

A Comprehensive CT Dataset for Liver Computer Assisted Diagnosis

Qingsen Yan¹

qingsenyan@gmail.com

Bo Wang²⁵

wang-b17@mails.tsinghua.edu.cn

Dong Gong¹

edgong01@gmail.com

Dingwen Zhang⁴

zhangdingwen2006yyy@gmail.com

Yang Yang⁴

tp030ny@mail.nwpu.edu.cn

Zheng You²

yz-dpi@mail.tsinghua.edu.cn

Yanning Zhang³

ynzhang@nwpu.edu.cn

Javen Qinfeng Shi¹

javen.shi@adelaide.edu.au

¹ The Australian Institute for Machine Learning

The University of Adelaide
South Australia, Australia

² The Innovation Center for Future Chips
Tsinghua University
Beijing, China

³ The School of Computer Science
Northwestern Polytechnical University
Xi'an, China

⁴ The School of Automation
Northwestern Polytechnical University
Xi'an, China

⁵ The Beijing Jingzhen Medical
Technology Ltd.
Beijing, China

Abstract

Automated severity assessment of disease in computed tomography (CT) images plays an essential role for regional assessment of disease that is in great need of liver-related computer assisted diagnosis. An effective dataset is a cornerstone of the deep learning-based method. However, the popular liver-related datasets only focus on specific tasks and fail to guide advanced liver diagnosis tasks. To bridge the gap, we construct the first comprehensive liver dataset, called **ComLiver**, which consists of multiple liver-related tasks and manually marks the elaborate labels for each task (500 cases). Note that the images in the ComLiver dataset are all thin thickness data that have significant clinical implications. In addition, liver and liver lesions segmentations have attracted substantial interest and achieved approving progress, other liver-related tasks of the liver are still under-studied due to various challenges (*e.g.* low contrast, vascular complexity, etc.), despite its significance in assisting preoperative planning. To better exploit the advanced tasks in liver therapy, we introduce vessel instance segmentation, couinaud segmentation tasks based on previous tasks. Finally, we perform a thorough evaluation of the state-of-the-art methods of each task on the proposed ComLiver dataset and obtain a number of interesting findings. Results show its challenging nature, unique attributes and present definite prospects for novel, adaptive, and generalized liver-related segmentation methods. We hope this dataset could advance research towards liver-related computer assisted diagnosis.

⁰The first two authors contributed equally.

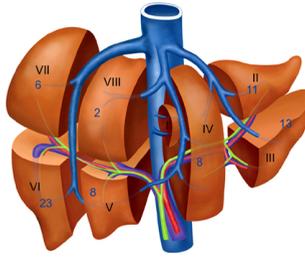


Figure 1: Anatomy of the liver [0].

1 Introduction

The liver is a vital organ with a variety of metabolic functions in the human body. It not only plays an important role in the metabolism of sugar, protein and hormones, but also has crucial functions such as secretion, excretion and biotransformation. When the liver function is abnormal, it can disturb the body’s metabolism and cause diseases. Hepatitis B virus (HBV) is one of the common liver diseases and threatens human health and safety. There are 300 million HBV-infected patients in the world, a large proportion of these infected people die of liver cirrhosis and liver cancer. Surgical resection is currently the preferred choice to treat these diseases. For surgeons, there are three goals that need to be achieved in the liver resection process (See Figure 1), which guarantee the maximum benefit for patients in the process. (i) All the target lesions should be completely removed, (ii) the remaining liver is structurally intact and the functional regions are maximized, (iii) surgical bleeding and systemic trauma invasion are controlled to the greatest extent.

Understanding computed tomography (CT) imaging of the liver diseases will help to detect infections early and assess the disease progression. Especially, automated severity assessment of disease in CT images [27, 81, 63] plays an essential role for liver-related computer assisted diagnosis. Traditional methods often use hand-crafted features to achieve the corresponding tasks, but have huge room for improvement. A recent twist to the low precision problem is useful information extracted by AI technologies, in particular, deep neural networks (DNNs) [24, 28, 29, 61, 62]. At present, most medical image processing tasks use supervised deep learning approaches, an effective dataset is a cornerstone for supervised learning. A number of datasets [0, 9, 10, 13, 15, 21, 23] have been constructed for the liver-related task (*e.g.* liver segmentation, lesion segmentation, vessel segmentation etc.), and on top of which various models [8, 13, 20] have been proposed.

Such popular datasets [0, 10, 15, 23], however, cannot serve as ideal testbeds to evaluate the processing ability of the existing models. First, the existing datasets are typically simple and small, focusing mainly on some single and specific tasks, such as liver segmentation or vessel segmentation. Second, many images in the existing datasets contain only limited information (*e.g.* lower resolution, thick thickness) and cannot meet the demands of the advanced liver-related tasks. The thick layer data can be used for liver and lesions segmentation tasks, but it will generate inaccurate results for vessel or couinaud segmentation. For example, the thick layer data often discards the abundant minor vessels, but the insufficient vascular system is useless for surgical planning. Third, the popular liver-related datasets only focus on the specific tasks and fail to guide advanced liver diagnosis tasks.

To tackle the aforementioned problems, we propose a new challenging dataset that consists of multiple tasks for promoting the development of liver-related computer assisted diagnosis. Our dataset is constructed on real-world images, which consists of 500 cases of CT images and their corresponding ground truth images of the same patients. The key novelty of our dataset lies in new tasks for live remedy, *i.e.* vessel instance segmentation, couinaud segmentation. Different from vessel segmentation, the **vessel instance segmentation** can separate the hepatic (blue color in Figure 1) and portal vein (purple color in Figure 1) to guide surgical planning which is a crucial step in the surgery planning. The **couinaud segmentation** is currently the most widely used system to describe functional liver anatomy, which divides the liver into eight independent functional parts (labeled with I-VIII in Figure 1) allowing resection of segments without damaging other segments [14]. Specifically, we not only annotate the label of the common tasks (*e.g.* liver, tumors, vessels), but also annotate the label of advanced tasks (hepatic vein, portal vein, functional regions) by the medical rules. Along with the dataset, a benchmark is proposed on these tasks, which evaluates several state-of-the-art methods of each task on CT sequences. With the new dataset and tasks, we deem the liver-related computer assisted diagnosis will obtain great progress. The contributions of this paper can be summarized as follows: (1) We build a new dataset on top of real images, named ComLiver, which assembles the several liver-related tasks to promote the development of computer assisted diagnosis. (2) Based on the proposed dataset, we introduce two tasks, *i.e.* vessel instance segmentation, couinaud segmentation and resected parts assessment tasks, which requires a model to segment the hepatic and portal vein, or functional regions. (3) We conduct a comprehensive evaluation on the ComLiver dataset via several state-of-the-art methods, from which we find that the existing methods still leave much room for improvement.

2 Related Work

2.1 Liver Related Tasks

Liver Segmentation. The accurate liver shape measurements from computed tomography (CT) images are the cornerstone of the following tasks. Traditionally, the liver is often delineated by radiologists on a slice-by-slice basis, which is time-consuming. To handle this problem, many methods have been proposed to automatically segment the liver. Considering the relationship between each slice, 3D FCNs are proposed which 2D convolutions in FCN are replaced by 3D convolutions. Dou *et al.* [5] presented a novel 3D deeply supervised network which takes advantage of a fully convolutional architecture and performs efficient end-to-end learning and inference. Tang *et al.* [25] introduced an edge enhanced network for more accurate liver segmentation.

Vessel Segmentation. Liver vessel extraction from CT images is essential in liver surgical planning, *e.g.* path planning and guidance in liver surgery. Without accurate vessel segmentation, the large blood supply vessels cannot be avoided reasonably, which may lead to hemorrhage during the operation and has a greater risk. Kitrungrotsakul *et al.* [14] proposed a CNNs for vessel segmentation from CT volume, which consists of three deep convolutional neural networks to extract features from different planes of CT data. Huang *et al.* [8] designed a loss function based on a variant of the dice coefficient to increase the penalties for misclassified voxels. Yu *et al.* [32] employed the residual module into the 3D-UNet to segment the liver vessel, which enhances the representation for discriminative features.

Table 1: Comparing ComLiver dataset with other liver imaging datasets. H./P. veins denotes the hepatic and portal veins. Resol. denotes resolution, Thick. represents thickness.

Dataset	3DIRCAD [10]	Sliver07 [23]	Image CLEF [10]	LiTs [15]	Jiang [13]	Ouhmich [20]	Decathlon [8]	IRB [10]	CT- ORG [9]	Ours
Liver Cases	20	30	60	201	117	13	201	50	140	500
Vessel Cases	20	0	0	0	0	0	443	50	0	500
Couinaud Cases	0	0	0	0	0	0	0	0	0	500
Tumor Cases	15	0	0	194	117	13	194	0	136	144
Thick.(mm)	1.25- 4.0	1.0- 3.0	0.399- 2.5	0.45- 6.0	1.25- 5.0	0.7- 1.25	2.50- 5.0	2.50- 5.0	0.45- 6.0	0.63- 2.5
Resol.(mm)	0.56- 0.87	0.55- 0.8	0.67- 1.01	0.56- 1.0	0.71- 1.17	0.66- 0.97	- 0.98	0.54- 0.98	0.56- 1.0	0.55- 0.76
Liver	✓	✓	✓	✓	None	None	✓	✓	✓	✓
Vessel	✓	None	None	None	None	None	✓	✓	None	✓
Tumor	✓	None	BBox	✓	✓	✓	✓	None	✓	✓
H./P. veins	None	None	None	None	None	None	None	None	None	✓
Couinaud	None	None	None	None	None	None	None	None	None	✓

Couinaud Segmentation. Liver couinaud segmentation (LCS) is currently the most widely used system to describe functional liver anatomy [10]. The LCS is divided into eight segments. Since each segment is a completely independent unit, each segment can be resected without affecting others. In order to ensure the survival of the remaining liver, the resection must be performed along the surrounding blood vessels of each segment. Oliveira *et al.* [19] proposed a methodology to segment the liver, its vessels and nodules from computer tomography images for surgical planning. Huang *et al.* [9] introduced a method based on 3D thinning, vascular tree pruning, classification and projection.

Lesions Segmentation. Liver cancer is one of the leading causes of cancer death in the world. To assist doctors in hepatocellular carcinoma diagnosis and treatment planning, an accurate and automatic liver and tumor segmentation method is highly demanded in clinical practice. Recently, FCNs including 2D and 3D FCNs, serve as the backbone in volumetric image segmentation. Vorontsov *et al.* [26] evaluated an automated lesion detection method using an FCN that was trained. Dey *et al.* [9] proposed a cascaded system that combines both 2D and 3D convolutional neural networks to effectively segment hepatic lesions.

2.2 Liver Related Datasets

The liver medical image dataset is essential to train a network for evaluating current liver, vessel and liver tumor tasks. A number of public datasets (*i.e.* 3D-IRCADb [10], Sliver07 [23], ImageCLEF [10], LITS [15], Jiang [13], Ouhmich [20], Decathlon [8], IRB [10] and CT-ORG [9]) have been constructed for the liver-related task. The overview of publicly available datasets of medical liver images is reported in Table 1. These datasets partially provide liver contour annotations, liver lesions or liver vessel masks which are labeled by experts. However, none of them can cover all the tasks (*i.e.* liver contour, vessels, lesions, hepatic vein, portal vein, functional regions) in one dataset.

3 The Comprehensive Liver Dataset

The goal of semantic segmentation of the liver is to associate a pixel with a label in the liver without human initialization. The success of semantic segmentation algorithms is con-

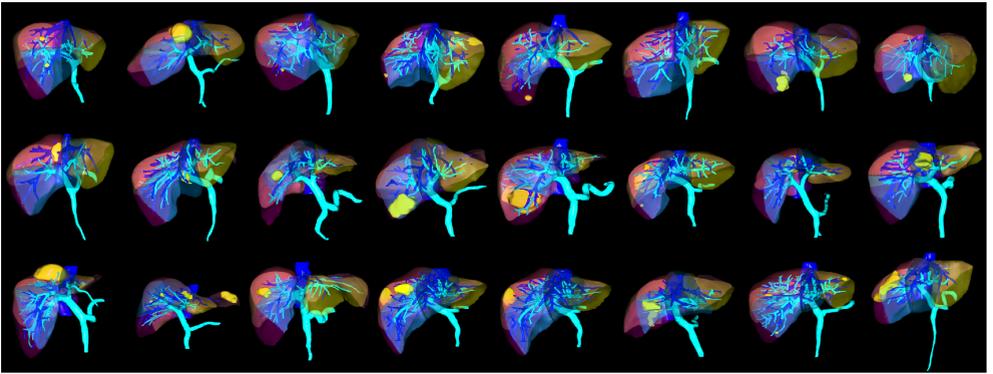


Figure 2: 3D models of liver segments with segmentation annotation in our dataset. Different colors denote the different categories which include vessel, tumors and functional segments.

tingent on the availability of high-quality imaging data with corresponding labels provided by experts. In this section, we will introduce a large collection of annotated liver CT datasets of various clinically relevant anatomies available to facilitate the development of liver-related computer assisted diagnosis.

3.1 Data Collection

This dataset is created in collaboration with six clinical centers and manually blind reviewed by independent twelve radiologists. This dataset is approved by the medical ethics committees of the participating hospitals. In addition, we have deleted the sensitive information for the annotation and sharing of data. 500 portal venous phase CT scans are obtained with 4, 16, and 64 detector rows of different manufacturers. All datasets are acquired contrast-dye-enhanced in the central venous phase. Images are reconstructed at a section thickness that varies from 0.55 to 0.76 mm in x/y-direction, and slice thicknesses vary from 0.625 mm to 2.5 mm. We deem this variety helped to verify the generalizability of the deep learning model. In addition, we removed the personally identifiable information (PII) from all CT scans for patients' privacy. One of the used protocols required patients to lie on their side, *e.g.* the entire anatomy is rotated around the z-axis. Most images in the study were pathologic and included tumors, metastasis and cysts of different sizes.

The proposed dataset has two characteristics: 1) larger and diverse. The amount of data is large compared to other liver released datasets. In addition, it includes diverse shapes and scales of the liver as well as different amounts of vessel branches (See Figure 2). 2) high-quality. As shown in Table 1, the accuracy of resolution and thickness of each case can meet the demand of real-case tasks. For example, the thickness of Decathlon [10] is 2.5-5 mm, which discards too many details of the liver.

3.2 Annotation

To obtain the ground truth of each sample, twelve senior radiologists manually annotate liver contours in a transversal slice-by-slice fashion. Specifically, all samples are pre-labeled using ITKSNAP software to mark the areas of the liver, vessel, and lesions. In case of leakage or inaccurate boundaries, the labels are refined by drawing freehand contours for



Figure 3: Statistics of annotated models.

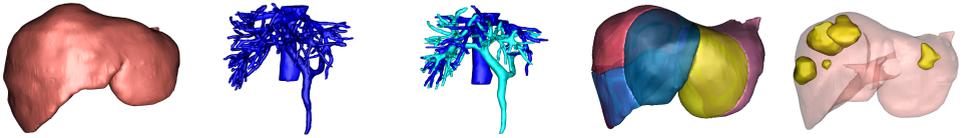


Figure 4: Illustration of the different tasks. H./P. denotes the hepatic and portal veins.

the affected parts. The employed segmentation protocol defined the segmentation mask as the entire liver tissue including all internal structures like vessel systems, tumors, functional parts etc. We divide the senior radiologists team into half and half. One group made all the initial annotations, and another group to check the quality, then the first group revised the results and check it again, until all of them passed the quality test.

Experts collected out datasets instead of regular people. All the 3D models of the liver tissue have to be reconstructed from labeled data manually, as shown in Figure 2. Furthermore, the annotation of the labels of the liver (e.g. contour, vessel, lesions, function parts) requires years of clinical experience in complex situations. The liver-related 3D models have to be restored from the initial label manually. A senior radiologist completed it in 7 hours.

3.3 Statistics of the Dataset

The proposed dataset is composed of the 3D CT scans of 221 females and 279 males. The statistics of our dataset are shown in Figure 3. Figure 3 (a) shows the age distribution of the proposed dataset. Our dataset includes all types of ages in medicine: infant, children, youth, midlife, and elderly people. We also calculate the liver and tumor sizes of samples (Figure 3 (b) and (c)) in the proposed dataset to some extent express the difference of the shapes, since the livers and tumors are mostly unevenly distributed on the size.

4 The Liver Related Tasks

Task 1: Liver Segmentation As shown in Figure 4 (a), the delineation of the liver from abdominal 3D CT images is a crucial prerequisite for computer-assisted liver surgery planning and hepatic disease. If the segmentation can be performed rapidly, the results can also be used in intraoperative guidance [9]. Manual annotation is tedious, error-prone and time-consuming. Automatic liver segmentation from CT volumes is therefore highly demanded. However, automatic and accurate segmentation remains challenging.

Challenges: Complex backgrounds, ambiguous boundaries, heterogeneous appearances and highly varied shapes of the liver.

Task 2: Vessel Segmentation *Vessel Semantic Segmentation* Since the vessels of the liver have been acknowledged as an indispensable element in liver disease diagnosis, the ac-

curate segmentation of the liver vessel tree has become the prerequisite step for automated or computer-aided diagnosis systems. The attributes of the liver blood vessels including length, width, tortuosity, branching pattern, and angles will contribute to the diagnostic result. To assist radiologists with this complex and tedious work, the demand for the fast automated segmentation of the vessel from CT images arises (See Figure 4 (b)).

Challenges: Inconsistent shapes, tiny vessels, low contrast, boundary areas.

Vessel Instance Segmentation Though vessel semantic segmentation provides the information of the venous system, vessel instance segmentation is an effective approach to achieve the couinaud segmentation. The venous system consists of a hepatic and portal vein. The hepatic vein is one of the most important veins that receive blood from the body and transports it into the liver for filtration and processing, and the portal vein drains blood from the gastrointestinal tract and spleen to the liver. However, manual segmentation of hepatic and portal vein (Figure 4 (c)), although possible, is time-consuming and repetitive work, and it also requires professional skills.

Challenges: Complex structure, indistinguishable, tiny vessels, low contrast, boundary areas.

Task 3: Couinaud Segmentation The first step of large liver resections or segmental liver-directed treatments is to effectively achieve liver segmental anatomy. As illustrated in Figure 4 (d), The couinaud segmentation of the liver divides the liver into eight functionally independent segments. Each segment has its own vascular inflow, outflow and biliary drainage. In the center of each segment, there is a branch of the portal vein, hepatic vein. In the periphery of each segment, there is vascular outflow through the hepatic veins.

Challenges: Boundaries of eight independent segments, complexed vasculature, structured relationships.

Task 4: Lesion Segmentation Liver lesions are groups of abnormal cells in the liver and some of them lead to cancer (See Figure 4 (e)). The segmentation of liver tumors in CT is required for assessment of tumor load, treatment planning, prognosis, and monitoring of treatment response.

Challenges: Heterogeneous and diffusive appearance, varying sizes.

5 Experiments

5.1 Experimental Settings

Implementation Details. We select the state-of-the-art methods as the benchmarks of segmentation of our dataset. We implement dataset interfaces to the original implementations by the authors and kept the same hyper-parameters and loss functions of models as in the original papers. We train all the models with Adam optimizer until the accuracy of the validation set stops improving. We estimate these methods by 3-fold cross-validation. The proposed dataset consists of 400 CT scans for training and 100 for testing. However, not all of the samples had lesions in our dataset, there are 110 CT scans for training and 34 for testing on lesions segmentation tasks. In our experiments, all the models are trained on training samples with a size of $128 \times 128 \times 96$. The experiments are carried out on two workstations equipped with 16 NVIDIA TITAN RTX (24G) GPUs. The net training time of all methods is over 16 days.

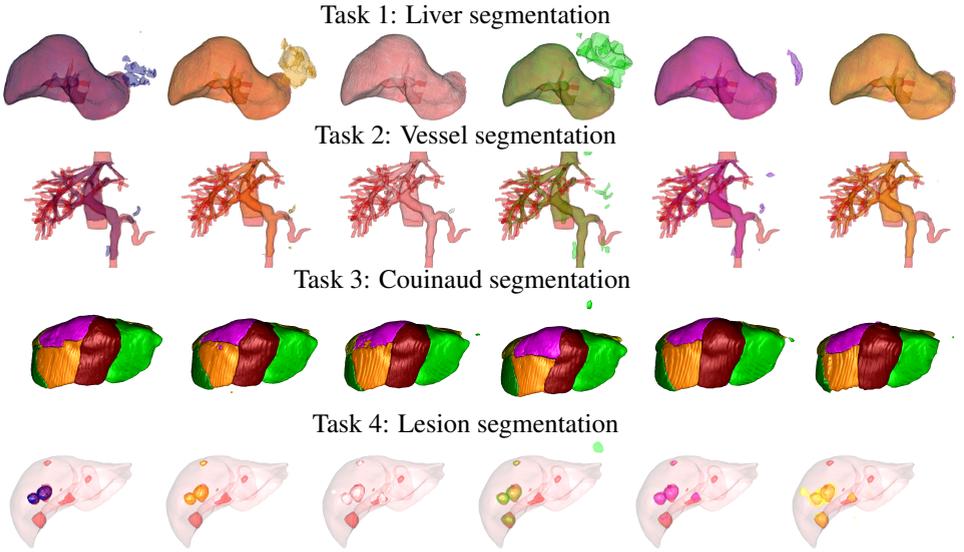
Baselines. Since 2D network methods do not consider any global information – an important cue for CT images, existing segmentation studies based on the 3D network have been

Table 2: Segmentation results of each method on test set.

Tasks	Methods	Dice	Recall	Accuracy	Specificity	Precision
Liver	FCN [16]	0.9602±0.0969	0.9691±0.0707	0.9973±0.0029	0.9982±0.0027	0.9546±0.0994
	UNet [22]	0.9623±0.0968	0.9702±0.0733	0.9975±0.0028	0.9983±0.0028	0.9575±0.0994
	UNet++ [65]	0.9606±0.0959	0.9546±0.0671	0.9974±0.0026	0.9989±0.0024	0.9708±0.0970
	nnUNet [10]	0.9271±0.1123	0.9703±0.0724	0.9940±0.0073	0.9947±0.0076	0.8962±0.1379
	AUNet [18]	0.9678±0.0962	0.9808±0.0680	0.9979±0.0028	0.9983±0.0027	0.9581±0.0969
	PraNet [6]	0.9688±0.0961	0.9757±0.0680	0.9980±0.0025	0.9987±0.0025	0.9650±0.0972
Vessel	FCN [16]	0.8294±0.0528	0.8043±0.0929	0.9936±0.0023	0.9976±0.0011	0.8666±0.0508
	UNet [22]	0.8256±0.0526	0.7895±0.0939	0.9936±0.0022	0.9978±0.0010	0.8775±0.0481
	UNet++ [65]	0.8203±0.0660	0.7961±0.1054	0.9933±0.0025	0.9974±0.0011	0.8598±0.0468
	nnUNet [10]	0.8174±0.0377	0.8400±0.0729	0.9928±0.0021	0.9960±0.0013	0.8038±0.0598
	AUNet [18]	0.8245±0.0556	0.7814±0.0961	0.9935±0.0025	0.9979±0.0008	0.8846±0.0432
	PraNet [6]	0.8251±0.0498	0.8007±0.0891	0.9934±0.0025	0.9974±0.0010	0.8617±0.0506
Hepatic	FCN [16]	0.8083±0.0577	0.7491±0.0991	0.9961±0.0015	0.9990±0.0006	0.8921±0.0579
	UNet [22]	0.8141±0.0555	0.7688±0.0949	0.9961±0.0014	0.9988±0.0006	0.8772±0.0554
	UNet++ [65]	0.7944±0.0585	0.8109±0.0843	0.9953±0.0008	0.9976±0.0007	0.7890±0.0836
	nnUNet [10]	0.8119±0.0508	0.8057±0.0823	0.9959±0.0014	0.9982±0.0008	0.8283±0.0725
	AUNet [18]	0.8014±0.0564	0.7659±0.0907	0.9957±0.0016	0.9985±0.0005	0.8514±0.0635
	PraNet [6]	0.8018±0.0586	0.7830±0.0915	0.9957±0.0016	0.9982±0.0007	0.8322±0.0696
Portal	FCN [16]	0.8230±0.0585	0.7653±0.0989	0.9974±0.0013	0.9993±0.0004	0.9038±0.0471
	UNet [22]	0.8215±0.0589	0.7593±0.0985	0.9973±0.0013	0.9994±0.0004	0.9086±0.0465
	UNet++ [65]	0.7998±0.0642	0.7536±0.1000	0.9976±0.0004	0.9990±0.0004	0.8646±0.0642
	nnUNet [10]	0.8150±0.0576	0.7562±0.0943	0.9972±0.0013	0.9993±0.0004	0.8962±0.0519
	AUNet [18]	0.8055±0.0645	0.7619±0.1015	0.9970±0.0013	0.9990±0.0004	0.8667±0.0601
	PraNet [6]	0.8058±0.0638	0.7905±0.0983	0.9969±0.0013	0.9987±0.0006	0.8323±0.0710
Couinaud	FCN [16]	0.8260±0.0429	0.8302±0.0433	0.9895±0.0033	0.9949±0.0016	0.8413±0.0370
	UNet [22]	0.8380±0.0401	0.8535±0.0382	0.9903±0.0029	0.9949±0.0015	0.8387±0.0373
	UNet++ [65]	0.8269±0.0457	0.8444±0.0414	0.9895±0.0034	0.9945±0.0017	0.8285±0.0420
	nnUNet [10]	0.8170±0.0412	0.8345±0.0388	0.9889±0.0029	0.9942±0.0015	0.8202±0.0377
	AUNet [18]	0.8231±0.0423	0.8516±0.0390	0.9891±0.0031	0.9941±0.0016	0.8167±0.0389
	PraNet [6]	0.8303±0.0428	0.8402±0.0415	0.9898±0.0031	0.9948±0.0015	0.8374±0.0396
Lesions	FCN [16]	0.5227±0.3065	0.4752±0.3018	0.9972±0.0024	0.9993±0.0001	0.8007±0.2544
	UNet [22]	0.6501±0.2375	0.6083±0.2523	0.9971±0.0039	0.9991±0.0018	0.7859±0.2612
	UNet++ [65]	0.4792±0.3154	0.4877±0.3269	0.9948±0.0086	0.9980±0.0025	0.5734±0.3458
	nnUNet [10]	0.5987±0.2878	0.7225±0.2704	0.9962±0.0038	0.9972±0.0036	0.5628±0.3159
	AUNet [18]	0.5277±0.2905	0.5014±0.2987	0.9952±0.0088	0.9991±0.0014	0.6609±0.3152
	PraNet [6]	0.4749±0.2985	0.7628±0.2936	0.9878±0.0187	0.9886±0.0186	0.4022±0.2869

widely studied. We only choose the DNN methods with 3D convolution as our benchmark for references, such as FCN [16], UNet [22], UNet++ [65], nnUNet [10], AUNet (*i.e.* Attention UNet) [18], PraNet [6]. Note that the 2D convolutions of each method are replaced by 3D convolutions. These methods achieve promising performance in medical image segmentation.

Evaluation Metrics. The final results of each task are evaluated based on a sequence of segmentation metrics. They are comprised of Dice score, recall, accuracy, specificity and precision. Dice coefficient is 2 times of the area of overlap divided by the total number of pixels in both images. When applied to a binary segmentation task, it evaluates the degree of overlap between the predicted segmentation mask and the reference segmentation mask. Poor segmentation manifests in a poor overlap as measured by the Dice score. To represent how good a result is at detecting the positives, recall calculates the proportion of positive labels that are correctly identified. To denote how good a result is at avoiding false alarms, specificity tells us the proportion of correctly identified negative labels among all the negative labels. Precision is the proportion of correctly identified positive labels among all the predicted positive labels, which address how many of the positively classified were relevant. Accuracy is the proportion of the correctly predicted labels among all our predictions. The final results are the mean values of each result.



(a) FCN [14] (b) UNet [23] (c) UNet++ [65] (d) nnUNet [14] (e) AUNet [18] (f) PraNet [9]
 Figure 5: The comparisons on the liver-related segmentation tasks. Red color represents the ground truth. The differences with ground truth demonstrate the performance of each method.

5.2 Results and Analysis

5.2.1 Task 1: Liver Segmentation

The 3D results of each method are shown in Figure 5 (first row), to better compare the differences, the ground truth is represented by red color. UNet++ employs the nested and dense skip connections to capture abundant features, PraNet uses a parallel reverse attention network for accurate segmentation. Thus, the visual effects of the results are better than others. FCN, UNet, nnUNet and AUNet use inadequate features, the results often include the failed regions in the background. Table 2 reports the performance of each method, all the methods can achieve approving results. We deem that some existing state-of-the-art methods can solve the major challenges in liver segmentation.

5.2.2 Task 2: Vessel Segmentation

Vessel segmentation includes semantic segmentation and instance segmentation. Due to space constraints, we only show the results of vessel semantic segmentation, more results are shown in supplementary materials. The 3D semantic segmentation results are displayed in Figure 5 (second row), from which we can find that vessel semantic segmentation is a challenging task. All the methods can segment the main branch of vessels, but the minor vessels are neglected. Since FCN tends to use semantic information to segment vessels, the results have the fragmented vessels and fewer small vessels. Although the UNet++ captures more features than UNet, the results have slight differences. It means that we need to design the specific module for this task. Compared with other methods, nnUNet can segment the main branch of vessels. From Table 2, AUNet obtains better performance than other methods. This can be attributed to the attention module, a specific attention module may have

the ability to improve the performance of the vessel segmentation task. As discussed above, vessel semantic segmentation is an important task and has great research value, such as how to accurately segment minor vessels.

5.2.3 Task 3: Couinaud Segmentation

Since the shape of the liver functional affects each other and their borders relate to the appearance of other structures (*e.g.* vessels), we design two settings (*i.e.* liver with/without hepatic and portal veins). Here we only show the results without hepatic and portal veins. Compared with popular liver segmentation datasets, the proposed dataset is the first to provide liver and vasculature labels for couinaud segmentation. As shown in Figure 5 (third row), all the methods can distinguish each part of the liver, but the accuracy of the results is not enough for couinaud segmentation task. For example, the boundary of each part is not clear, the purple parts of the UNet++’s result are confused with the golden yellow parts. Results are shown in Table 2, UNet achieves better results than others. For this task, we consider that such structural relationships play a critical role in the accurate delineation of the couinaud segmentation. The compared method is not designed for the couinaud segmentation task, thus there is much room for improvement.

5.2.4 Task 4: Lesion Segmentation

The visualization and quantification of lesions segmentation are reported in Figure 5 (last row) and Table 2. Compared with others, FCN is the worst one for lesion segmentation, due to the lesions have different scales. UNet or UNet++ employ multi-scale features and fused multi-scale features to predict the lesions, the results can be improved for small lesions. Although nnUNet has the ability to segment more lesions, it tends to generate over-segmented results. The results of AUNet is on-par with PraNet’s results, both of them introduce an attention modules. But AUNet’s results are under-segmented and PraNet’s results are over-segmented. In addition, there still are small lesions which are not segmented using the existing methods.

Based on the results of the above tasks, we can calculate the percentage of the liver after lesion resection, which provides guidance for designing the plan. In addition, we also can calculate the percentage of the lesions in each functional part of the liver.

6 Conclusion

We believe the underlying research progress is mainly obstructed by the lack of an adequately annotated dataset. To this end, we have presented the ComLiver, a new large-scale benchmark for liver-related tasks. It is the largest such dataset to date and presents several real-world challenges that do not present in previous datasets, such as vessel instance segmentation, couinaud segmentation. Based on the ComLiver dataset, we conduct a comprehensive performance evaluation of the current state-of-the-art methods and show that there is still much room for improvement. It is our hope that ComLiver will help to propose technology and promote research to address these new challenges.

Acknowledgements

This work was partially supported by ARC (DP140102270, DP160100703), NSFC (61871328, 61971273, 61901384).

References

- [1] 3Dircadb. <https://www.ircad.fr/research/3d-ircadb-01>.
- [2] CT-ORG. <https://wiki.cancerimagingarchive.net/display/Public/CT-ORG%3A+CT+volumes+with+multiple+organ+segmentations#61080890171ba531fc374829b21d3647e95f532c>.
- [3] MICCAI 2018 Medical Segmentation Decathlon. <http://medicaldecathlon.com/>.
- [4] Raunak Dey and Yi Hong. Hybrid cascaded neural network for liver lesion segmentation. In *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, pages 1173–1177. IEEE, 2020.
- [5] Qi Dou, Hao Chen, Yueming Jin, Lequan Yu, Jing Qin, and Pheng-Ann Heng. 3d deeply supervised network for automatic liver segmentation from ct volumes. In *International conference on medical image computing and computer-assisted intervention*, pages 149–157. Springer, 2016.
- [6] Deng-Ping Fan, Ge-Peng Ji, Tao Zhou, Geng Chen, Huazhu Fu, Jianbing Shen, and Ling Shao. Pranet: Parallel reverse attention network for polyp segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 263–273. Springer, 2020.
- [7] Jean-Francois Gigot, David Glineur, Juan Santiago Azagra, Martine Goergen, Marc Ceuterick, Mario Morino, José Etienne, Jacques Marescaux, Didier Mutter, Ludo van Krunckelsven, et al. Laparoscopic liver resection for malignant liver tumors: preliminary results of a multicenter european study. *Annals of surgery*, 236(1):90, 2002.
- [8] Qing Huang, Jinfeng Sun, Hui Ding, Xiaodong Wang, and Guangzhi Wang. Robust liver vessel extraction using 3D U-Net with variant dice loss function. *Computers in biology and medicine*, 101:153–162, 2018.
- [9] Shao-hui Huang, Bo-liang Wang, Ming Cheng, Wei-li Wu, Xiao-yang Huang, and Ying Ju. A fast method to segment the liver according to couinaud’s classification. In *International Conference on Medical Imaging and Informatics*, pages 270–276. Springer, 2007.
- [10] ImageCLEF. <https://www.imageclef.org/2015/liver>.
- [11] IRB. <https://www.synapse.org/#!/Synapse:syn3193805/wiki/217789>.

- [12] Fabian Isensee, Jens Petersen, Andre Klein, David Zimmerer, Paul F Jaeger, Simon Kohl, Jakob Wasserthal, Gregor Koehler, Tobias Norajitra, Sebastian Wirkert, et al. nnu-net: Self-adapting framework for u-net-based medical image segmentation. *arXiv preprint arXiv:1809.10486*, 2018.
- [13] Huiyan Jiang, Tianyu Shi, Zhiqi Bai, and Liangliang Huang. Ahcnet: An application of attention mechanism and hybrid connection for liver tumor segmentation in ct volumes. *IEEE Access*, 2019.
- [14] Titinunt Kitrungrotsakul, Xian-Hua Han, Yutaro Iwamoto, Amir Hossein Foruzan, Lanfen Lin, and Yen-Wei Chen. Robust hepatic vessel segmentation using multi deep convolution network. In *Medical Imaging 2017: Biomedical Applications in Molecular, Structural, and Functional Imaging*, volume 10137, pages 11–16. International Society for Optics and Photonics, 2017.
- [15] LITS. <https://competitions.codalab.org/competitions/17094>.
- [16] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.
- [17] Michael C Lowe and Michael I D’Angelica. Anatomy of hepatic resectional surgery. *Surgical Clinics*, 96(2):183–195, 2016.
- [18] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, et al. Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*, 2018.
- [19] Dário AB Oliveira, Raul Q Feitosa, and Mauro M Correia. Segmentation of liver, its vessels and lesions from ct images for surgical planning. *Biomedical engineering online*, 10(1):1–23, 2011.
- [20] Farid Ouhmich, Vincent Agnus, Vincent Noblet, Fabrice Heitz, and Patrick Pessaux. Liver tissue segmentation in multiphase ct scans using cascaded convolutional neural networks. *International Journal of Computer Assisted Radiology and Surgery*, 2019.
- [21] Shima Rafiei, Ebrahim Nasr-Esfahani, Kayvan Najarian, Nader Karimi, and S.M. Reza Soroushmehr. Liver segmentation in ct images using three dimensional to two dimensional fully convolutional network. In *25th IEEE International Conference on Image Processing (ICIP)*, 2018.
- [22] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention*, pages 234–241. Springer International Publishing, 2015.
- [23] Sliver07. <http://www.sliver07.org>.
- [24] Shaolin Su, Qingsen Yan, Yu Zhu, Cheng Zhang, Xin Ge, Jinqiu Sun, and Yanning Zhang. Blindly assess image quality in the wild guided by a self-adaptive hyper network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3667–3676, 2020.

- [25] Youbao Tang, Yuxing Tang, Yingying Zhu, Jing Xiao, and Ronald M Summers. E2net: An edge enhanced network for accurate liver and tumor segmentation on ct scans. *arXiv preprint arXiv:2007.09791*, 2020.
- [26] Eugene Vorontsov, Milena Cerny, Philippe Régnier, Lisa Di Jorio, Christopher J Pal, Réal Lapointe, Franck Vandenbroucke-Menu, Simon Turcotte, Samuel Kadoury, and An Tang. Deep learning for automated segmentation of liver lesions at ct in patients with colorectal cancer liver metastases. *Radiology: Artificial Intelligence*, 1(2):180014, 2019.
- [27] Bo Wang, Shuo Jin, Qingsen Yan, Haibo Xu, Chuan Luo, Lai Wei, Wei Zhao, Xuexue Hou, Wenshuo Ma, Zhengqing Xu, et al. Ai-assisted ct imaging analysis for covid-19 screening: Building and deploying a medical ai system. *Applied Soft Computing*, 98: 106897, 2021.
- [28] Qingsen Yan, Dong Gong, Qinfeng Shi, Anton van den Hengel, Chunhua Shen, Ian Reid, and Yanning Zhang. Dual-attention-guided network for ghost-free high dynamic range imaging. *International Journal of Computer Vision*. doi: 10.1007/s11263-021-01535-y.
- [29] Qingsen Yan, Dong Gong, and Yanning Zhang. Two-stream convolutional networks for blind image quality assessment. *IEEE Transactions on Image Processing*, 28(5): 2200–2211, 2018.
- [30] Qingsen Yan, Dong Gong, Qinfeng Shi, Anton van den Hengel, Chunhua Shen, Ian Reid, and Yanning Zhang. Attention-guided network for ghost-free high dynamic range imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1751–1760, 2019.
- [31] Qingsen Yan, Bo Wang, Wei Zhang, Chuan Luo, Wei Xu, Zhengqing Xu, Yanning Zhang, Qinfeng Shi, Liang Zhang, and Zheng You. An attention-guided deep neural network with multi-scale feature fusion for liver vessel segmentation. *IEEE Journal of Biomedical and Health Informatics*, 2020.
- [32] Qingsen Yan, Lei Zhang, Yu Liu, Yu Zhu, Jinqiu Sun, Qinfeng Shi, and Yanning Zhang. Deep hdr imaging via a non-local network. *IEEE Transactions on Image Processing*, 29:4308–4322, 2020.
- [33] Qingsen Yan, Bo Wang, Dong Gong, Chuan Luo, Wei Zhao, Jianhu Shen, Jingyang Ai, Qinfeng Shi, Yanning Zhang, Shuo Jin, et al. Covid-19 chest ct image segmentation network by multi-scale fusion and enhancement operations. *IEEE Transactions on Big Data*, 7(1):13–24, 2021.
- [34] Wei Yu, Bin Fang, Yongqing Liu, Mingqi Gao, Shenhai Zheng, and Yi Wang. Liver vessels segmentation based on 3d residual u-net. In *IEEE International Conference on Image Processing (ICIP)*, pages 250–254. IEEE, 2019.
- [35] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: A nested u-net architecture for medical image segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pages 3–11. Springer, 2018.