follows:

1) We present a novel multi-task deep learning framework characterized by its simultaneous detection of critical landmarks and segmentation of Couinaud liver sections. This integrated approach within a single model results in the generation of highly accurate Couinaud liver segments.
2) We defined seven landmarks on the liver, which are essential for automatically generating segments II-VIII of the Couinaud liver sections. The accuracy of these key landmarks was meticulously assessed to ensure precise segment delineation by the radiologist.
3) Our research included patients with various liver conditions and used multiple imaging setups with different field strengths, equipment, and contrast agents. This comprehensive approach highlights the robustness and versatility of our findings, demonstrating their significant clinical utility in diagnosing and monitoring a wide range of liver diseases across diverse clinical settings.

II. RELATED WORKS

The current methods for the automatic segmentation of Couinaud segments can be categorized into three main types:

IEEE Journal of Translational Engineering in Health and Medicine

1) vessel identification-based methods, 2) segmentation-based methods, and 3) landmark detection-based methods. This section will present representative works of these approaches and discuss their advantages and limitations.

A. VESSEL IDENTIFICATION-BASED METHODS

Couinaud subdivided the liver into eight functionally independent segments according to hepatic and portal veins of liver vasculature. If the hepatic vein and portal vein can be extracted from the liver, it is relatively easy to identify liver segments based on anatomical planes. Therefore, the problem of segmenting the Couinaud segments shifts to accurately identifying hepatic and portal veins. We refer to such methods as vessel identification-based (ViB) methods.

Selle et al. [18] improved a threshold-based region-growing algorithm to segment intrahepatic vessels, and the segmented vessels were converted to directed acyclic maps for vessel separation and labeling based on the anatomical properties of the liver vessels due to the complexity of the vessel branching, finally the authors used two methods to generate into hepatic segments, the first one is the nearest-neighbor segment approximation method and Laplacian segment approximation to generate the final segments. Yang et al. [19] applied a customized level-set method to extract the liver, portal vein, and hepatic vein from CT images. They applied a semiautomatic approach to separate the main portal and hepatic veins. Constrained by the complexity of the blood vessels, the above two methods require manual involvement in the extraction or separation of blood vessels; Lebre et al. [9] proposed a complete automatic framework that automatically segments the liver segments mainly by shape model-based liver segmentation, followed by skeletonization process to extract the intrahepatic vessels automatically, and then finally, by detecting the four main directions on the hepatic blood vessels to construct planes for the liver to be divided into different liver segments. Alirr and Rahni [20] applied the atlas-based method to localize the portal and the hepatic veins bifurcations and used a fast marching scheme to extract the centerline to define the trajectories to define anatomical segments.

Despite the success of ViB methods in liver segment extraction, two significant drawbacks remain: 1) it lacks an objective assessment mechanism, with results evaluated manually by physicians, thus limiting precision due to individual subjectivity and experience; 2) the approach often necessitates iterative isolation and identification of intrahepatic blood vessels from images, resulting in longer inference times compared to end-to-end deep learning models.

B. SEMANTIC SEGMENTATION BASED METHODS

The success of deep learning semantic segmentation models, particularly UNet [21], in recent years has inspired researchers to apply deep neural networks to segment liver segments. UNet employs a symmetric encoder-decoder architecture to predict classes for voxels in medical images, demonstrating effective performance in this domain. Tian et al. [14] proposed a 2.5D class-aware segmentation model to obtain Couinaud segmentation from CT images. They collected 112 CT images from a publicly available dataset. The inference time of the model is only about 5 seconds. The results show that deep learning can make automatic liver segmentation more accurate and efficient; Jia et al. [22] built a 2D UNet based on Spatial and Semantic attention mechanisms for the segmentation of liver segments, which collected 59 MRI images; Han et al. [15] collected more MRIs (367 for training, 157 for validation, and 158 for testing) to train the 3D UNet model and tested it on three different devices, but did not incorporate comparisons of multiple contrast agents.

Deep learning segmentation networks can indeed segment liver segments efficiently and accurately. However, deep segmentation models usually require a considerable amount of manually labeled labels to train the model, and labeling liver segments requires some clinical experience, which burdens doctors. In addition, deep segmentation models operate on pixels and voxels, meaning that deviations within a region—such as tumors or artifacts—can lead to significant errors in segmentation. These models may struggle to accurately delineate boundaries in the presence of such irregularities, resulting in disproportionate inaccuracies.

C. LANDMARK DETECTION BASED METHODS

Inspired by the recent deep learning methods to detect landmarks, Arya et al. [23] first defined eight landmarks in liver vessels according to clinical anatomy guidelines and then used these eight landmarks to generate planes to divide the liver into different liver segments; To automate the whole process, they trained a 3DUNet to locate these eight landmarks, and the detection method was carried out through the heat map of regression points. They collected 122 cases of MRI data and tested them with 23 cases. However, their point-based method could only generate segments II to VIII and failed to predict segment I. Wang et al. [17] improved this approach by refining the eight points down to six and designing a method to generate the caudate lobe by drawing circles layer by layer. They applied their proposed ARH-CNet to a dataset of 193 CT cases. While their method yielded more accurate landmark detection, the reduced number of points did not translate to improved segmentation accuracy; in addition, the highly irregular and flexible contour of the caudate lobe posed limitations for their segmentation method.

III. METHODOLOGY

A. LANDMARKS AND COUINAUD SEGMENTS GENERATION

1) LOCATION OF SEVEN LANDMARKS

According to the definition from Germain et al. [24], we define seven landmarks for generating hepatic segments in the hepatic vein and portal vein, indicated by the orange dots in Fig. 1. These landmarks are: 1) Inferior vena cava inferior

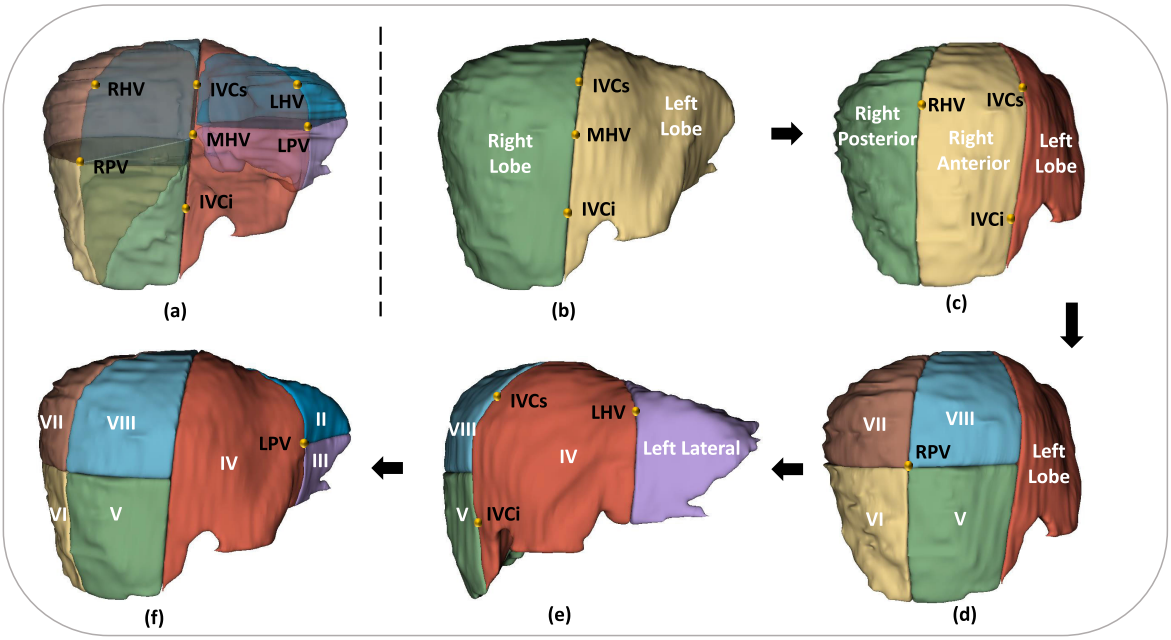


FIGURE 2. The whole process of generating Couinaud II-III segments based on defined seven landmarks, (a) showing the position of the plane on which the seven landmarks are located; (b)-(f) showing the area delimited during the generation process along with the corresponding landmarks.

part (IVCi); 2) Inferior vena cava superior part (IVCs); 3) Middle hepatic vein (MHV); 4) Right hepatic vein (RHV); 5) Left hepatic vein (LHV); 6) Right portal vein (RPV), and 7) Left portal vein (LPV).

2) COUINAUD SEGMENTS GENERATION

Fig. 2 illustrates the locations of the predefined seven landmarks marked on the axial plane (designated as (a)) and the overall process of liver segment generation (designated as (b-f)).

Initially, the liver is partitioned into a left lobe and right lobe along a plane determined using three landmarks: the IVCs, the IVCi, and the MHV as depicted in Fig. 2 (b). Subsequently, the right lobe is divided into the right posterior sector and right anterior sector along the plane defined by the points at IVCs, IVCi, and the RHV, as shown in Fig. 2 (c). In the next step, depicted in Fig. 2 (d), the right lobe is further partitioned into Segments V-VIII along the horizontal plane defined by the RPV. For the left lobe, as shown in Fig. 2 (e) and (f), it is initially divided into segments IV and left lateral along the plane formed by the IVCs, IVCi, and LHV, and then eventually into segments II and III through the horizontal plane formed by the LPV. Given the complexity and irregular boundaries of Segment I (caudate lobe), it was not partitioned via point-generated means. Instead, manual delineations were made within the dataset.

3) LANDMARK HEATMAP GENERATION

In this study, the initial step was to convert the manually annotated landmarks into a heat map representation following a Gaussian distribution. Consider the scenario where the original label is in the resolution $(1, H, W, D)$, with the digits 1 to 7 labeling seven pixels. Upon transformation, this format is modified to $(7, H, W, D)$, implying that an individual heat map now represents each pixel. The mathematical representation of this transformation is as follows:

$$L(x, y, z) = exp(-\frac{(x - l_x)^2 + (y - l_y)^2}{2\sigma_1^2} - \frac{||\, z - l_z \,||_2^2}{2\sigma_2^2}) \quad (1)$$

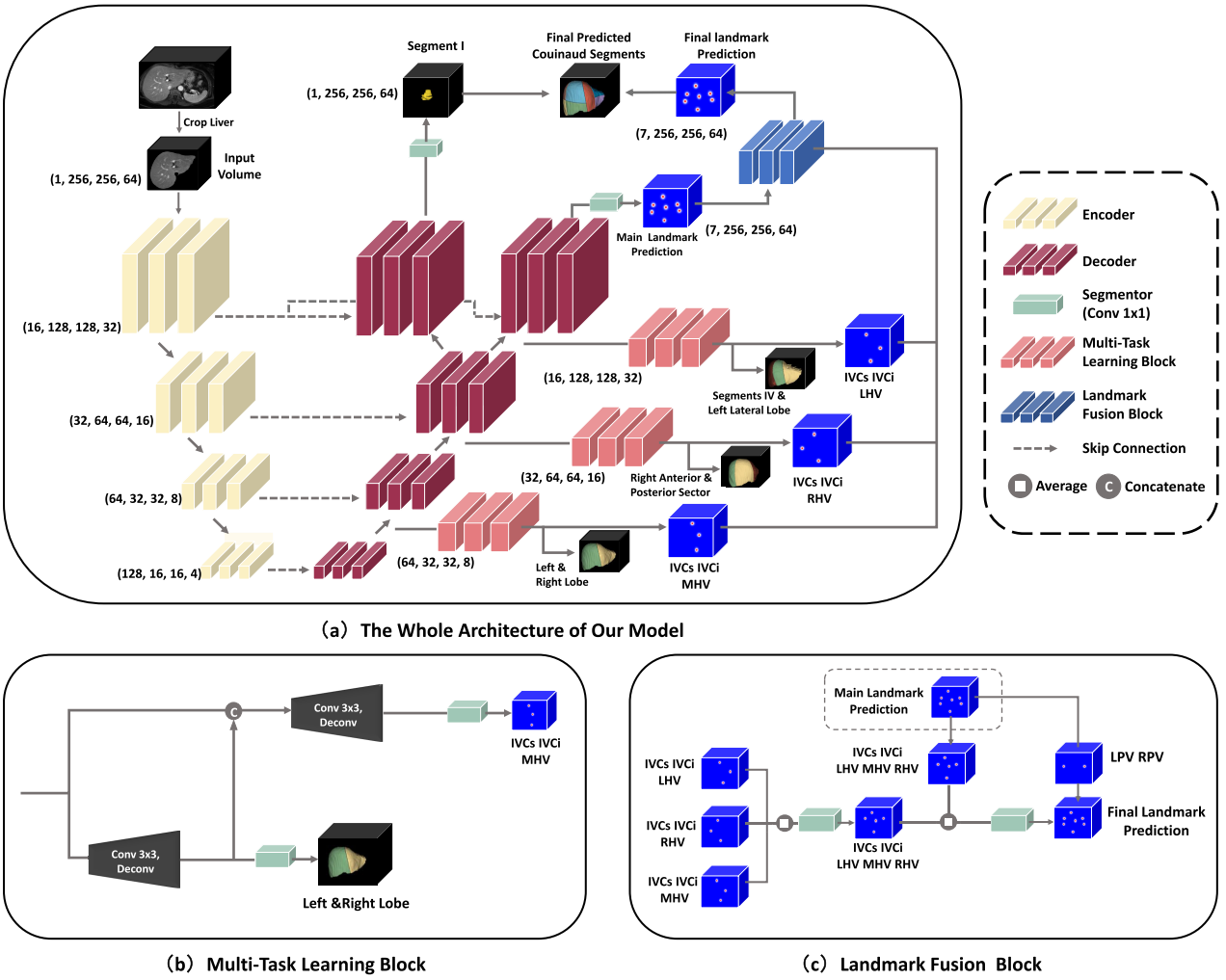
where (x, y, z) represents the coordinate of each pixel; (l_x, l_y, l_z) represents the coordinate of landmarks. σ is used to control the spread of each landmark to tackle the uncertainty in its representation. Where σ is used to control the spread of each landmark to ensure that all axes cover roughly the same range, based on the voxel sizes of the image(0.92, 0.92, 2.5), we set $\sigma_1 = 8$ for the x and y-axis, $\sigma_2 = 2$ for the z-axis.

This methodological adaptation facilitates reconfiguring the significant point detection task into a segmentation task. Consequently, the focus of the network shifts to predicting a heat map for each landmark directly rather than merely identifying their coordinates.

B. MULTI-TASK NETWORK ARCHITECTURE

The overall structure of the model is shown in Fig. 3 (a), which mainly consists of a CNN-based encoder and decoder; in each layer of the Decoder, we also add a multi-task block to associate specific Segments with their corresponding Landmarks for prediction, and then a landmark fusion block fuses the output of the model to generate the final landmark

IEEE Journal of Translational Engineering in Health and Medicine



(a) The Whole Architecture of Our Model

(b) Multi-Task Learning Block

(c) Landmark Fusion Block

FIGURE 3. An overall visualization of our automatic Couinaud segments detection model: (a) A diagram of the entire architecture of our proposed model; the model consists of three parts: an encoder that extracts the features, a decoder that decodes the features, and the multi-task learning block that follows the decoder at each level; both the decoder and encoder consist of a residual two-layer convolutional block followed by a Conv 2 × 2 or a Deconv 2 × 2 layer to compress and recover features; the penultimate layer in the decoder is divided into two branches simultaneously segmenting Segments I and detecting landmarks. (b) The multi-task learning block, here we use Left& Right Lobe and IVCs, IVCi, and MHV as examples to illustrate, it consists of several Conv 3 × 3 and Deconv 2 × 2 designed by us to first form a segmentation map, and then use the features in the segmentation map to guide the landmark heatmap for the detection of landmarks; (c) The landmark fusion process. The three heat maps in the MTLB are first fused by Conv 1 × 1, then averaged with the "Main Prediction", and then followed by Conv 1 × 1 to generate the final landmark prediction. Finally, the predicted landmarks will be used to generate Couinaud Segments II-VIII with the process in Fig. 2.

heatmap. In addition, to simultaneously detect Segment I, we set up two Decoders in the last layer to segment Segment I and predict the landmark heatmap.

Finally, we overlay the pixel-level segmentation mask for Segment I onto the combined masks for Segments II-VIII from the landmarks. We will elaborate on the details of each section below:

1) THE ENCODER AND DECODER

The underlying architecture of our network is inspired by the UNet-like encoder-decoder framework. As depicted in Fig 3(a), for a given liver CE-MRI scan, we initially utilized a pre-trained UNet model for liver auto-segmentation (with the DSC of 97%) to acquire a liver mask. After this, the image was cropped and padded to achieve a resolution of (256, 256, 64).

Each phase of the encoder primarily comprises 4 successive 3 × 3 convolutional operations, supplemented with a Parametric Rectified Linear Unit (PRelu) activation layer and a LayerNorm layer. These transformations adjust the number of channels in the feature maps to progressively higher resolutions throughout the encoding stage, resulting in a sequence of resolutions of (16, 32, 64, 128). In addition to these transformations, each encoding phase incorporates a 2 × 2 convolutional layer, which explicitly downsample the resolution of the feature map by a factor of 1/2.

The inputs for the decoder are derived from the output of the preceding layer and the merged features of the

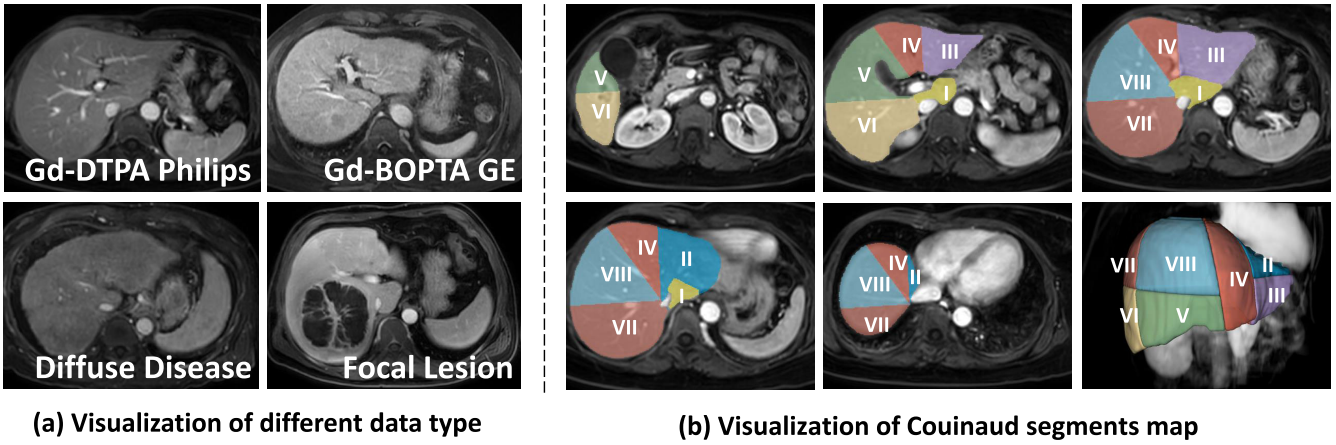


FIGURE 4. Visualization of our research dataset. (a) illustrates four different types of images from our dataset. The upper left image is a liver CE-MRI scan obtained using a 3.0T Philips CE-MRI scanner with the gadopentetate dimeglumine (Gd-BOPTA) contrast agent. The upper right image is another CE-MRI scan of a normal liver obtained using a 1.5T GE CE-MRI scanner with the gadopentetate (Gd-DTPA) contrast agent. The lower left shows a CE-MRI scan of cirrhosis, and the lower right shows a CE-MRI scan of hepatocellular carcinoma. (b) illustrates the Couinaud segments generated by the landmarks in the transverse and 3D view.

corresponding level from the encoder via Skip Connection. This design aims to preserve higher-resolution features from earlier in the network, enabling more detailed reconstructions in the decoder stages. By maintaining these features, the network can achieve finer granularity in the output. The decoder encompasses a 2×2 deconvolution operation, which serves to heighten the resolution of the feature map by doubling it while concurrently halving the count of channels. After this deconvolution, a set of 3×3 convolutions is implemented. The architectural configuration of these convolutions mirrors that of the encoder and includes a PRelu activation layer and a LayerNorm layer. These combined operations gradually transfigure the feature map into a target step-by-step. To concurrently predict seven landmarks and Segment I, we have incorporated two distinct yet parallel decoders within the two terminal layers of our architecture. This dual-decoder configuration caters explicitly to the two divergent tasks, thereby optimizing the performance yield for both objectives within an unified model.

Finally, a Segmentor consisting of a 1×1 convolution maps the channels to the target number to complete the final prediction.

2) MULTI-TASK LEARNING BLOCK

In Section III-A.2, we detail the process of generating liver segments II-VIII using individual landmarks. Each landmark plays a distinct role; for example, the IVCs and IVCi contribute to the formation of all liver segments, while the MHV is specifically responsible for dividing the liver into the left and right lobes.

Given these distinct functions of each landmark, it is intuitive to contemplate whether explicitly modeling the interdependencies amongst these components within the network and subsequently establishing a loss function to supervise this interaction would potentially enhance the detection accuracy of these landmarks. As shown in Fig. 3(a), we add the Multi-task Learning Block (MTLB) behind each of the decoders in levels 1-3 to predict both the partial landmarks as well as the hepatic segmentation region they are responsible for. The hepatic region predicted above in MTLB is to enhance the prediction accuracy of the corresponding landmarks that will not be involved in the forward propagation with the final Couinaud Segments generated.

More specifically, as shown in Fig. 3 (b), here we use Left& Right Lobe and IVCs, IVCi and MHV as examples to illustrate. Specifically, we perform successive Conv 3×3 and Deconv 2×2 oversampling of the output of the first layer of the decoder to a size consistent with the original image. Subsequently, we perform Conv 1×1 to separate the left and right lobes of the liver to extract the feature weights of these regions. We then merge these features with another parallel obtained up-sampled feature and subsequently apply Conv 3×3 to fuse them sequentially to generate a combined representation of the left and right lobes. This fusion process resulted in the generation of a predicted Landmarks Heatmap that encapsulates information about the left and right lobes of the liver. Similarly, we incorporate the segmentation of the Right Anterior and Posterior sectors into the second layer of the Decoder to guide the detection of IVCs, IVCi, and RHV. Additionally, we incorporate the segmentation of Segments IV and the Left Lateral in the third layer to enhance the detection of the IVCs, IVCi, and LHV.

RPV and LPV were excluded from additional detection efforts, given their relative ease of detection compared to the aforementioned points. Moreover, our experiments demonstrated that the improvements garnered for RPV and LPV were relatively insignificant.

IEEE Journal of Translational Engineering in Health and Medicine

3) LANDMARKS FUSION BLOCK

Upon securing these heatmaps, as discussed in the preceding section, it becomes imperative to integrate them with the final predicted landmarks heatmap. As illustrated in Fig. 3 (c), we first take the three heatmaps predicted by the Multi-Task Learning Block in the decoding branch, each with a resolution of (3, 256, 256, 64). The first two channels of each heatmap predict IVCs and IVCi. We begin by averaging these first two channels across the three heatmaps and then concatenate them with MHV, RHV, and LHV along the channel dimension. This concatenated feature map is then passed through a 1×1 convolution layer to obtain a fused representation with the same output dimension. Next, we compute the average of the above outputs with the corresponding five landmarks from the ''Main Landmark Prediction'' predicted by the backbone network shown in Fig. 3 (a) and followed by another 1×1 convolution. Finally, we concatenate the result with LPV and RPV from the backbone network to form the final landmark prediction.

Through this fusion block, we aim to enhance the accuracy of the predicted landmarks, thereby improving the generation of the final Couinaud Segments.

C. COMBINED LOSS FUNCTIONS

In terms of the loss function of the network, for the landmark detection task, we employed a loss function that combines the DICE with Kullback-Leibler (KL) Divergence Loss. This combination helps produce an output that closely resembles the distribution of the heatmap, as KL divergence ensures that the predicted probabilities align with the actual distribution patterns observed in the data, making it suitable for landmark heatmap predictions. For the segmentation task within the network, we utilized a loss function that merges the DICE with Cross-Entropy (CE) Loss to generate the segmentation map for the target region. By integrating these losses, the model benefits from the class-agnostic properties of DICE and the probabilistic classification capabilities of Cross-Entropy. This approach leads to models that are not only accurate in pixel-wise classification but also robust against varying class sizes and distributions. Combining these losses has the potential to yield models with improved generalization capabilities on unbalanced data [25], [26]. The formula is as follows:

$$\mathcal{L}_{DICE} = 1 - \frac{2\sum_{i=1}^{N} p_i q_i}{\sum_{i=1}^{N} p_i^2 + \sum_{i=1}^{N} q_i^2} \tag{2}$$

$$\mathcal{L}_{KL} = \sum_{i=1}^{N} p_i \log\left(\frac{p_i}{q_i}\right) \tag{3}$$

$$\mathcal{L}_{CE} = -\sum_{i=1}^{N} p_i \log(q_i) \tag{4}$$

where p and q denote the predicted and true labeling probabilities; N denotes the number of classes.

$$\mathcal{L}_{Land} = \mathcal{L}_{DICE} + \mathcal{L}_{KL} \tag{5}$$

$$\mathcal{L}_{Seg} = \mathcal{L}_{DICE} + \mathcal{L}_{CE} \tag{6}$$

It is worth noting that in the MTLB, we assigned equal weights of 0.5 to the aforementioned loss functions to ensure a balanced learning process and optimize the overall performance.

IV. EXPERIMENTS

A. DATASET AND METRICS

1) DATASET DETAILS

To enhance the robustness of our research, we integrate multiple types of data into our dataset. Specifically, we collected 113 CE-MRI scans from the Department of Radiology, the First Hospital of Dalian Medical University. Among these scans, 50 administered Gadopentetic acid (Gd-DTPA) as a contrast agent; 29 of these scans used a 3.0T scanner and the remaining 21 scans used a 1.5T scanner. Both scanners used the Signa Excite HDxt model from General Electric Healthcare. The remaining 63 patients were scanned with a 3.0T MR Ingenia CX scanner from Philips Healthcare and administered Gadobenate Dimeglumine (Gd-BOPTA) as a contrast agent.

All data were labeled by a radiologist with three years of experience with seven landmarks as defined by us and confirmed by a senior radiologist (more than eight years of clinical experience). In the experimental setup, we divided the different data types into train (78), validation (10), and test sets (25) according to their proportions, making sure that all types of images are included in each subset. Table 1 shows the details of our whole dataset.

In Table 1, we present a detailed information of our dataset, where the patient type includes three categories of liver conditions - Normal Liver, Hepatic Focal Lesions (incorporating Hepatocellular Carcinoma, Liver Hemangioma, Liver Cyst, and Intrahepatic Cholangiocarcinoma), and Hepatic Diffuse Diseases (encompassing Cirrhosis and Steatohepatitis). Fig. 3 illustrates a schematic representation of the different data types in our dataset and the Couinaud segments generated by the landmarks in the transverse and 3D view.

In terms of parameters, voxel sizes in the x and y axes ranged from $[0.7813, 0.7813]$ to $[0.9259, 0.9259]$, with a consistent slice thickness of 2.5 *mm* across all instances. The resolution in the X and Y range (432, 432) from (512, 512). Z-axis slices varied from 72 to 100. This width of collected data fosters comprehensive understanding and bolsters the capabilities of the derived model, underscoring the potential for practical applicability in real-world clinical contexts. All data involved in this study have been approved by the Ethics Committee on Biomedical Research, the First Affiliated Hospital of Dalian Medical University, ethics number PJ-KS-KY-2019-167.

2) EVALUATION METRICS

For the landmark detection task, we calculated the Euclidean Distance (ED) between the model output and ground truth

TABLE 1. The Details of our study dataset.

		Train(78)	Val(10)	Test(25)
Contrast Agent	Gd-BOPTA[a]	35	5	10
	Gd-DTPA[b]	43	5	15
MRI Scanner	GE	35	5	10
	Philips	43	5	15
Field Strength	1.5T	13	3	5
	3.0T	65	7	20
Liver Conidition	Normal Liver	7	1	2
	Focal Lesion	71	9	22
	Diffuse Disease[c]	20	2	9

[a] Gadobenate Dimeglumine [b] Gadopentetic acid
[c] A total of 30 scans presented both diffuse disease and localized lesions.

points:

$$ED(p, g) = \sqrt{\sum_{i=1}^{D=3}(p_i - g_i)^2} \tag{7}$$

where $p = (p_1, p_2, p_3)$ and $g = (g_1, g_2, g_3)$ represent the coordinates of the predicted points and the ground truth points in Euclidean space.

For the Couinaud segments evaluation, We quantitatively evaluated the model output on test sets by computing the mean value and standard deviation of two metrics: 1) Dice similarity coefficient (DSC) score to assess the overlapping volume between the predicted and the ground truth and 2) Average Surface Distance (ASD) to assess the distance between the surfaces of the predicted and the ground-truth. The formula is as follows:

$$DSC = \frac{2 \mid V_p \bigcap V_g \mid}{\mid V_p \mid \bigcup \mid V_g \mid} \tag{8}$$

$$ASD = \frac{\sum_{p \in P} min_{g \in G}(ED(p, g))}{|P|} \tag{9}$$

where V_p and V_g denote the volume of the output mask and the ground-truth mask, respectively; P and G are the boundary point sets of the predicted mask and ground truth.

B. IMPLEMENTATION DETAILS

We mainly implement our model using the deep learning framework PyTorch and MONAI [27]. All models are trained with the batch size of 2, using the AdamW [28] optimizer with an initial learning rate of $1e^{-4}$ for 5, 000 epochs. During training, the learning rate decays with the cosine annealing strategy. All models are trained using one NVIDIA RTX 3090. Due to the varying times for data loading and preprocessing across different hardware platforms, we report only the inference time per test sample during the model evaluation phase, which is 12.42 seconds. In a batch size of 2, the memory required for training is approximately 18.2 GB with Nvidia RTX 3090.

For the data preprocessing, the liver regions are first cropped from the entire CE-MRI scans by an automatic liver segmentation model, concentrating the focus on the areas of interest. To bring uniformity to the input of the model and to consistently envelop the complete liver area across all images, every image was subjected to a central crop, aligning them to a defined dimension of (256, 256, 64). This operation was based on the maximum liver size found within our dataset, thus ensuring the preservation of crucial liver information across all instances. Finally, intensities within all images were standardized by transforming the grayscale intensities to conform to a standard normal distribution. This step is generally beneficial to generalize the data, avoid any potential bias caused by the variability of intensities across different CE-MRI scans, and provide stability in training deep learning models. For data augmentation, we apply augmentation strategies such as random rotation of 90^o, 180^o, and 270^o degrees, random flip in axial, sagittal, and coronal views, and random scale and shift intensity. The random augment probability (whether or not to do augmentation) is set to 0.2.

C. QUANTITATIVE RESULTS

To demonstrate the efficacy of our proposed multi-task learning framework, we conduct a comprehensive comparative analysis with a variety of state-of-the-art medical image segmentation networks for both landmark detection and Couinaud segment segmentation tasks. Specifically, the competitive networks that we compare against include:

1) CONVOLUTION-BASED

3DUNet [10] and VNet [10] are among the most prevalent convolution-based segmentation networks utilized in the field of medical image analysis. These networks typically employ a down-sampling encoder block for feature extraction, followed by an up-sampling decoder for generating the segment map. Additionally, skip connections are incorporated to establish a link between the encoder and decoder, compensating for information loss resulting from the down-sampling process. A noteworthy contribution to this field is nnUNet [11], which introduced a self-adaptive framework based on vanilla U-Nets, enabling automatic adaptation to specific datasets.

2) TRANSFORMER-BASED

The challenge of learning global features in complex natural images has spurred the development of the Vision Transformer (ViT) [29] and its variant, the SwinTransformer [30]. Meanwhile, in the field of medical image segmentation, novel state-of-the-art networks such as UNetR [13] and Swin-UNetR [31] have also emerged, leveraging the power of the ViT and SwinTransformer respectively.

Table 2 presents a comparative analysis of our segmentation model to existing counterparts, all of which output nine-channel binary prediction maps encompassing background and segments I through VIII. Our method surpasses

IEEE Journal of Translational Engineering in Health and Medicine

TABLE 2. Quantitative comparison between segmentation-based methods and ours. (Segments I-VIII).

Segmentation Models	DSC(%) [mean±std]	ASD(mm) [mean±std]
3DUNet [21]	80.71 ± 10.57	3.63 ± 4.84
VNet [10]	82.17 ± 7.69	3.09 ± 2.13
nnUNet [11]	80.66 ± 9.83	3.85 ± 4.20
UNetR [13]	78.25 ± 10.88	5.28 ± 4.16
SwinUNetR [31]	78.95 ± 14.39	4.27 ± 4.60
Ours	$\mathbf{85.29 \pm 5.10}$	$\mathbf{2.08 \pm 1.51}$

TABLE 3. Quantitative comparison between landmark detection-based methods and ours. (Segments II-VIII).

Landmark Models	ED(mm) [mean±std]	DSC(%) [mean±std]	ASD(mm) [mean±std]
3DUNet [21]	18.66 ± 8.47	66.05 ± 14.09	5.79 ± 3.97
VNet [10]	10.86 ± 6.33	82.75 ± 9.66	2.47 ± 1.87
nnUNet [11]	14.42 ± 7.48	69.91 ± 14.11	5.81 ± 4.43
UNetR [13]	17.05 ± 10.40	68.59 ± 23.04	7.11 ± 13.06
SwinUNetR [31]	13.37 ± 6.95	75.16 ± 13.51	4.05 ± 3.84
Ours	$\mathbf{7.81 \pm 3.84}$	$\mathbf{85.29 \pm 5.10}$	$\mathbf{2.08 \pm 1.51}$

all currently available models in DSC and ASD. To elaborate, our proposed model delivers a mean DSC of 86.15% and a mean ASD of 2.08% for segments I through VIII. VNet achieves the second-best results with a mean DSC of 82.17% and a mean ASD of 3.09%. Notably, models relying on the Vision Transformer [32] failed to deliver competitive results, as is evident from the mean DSC of less than 80.00% achieved by both UNetR [13] and SwinUNetR [31].

In Table 3, the performance comparison between our model and the models mentioned above continues, but the focus shifted to the detection of liver segments using the significant point approach. Our model maintains its superior performance in this context, demonstrating an ED of 7.81 mm from the seven manually labeled significant points detected, with a standard deviation of 3.84 mm. The mean DSC for liver segments II through VIII, constructed from these seven landmarks, stands at 85.23%, with a mean ASD of 2.12 mm. Among the other models, VNet [10] continues to secure top-tier results, which may be attributed to the substantial number of filters in its dense connections, with a mean ED of 10.86 mm. Interestingly, Transformer-based models exhibit contrasting results from the segmentation task. While UNetR [13] sustains its underperformance, SwinUNetR [31] delivers the third-best result.

Table 4 illustrates the results between the predicted landmarks of each model and the corresponding ground truth. Considering the distinctive responsibilities and individual characteristics of each landmark, our model consistently show superior performance by minimizing the overall ED

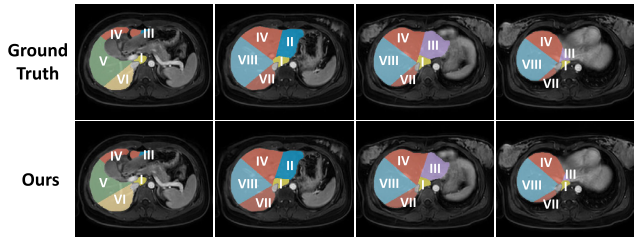


FIGURE 5. Visualization in the 2D transverse view of the ground truth Couinaud segments compared with the prediction of our model.

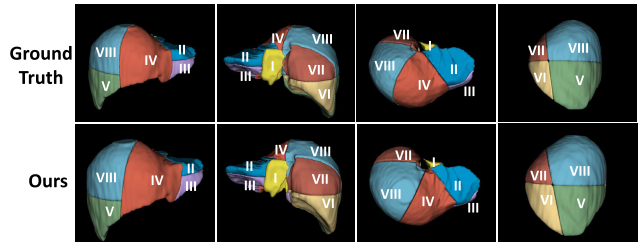


FIGURE 6. Visualization in the 3D view of the ground truth Couinaud segments compared with the prediction of our model.

in landmark prediction. Notably, IVCi was particularly well-predicted with an impressive ED of 4.60 *mm*. Regarding the hepatic vein system, it is worth noting that the three landmarks—MHV, RHV, and LHV exhibited relatively larger errors compared to others, with our model yielding ED of 10.36*mm*, while other models surpassed 15*mm*.

Table 5 simultaneously presents the DSC for liver segmentations generated based on landmark detection models (marked with '*') and segmentation-based approaches (segments I-VIII). Once again, our model demonstrates competitive results across different segments, with higher DSC than all other models except for segments I and VIII. Particularly noteworthy is the performance of our model in segment II, where the landmark detection models struggled due to errors in detecting IVCs, IVCi, LHV, and LPV, leading to relatively low DSC. In contrast, our model recorded an impressive mean DSC of 83.84% in this segment.

Regarding segment I detection, our model records a mean DSC of 87.83%. While it may not exhibit the best results individually, the overall average remains competitive. This outcome can be attributed to integrating segmentation and landmark detection tasks into a single model. This amalgamation offers a significant advantage compared to other pure landmark detection models that cannot predict segment I.

D. QUALITATIVE RESULTS

To further illustrate the effectiveness of our model, Fig. 5 and Fig. 6 demonstrate a comparison between manually labeled Couinaud segments (marked as Ground Truth) from a sample in the test set and the 2D and 3D visualization of our predicted results. As evident from these multi-perspective outputs, our model can accurately delineate the boundary planes between each liver segment and appropriately segment the caudate

TABLE 4. Quantitative result comparison across seven landmarks between based methods and ours.

Models	ED(mm)[mean±std]						
	IVCs	IVCi	MHV	RHV	LHV	RPV	LPV
3DUNet [21]	13.54 ± 4.72	16.23 ± 5.23	15.58 ± 9.00	21.79 ± 11.94	14.82 ± 10.67	32.90 ± 9.34	15.81 ± 8.38
VNet [10]	9.26 ± 3.88	8.22 ± 4.88	12.82 ± 8.57	16.04 ± 8.10	11.81 ± 8.71	9.45 ± 6.85	8.45 ± 3.36
nnUNet [11]	11.54 ± 4.72	8.42 ± 3.28	17.11 ± 8.25	17.99 ± 6.82	15.53 ± 8.81	12.05 ± 6.51	18.32 ± 13.98
UNetR [13]	11.93 ± 10.40	12.13 ± 11.03	17.37 ± 12.57	23.65 ± 11.58	24.50 ± 8.53	17.13 ± 14.74	12.68 ± 5.91
SwinUNetR [31]	9.19 ± 4.74	7.91 ± 3.30	17.00 ± 11.58	20.92 ± 7.82	15.95 ± 10.64	8.75 ± 5.52	13.58 ± 5.09
Ours	$\mathbf{5.86 \pm 2.68}$	$\mathbf{4.60 \pm 2.71}$	$\mathbf{9.10 \pm 6.75}$	$\mathbf{10.36 \pm 7.63}$	$\mathbf{8.68 \pm 7.22}$	$\mathbf{8.06 \pm 6.04}$	$\mathbf{8.02 \pm 4.41}$

TABLE 5. Quantitative result comparison of Couinaud Segments across the landmark and segmentation methods. Note here " model* " represents the results of using the seven detected landmarks to generate liver segments II-VIII.

Models	DSC(%)[mean±std]							
	I	II	III	IV	V	VI	VII	VIII
3DUNet* [21]	-	42.31 ± 22.65	74.27 ± 9.85	82.06 ± 8.63	64.86 ± 15.25	54.03 ± 20.61	71.13 ± 14.41	73.68 ± 7.20
3DUNet [21]	88.43 ± 4.96	77.43 ± 11.02	84.86 ± 8.58	78.99 ± 12.03	82.32 ± 9.87	71.35 ± 16.89	78.49 ± 15.35	84.02 ± 5.90
VNet* [10]	-	72.01 ± 4.72	83.63 ± 10.04	85.51 ± 9.41	85.69 ± 6.09	79.10 ± 11.16	86.05 ± 7.49	$\mathbf{87.25 \pm 5.56}$
VNet [10]	$\mathbf{91.11 \pm 3.47}$	77.83 ± 11.11	85.03 ± 7.73	79.59 ± 11.31	84.21 ± 5.87	75.87 ± 9.82	80.26 ± 7.52	83.98 ± 4.98
nnUNet* [11]	-	55.46 ± 21.42	73.53 ± 12.85	79.29 ± 6.35	77.63 ± 11.33	52.23 ± 25.84	75.36 ± 11.84	75.86 ± 9.34
nnUNet [11]	86.69 ± 11.48	78.20 ± 10.91	84.71 ± 8.42	79.47 ± 9.13	82.78 ± 8.84	71.68 ± 14.20	78.95 ± 10.03	82.76 ± 5.53
UNetR* [13]	-	52.30 ± 22.12	74.47 ± 20.22	76.17 ± 21.43	65.52 ± 27.89	67.77 ± 23.36	75.39 ± 20.57	68.54 ± 25.66
UNetR [13]	81.45 ± 15.88	74.60 ± 11.30	82.55 ± 9.47	77.57 ± 8.60	80.43 ± 9.45	70.80 ± 14.14	77.13 ± 12.98	82.57 ± 5.87
SwinUNetR* [31]	-	57.43 ± 16.39	77.68 ± 10.87	77.21 ± 17.25	80.61 ± 9.99	72.36 ± 17.87	79.12 ± 13.76	81.69 ± 8.41
SwinUNetR [31]	83.72 ± 8.37	71.80 ± 18.48	79.85 ± 10.91	79.32 ± 11.48	79.88 ± 20.25	73.70 ± 21.31	80.92 ± 12.84	82.46 ± 11.50
Ours	87.83 ± 3.55	$\mathbf{83.84 \pm 8.28}$	$\mathbf{85.80 \pm 9.04}$	$\mathbf{86.02 \pm 8.05}$	$\mathbf{85.74 \pm 9.13}$	$\mathbf{82.49 \pm 9.56}$	$\mathbf{86.63 \pm 7.26}$	86.94 ± 5.61

lobe. This achieves a high level of consistency with the Ground Truth, supporting the efficacy of our methodology.

To provide an intuitive comparison of our model with other models that only rely on landmark detection and segmentation and to illustrate the features of these different modeling approaches, we have selected Couinaud Segments segmentation results from representative slices of two patients. Fig. 7 8 depicts (a) and (b) from a patient with cirrhosis and (c) and (d) from a patient with a sizable tumorous growth, respectively. Fig. 7 visually compares our model and the landmark detection model. The first column represents the Ground Truth, and the second column represents the predictions from our model, and columns three through seven show predictions from the other models. Fig. 7(a) showcases how our model, compared to the other models, could accurately detect the Left Hepatic Vein (LHV), allowing for a precise demarcation between the left medial and left lateral liver lobes. Only the VNet model, among the others, managed to segregate Segment III in one slice, yet its deviation from the Ground Truth was still considerable. Additionally, both the UNet and UNetR models incorrectly recognized Segments V and VI, indicating an inaccuracy in their Right Portal Vein (RPV) detection. Fig. 7 (b), (c), and (d) also demonstrate the accurate detection of the LHV by segregating Segments IV and III. Though models based on landmark detection can generate regular Couinaud Segments, as seen in the figures, they fail to generate Segment I, underscoring a limitation of such models.

Fig. 8 presents a visual contrast between our model and automatic segmentation models. The segmented illustrations from these models are less regular than those from the automatic segmentation ones. These models tend to make many errors, especially on slices where lesions or blood vessel delineation is not apparent, as seen in Fig. 8 (a) - (d). In Fig. 8 (a), despite accurately identifying Segments IV and III, they produce numerous errors in Segments V and VI. This pattern recurs in the following examples as well. In Fig. 8 (b), except for UNetR, all segmentation models mistakenly include a portion of Segment III in Segment II. For Fig. 8 (c), both nnUNet and UNetR wrongly predict areas meant for Segment III as Segment II. Concurrently, UNet, VNet, and SwinUNetR all indulge in minor areas of incorrect prediction. Fig. 8 (d) depicted acceptable overall predictions from the segmentation models, albeit with several mistakes.

IEEE Journal of Translational Engineering in Health and Medicine

TABLE 6. Ablation Study of our Multi-Task learning block. 'WithDe_*' signifies the inclusion of a secondary MTLB at the corresponding '*' level of the decoder.

Multi-Task	ED(mm)[mean ± std]							
	IVCs	IVCi	MHV	RHV	LHV	RPV	LPV	Avg
Without	10.63 ± 4.25	8.66 ± 4.08	12.57 ± 8.45	17.00 ± 7.61	11.20 ± 8.90	8.11 ± 6.64	8.05 ± 5.07	10.88 ± 6.43
WithDe_1	6.44 ± 2.39	5.48 ± 2.97	9.40 ± 6.38	16.32 ± 7.37	10.40 ± 7.38	8.23 ± 6.74	8.14 ± 5.32	9.20 ± 5.51
WithDe_2	$\mathbf{5.66 \pm 5.42}$	5.42 ± 3.32	12.14 ± 8.14	10.63 ± 4.25	11.17 ± 8.79	8.45 ± 6.17	8.12 ± 5.28	8.80 ± 5.91
WithDe_3	6.27 ± 3.25	4.98 ± 2.84	11.41 ± 9.49	16.66 ± 7.21	$\mathbf{8.65 \pm 7.15}$	8.29 ± 7.16	8.03 ± 4.97	9.18 ± 6.01
WithDe_23	$\mathbf{5.66 \pm 2.66}$	5.02 ± 3.32	11.74 ± 8.05	$\mathbf{10.22 \pm 9.30}$	8.88 ± 5.10	$\mathbf{8.06 \pm 6.33}$	$\mathbf{7.99 \pm 4.25}$	8.22 ± 5.72
WithDe_123	5.86 ± 2.68	$\mathbf{4.60 \pm 2.71}$	$\mathbf{9.10 \pm 6.75}$	10.36 ± 7.08	8.68 ± 7.63	$\mathbf{8.06 \pm 6.04}$	8.02 ± 4.41	$\mathbf{7.81 \pm 5.33}$

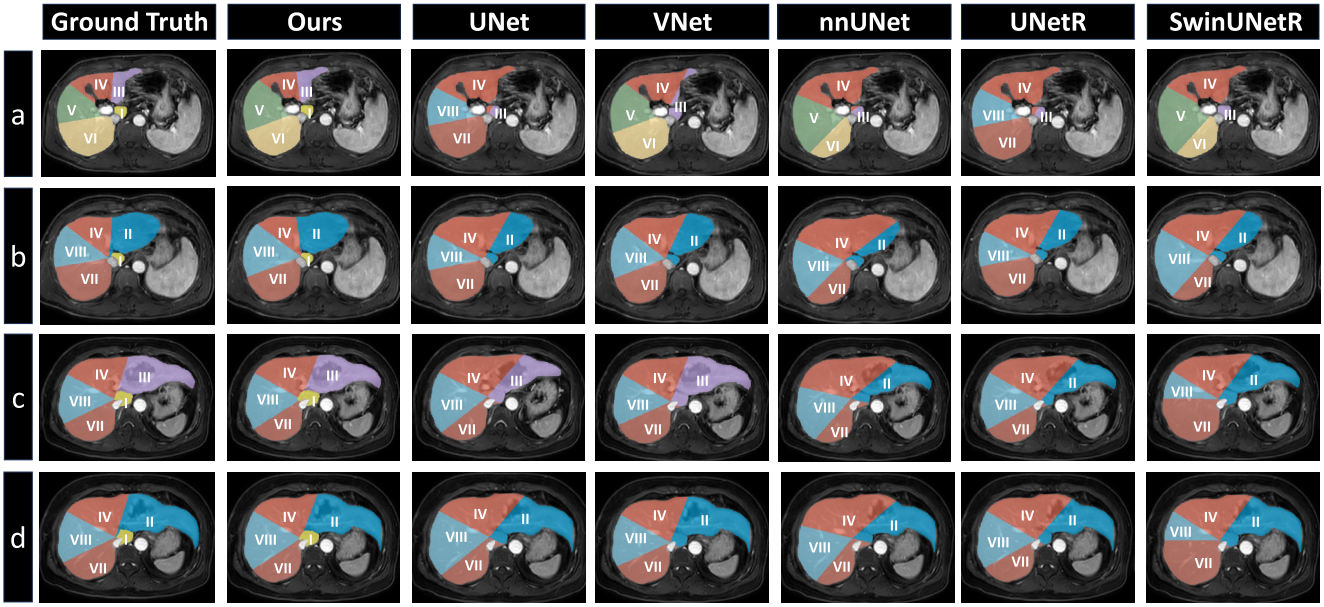


FIGURE 7. Visual comparison between the ground truth Couinaud segments and prediction of Landmark detection-based models.

VNet erroneously classifies the lesion area as background, and nnUNet includes a small portion of Segment I within Segment II.

E. ABLATION STUDY

1) MTLB BLOCK

Our designed MTLB is instrumental in delineating Couinaud liver segments. During the decoding stage, the MTLB predicts corresponding segmentation maps for landmarks such as the IVCs, IVCi, and MHV, thereby emulating the Couinaud segment generation process as informed by specific anatomical landmarks. We aim to integrate valuable a priori knowledge about Couinaud segment division into the decoder, enhancing the precision of landmark detection and ensuring more coherent segmentation of liver segments.

To verify the efficiency of the MTLB, we present a comparison under two conditions: in the absence of the MTLB and with a progressive inclusion of the MTLB. The results are elaborated in Table 6. In Table 6, the term 'Without' indicates landmark detection without the MTLB. 'WithDe_*' signifies the inclusion of a secondary MTLB at the corresponding '*' level of the decoder, and 'Avg' refers to average mED values across seven landmarks. The final row illustrates the result obtained from our complete model, achieving the lowest mED on average. To be more specific, the initial four rows evidence that the introduction of the MTLB at the first three decoder levels (De_1, De_2, De_3) enhances the detection accuracy of IVCs and IVCi, in addition to improving the precision for MHV, RHV, and LHV. Moreover, simultaneous implementation of the MTLB at Decoder levels 2 and 3 both improves RHV and LHV detection accuracy, although the impact on MHV detection is marginal. This highlights that the MTLB at each decoder level independently bolsters the accuracy of associated landmarks. Thus, we deployed MTLB across all three decoder levels in our final approach. Lastly, RPV and LPV were not configured with the MTLB due to

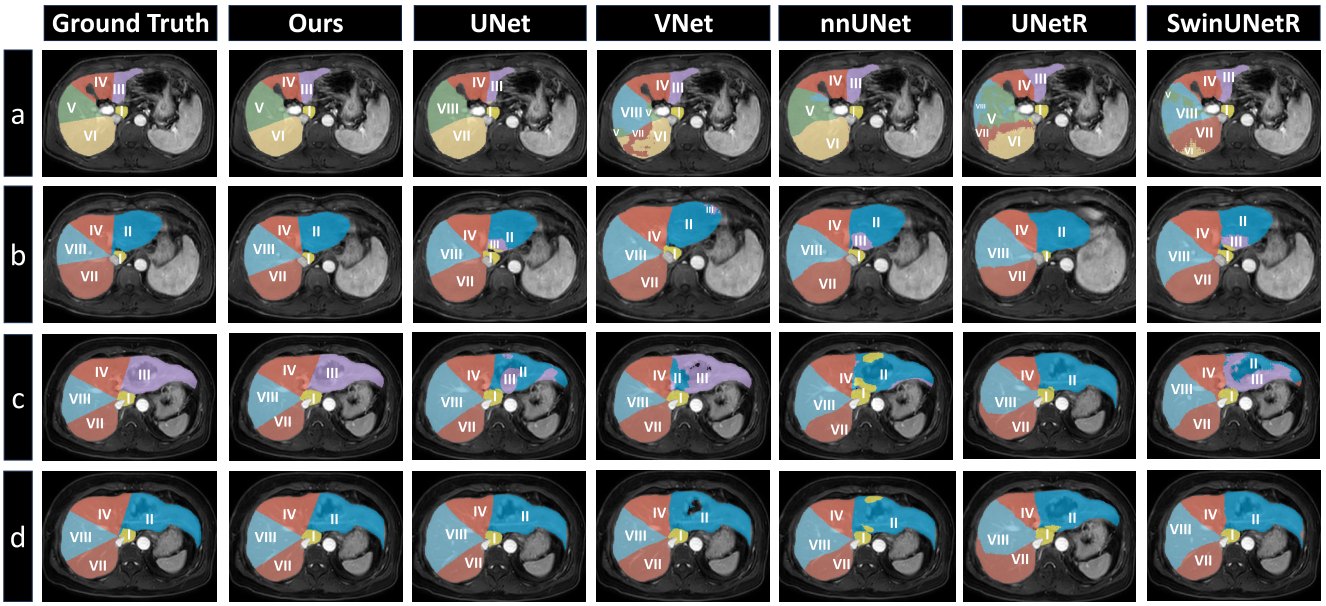


FIGURE 8. Visual comparison between the ground truth Couinaud segments and prediction of Segmentation-based models.

a lack of observed enhancement in detection accuracy. This results from the intrinsic ability of the model to accurately detect these two clear bifurcation points, making the MTLB application redundant in these specific instances.

2) LOSS FUNCTION

To evaluate the necessity of each term in our combined loss function while also explaining why we chose this loss. The experiments involved training our model using the following configurations: 1) only the CE loss, 2) only the KL loss, 3) only the Dice loss, 4) a combination of Dice and CE loss, and 5) a combination of Dice and KL loss. Figure 1 shows the results for the landmark detection task, and Figure 2 illustrates the results for the Segments I segmentation task. It is important to note that since our network needs to perform both landmark detection and the segmentation task simultaneously, the significant point detection results in Figure 1 were derived under the best segmentation task loss (DiceCE). Conversely, the results for Segments I in Figure 2 were obtained with the best landmark detection loss (DiceKL).

As illustrated in Fig. 9, DiceKL loss achieved the best results, yielding the minimal mean Euclidean Distance (mED) for seven landmarks and consequently attaining the highest DSC in Segments II-VII. In contrast, the results obtained using only CE loss or a combination of Dice and CE losses were the lowest and second lowest, respectively. We surmise that this discrepancy arises because CE loss tends to prioritize the background or more numerous classes, thereby impairing the performance in the landmark detection.

As evident from Fig. 10, unlike in landmark detection, KL loss exhibited poor performance in the segmentation task. Whether combined with Dice loss or used independently, the results did not surpass those obtained without KL loss. This

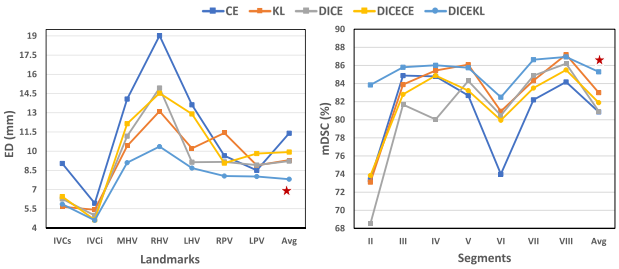


FIGURE 9. Visualization of ablation experiments on the loss function for the landmark detection task; the left shows the mED of the seven landmarks; the right shows the DSC for the liver segment generated from the seven landmarks.

may be attributable to the fact that segmentation tasks typically involve discrete pixel labels, whereas KL loss assumes continuous probability distributions. This mismatch could lead to ineffective handling of discrete labels. In contrast, the use of either Dice loss or CE loss yielded better results than those achieved with KL loss. The efficacy of combining Dice and CE losses has been validated in several segmentation studies [11], [31], [33], and our experimental results further corroborate this finding.

V. DISCUSSION

In this article, we explore the problem of automatic segmentation of Couinaud segments in the CE-MRI scans. By analyzing mainstream methods, including those based on landmark detection and automatic segmentation, we identify the strengths and limitations of each approach. While models based on landmark detection produce ‘‘regulated’’ sectional divisions of the liver by anatomical definitions, their results are heavily correlated with landmark detection

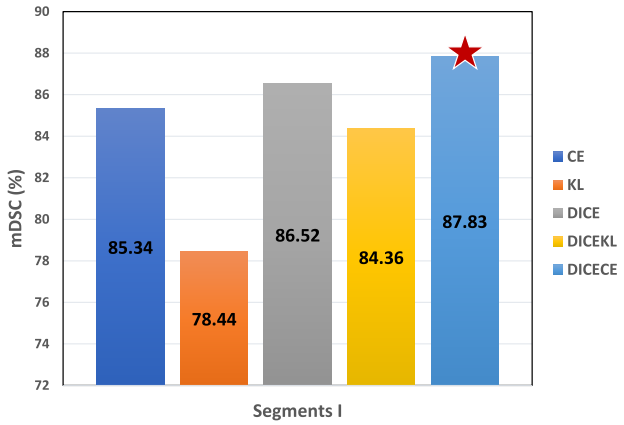


FIGURE 10. *Visualization of ablation experiments on the loss function for the SegmentsI segmentation task.*

accuracy. Consequently, lower detection accuracy leads to reduced overall precision and a decreased DSC. Additionally, landmark detection models struggle to accurately delineate Segment I. In contrast, segmentation models excel in detailed detection, yielding higher DSC values. However, they may produce significant errors when encountering abnormalities, such as pathological changes or cirrhosis, resulting in outputs that deviate from expected anatomical structures.

We defined seven landmarks based on the hepatic venous and portal venous anatomic systems of the liver and introduced a deep learning model that encapsulates segmentation and landmark detection. Within this model, we designed a multi-task block in the "Deep Supervision" paradigm, which constructs dependencies between each landmark and segment, aiming to boost the precision of the seven-point detection. Ablation experiments in the experimental section fully demonstrated the effectiveness of this design. Additionally, the model includes two branches, allowing both the detection of landmarks and the segmentation of Segment I concurrently.

To validate the effectiveness of our model, we employed a dataset encompassing various types and quantitatively evaluated the landmark detection and segmentation models using ED and DSC as our primary measurements. Our model achieved best-in-class performance in both metrics. Qualitatively, our model could generate a relatively superior predictive structure across different data types. The clinical applicability of our research is evident in its potential to enhance the efficiency of liver surgery planning and, consequently contribute to more precise surgical interventions and improved patient outcomes.

Nonetheless, our research presents certain limitations and opportunities for further improvements. Firstly, our segmentation results for the caudate lobe did not yield the best results, as we aim to predict all segments within a unified model. Although developing a separate model for the caudate lobe segmentation might slightly enhance the overall results, it would simultaneously introduce a larger number of parameters and prolong inference time. Secondly, our multi-task learning approach utilized distinct liver regions to offer auxiliary information for relevant landmark detection. Given that these landmarks are positioned on the hepatic vessels, it could be advantageous, intuitively, to establish accurate relationships between the vessels and these points for mutual support. However, due to the intricate nature of the vascular system and the difficulties associated with annotation, we refrained from adopting this strategy in our current investigation.

VI. CONCLUSION

This study presented an innovative deep learning framework incorporating multi-task learning for predicting Couinaud segments in CE-MRI scans. To the best of our knowledge, this is the first model that combines landmark detection and segmentation methodologies for this purpose. Our experiments, incorporating diverse data types, showed that the model exhibits strong quantitative and qualitative performance. The ability of our framework to automatically segment Couinaud segments offers a valuable tool that could significantly enhance surgical planning and improve clinical outcomes.

ACKNOWLEDGMENT

The authors would like to express their thanks to Prof. Hongkai Wang from the Faculty of Medicine, Dalian University of Technology, for his guidance on this work.

REFERENCES

[1] J. Hyun, J. Han, C. Lee, M. Yoon, and Y. Jung, "Pathophysiological aspects of alcohol metabolism in the liver," *Int. J. Mol. Sci.*, vol. 22, no. 11, p. 5717, May 2021.

[2] H. Qian et al., "Autophagy in liver diseases: A review," *Mol. Aspects Med.*, vol. 82, Dec. 2021, Art. no. 100973.

[3] C. Couinaud, "Liver anatomy: Portal (and suprahepatic) or biliary segmentation," *Digestive Surg.*, vol. 16, no. 6, pp. 459–467, 1999.

[4] M. Yamamoto, S. Katagiri, S. Ariizumi, Y. Kotera, Y. Takahashi, and H. Egawa, "Tips for anatomical hepatectomy for hepatocellular carcinoma by the glissonean pedicle approach (with videos)," *J. Hepato-Biliary-Pancreatic Sci.*, vol. 21, no. 8, pp. E53–E56, Aug. 2014.

[5] D. A. Oliveira, R. Q. Feitosa, and M. M. Correia, "Segmentation of liver, its vessels and lesions from CT images for surgical planning," *Biomed. Eng. OnLine*, vol. 10, no. 1, p. 30, 2011.

[6] Y.-C. Hsu, H.-Y. Lee, C.-M. Chang, C.-Y. Lin, Y.-S. Liu, and H.-S. Huang, "Clinical outcomes of percutaneous transhepatic biliary drainage at different Couinaud's hepatic entry segments for treating obstructive jaundice," *Frontiers Surg.*, vol. 10, Jan. 2023, Art. no. 1039106.

[7] S. Chatelin et al., "Influence of portal vein occlusion on portal flow and liver elasticity in an animal model," *NMR Biomed.*, vol. 34, no. 6, p. e4498, Jun. 2021.

[8] V. Vivaldi, S. Garbarino, G. Caviglia, M. Piana, and G. Sanbuceti, "Compartmental analysis of nuclear imaging data for the quantification of FDG liver metabolism," 2013, *arXiv:1305.7435*.

[9] M.-A. Lebre et al., "Automatic segmentation methods for liver and hepatic vessels from CT and MRI volumes, applied to the Couinaud scheme," *Comput. Biol. Med.*, vol. 110, pp. 42–51, Jul. 2019.

[10] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," 2016, *arXiv:1606.04797*.

[11] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, "NnU-Net: A self-configuring method for deep learning-based biomedical image segmentation," *Nature Methods*, vol. 18, no. 2, pp. 203–211, Feb. 2021.

[12] H. Cao et al., "Swin-Unet: Unet-like pure transformer for medical image segmentation," 2021, *arXiv:2105.05537*.

[13] A. Hatamizadeh et al., "UNETR: Transformers for 3D medical image segmentation," 2021, *arXiv:2103.10504*.

[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* (Lecture Notes in Computer Science), vol. 11861, H.-I. Suk, M. Liu, P. Yan, and C. Lian, Eds. Cham, Switzerland: Springer, 2019, doi: 10.1007/978-3-030-32692-0_32.

[15] X. Han et al., "Automated segmentation of liver segment on portal venous phase MR images using a 3D convolutional neural network," *Insights Imag.*, vol. 13, no. 1, p. 26, Dec. 2022.

[16] J. Jones et al., "Couinaud classification of hepatic segments," *Radiopaedia*, doi: 10.53347/rID-4474.

[17] 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. Med.*, vol. 144, May 2022, Art. no. 105363.

[18] D. Selle, B. Preim, A. Schenk, and H.-O. Peitgen, "Analysis of vasculature for liver surgical planning," *IEEE Trans. Med. Imag.*, vol. 21, no. 11, pp. 1344–1357, Nov. 2002.

[19] X. Yang et al., "Segmentation of liver and vessels from CT images and classification of liver segments for preoperative liver surgical planning in living donor liver transplantation," *Comput. Methods Programs Biomed.*, vol. 158, pp. 41–52, May 2018.

[20] O. I. Alirr and A. A. A. Rahni, "Automatic atlas-based liver segmental anatomy identification for hepatic surgical planning," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 15, no. 2, pp. 239–248, Feb. 2020.

[21] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning dense volumetric segmentation from sparse annotation," 2016, *arXiv:1606.06650*.

[22] 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, 2022.

[23] Z. Arya, G. Ridgway, A. Jandor, and P. Aljabar, "Deep learning-based landmark localisation in the liver for Couinaud segmentation," in *Proc. Annu. Conf. Med. Image Understand. Anal.*, 2021, pp. 227–237.

[24] T. Germain, S. Favelier, J.-P. Cercueil, A. Denys, D. Krausé, and B. Guiu, "Liver segmentation: Practical tips," *Diagnostic Interventional Imag.*, vol. 95, no. 11, pp. 1003–1016, Nov. 2014.

[25] J. Bertels et al., "Optimizing the dice score and Jaccard index for medical image segmentation: Theory & practice," 2019, *arXiv:1911.01685*.

[26] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. J. Cardoso, "Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations," 2017, *arXiv:1707.03237*.

[27] M. J. Cardoso et al., "MONAI: An open-source framework for deep learning in healthcare," 2022, *arXiv:2211.02701*.

[28] I. Loshchilov and F. Hutter, "Fixing weight decay regularization in Adam," *CoRR*, vol. abs/1711.05101, pp. 1–13, Nov. 2017.

[29] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30. Red Hook, NY, USA: Curran Associates, 2017, pp. 1–11.

[30] Z. Liu et al., "Swin transformer: Hierarchical vision transformer using shifted windows," 2021, *arXiv:2103.14030*.

[31] Y. Tang et al., "Self-supervised pre-training of swin transformers for 3D medical image analysis," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 20698–20708.

[32] A. Dosovitskiy et al., "An image is worth 16 × 16 words: Transformers for image recognition at scale," 2020, *arXiv:2010.11929*.

[33] J. Ma, Y. He, F. Li, L. Han, C. You, and B. Wang, "Segment anything in medical images," *Nature Commun.*, vol. 15, no. 1, p. 654, Jan. 2024.