Anatomical Prior-Inspired Label Refinement for Weakly Supervised Liver Tumor Segmentation with Volume-Level Labels

Fei Lyu[1]
feilyu@comp.hkbu.edu.hk
Andy J. Ma[2]
majh8@mail.sysu.edu.cn
Pong C. Yuen[1]
pcyuen@comp.hkbu.edu.hk

[1] Department of Computer Science, Hong Kong Baptist University
[2] School of Computer Science and Engineering, Sun Yat-sen University

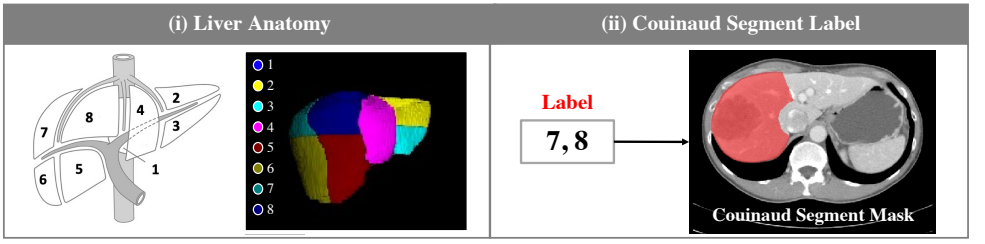
Abstract

Automatic liver tumor segmentation is important for assisting doctors in accurate diagnosis of liver cancer. Existing models for liver tumor segmentation usually require accurate pixel-level labels. However, acquiring such dense labels is laborious and costly. In this paper, we propose a weakly supervised method for liver tumor segmentation using volume-level labels, which can significantly reduce the manual annotation effort. Volume-level labels are propagated to image-level labels where all the slices in one CT volume share the same label, and pixel-level pseudo labels can be estimated from image-level labels. However, it will cause the label noise problem because not all slices contain tumors. To address this issue, we propose two label refinement strategies based on anatomical priors to reduce the training noise and improve the model performance. Evaluation experiments on two public datasets demonstrate that our proposed method can achieve competitive results compared to other methods with stronger supervision.

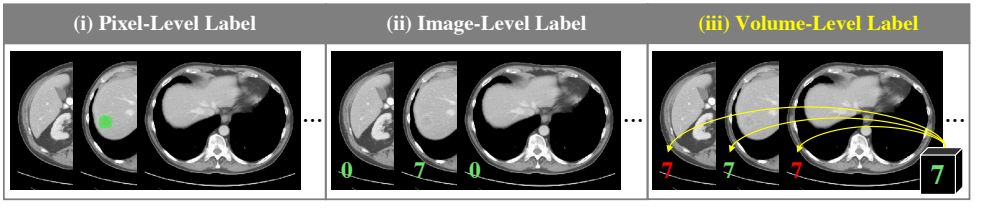
1 Introduction

Liver cancer is one of the leading cause of death globally, while early diagnosis can facilitate appropriate surveillance strategy and reduce cancer mortality [9, 13, 14, 20]. Automatic liver tumor segmentation is of great importance for assisting doctors in liver cancer diagnosis [7, 15, 17, 18, 26]. Deep learning based methods have been proposed for automatic liver tumor segmentation, which typically require pixel-level labels. However, obtaining such high-quality labels is extremely laborious and costly, which makes many successful deep learning techniques inapplicable.

Weak labels have been widely explored for medical image segmentation [1, 4, 6, 8, 12, 21, 22, 27]. As for liver tumor segmentation, using image-level labels based on Couinaud segment is proposed in [10], where Couinaud scheme is the widely used system in clinical practice which divides the liver into eight segments based on vascular structure [5, 11, 19, 25]. Couinaud segment is often used by radiologists to describe the localization



(a) Couinaud segment



(b) Comparing strong and weak labels for liver tumor segmentation

Figure 1: (a) Couinaud segment: i) The Couinaud classification of liver anatomy divides the liver into eight functionally independent segments based on its vascular structure. Automatic methods have been successfully applied to construct the Couinaud scheme from CT volumes without manual efforts. ii) The Couinaud segment label indicates a specific region of the liver, and can generate the Couinaud segment mask using existing automatic methods [10]. (b) Comparing strong and weak labels for liver tumor segmentation: i) Fully supervised pixel-level labels. ii) Weakly supervised image-level labels based on Couinaud segment. iii) Weakly supervised volume-level labels studied in our method, which can be treated as noisy image-level labels. Green: Correct labels. Red: Incorrect labels. Best viewed in color.

of liver tumors when they record their findings in radiology reports. For example, 'There is a new 13mm tumor with arterial enhancement in **segment VI** suspicious of recurrence'. Automatic methods have been successfully applied to construct the Couinaud scheme from CT volumes without manual efforts, as shown in Figure 1 (a)(i). Even Couinaud segment label is an image-level label, it indicates a specific region of the liver, and the Couinaud segment mask can be derived accordingly, as shown in Figure 1 (a)(ii). Inspired by [10], we are concerned with a more challenging weakly supervised segmentation setting, in which only volume-level labels are provided. Such a setting has great potential for practical applications since it requires significantly less annotation effort. Some pioneer works have investigated weakly supervised learning with volume-level labels. Most of these methods are designed for classification tasks [2, 23]. As for the segmentation task, a small amount of pixel-level labels is still required [24].

In this paper, we start the first attempt to utilize volume-level labels to train deep models for liver tumor segmentation. Comparison of different labels is shown in Figure 1 (b). Volume-level labels can be propagated to image-level labels and all the slices in one CT volume share the same label, but it will cause the label noise problem since not all slices contain tumors. To address this issue, we propose two label refinement strategies based on anatomical priors. We notice several distinct anatomical priors, *i.e.*, 1) All eight Couinaud segments of the liver can not exist on the same CT slice simultaneously; 2) Tumors should be continuous across adjacent slices. To incorporate these anatomical priors into our method,

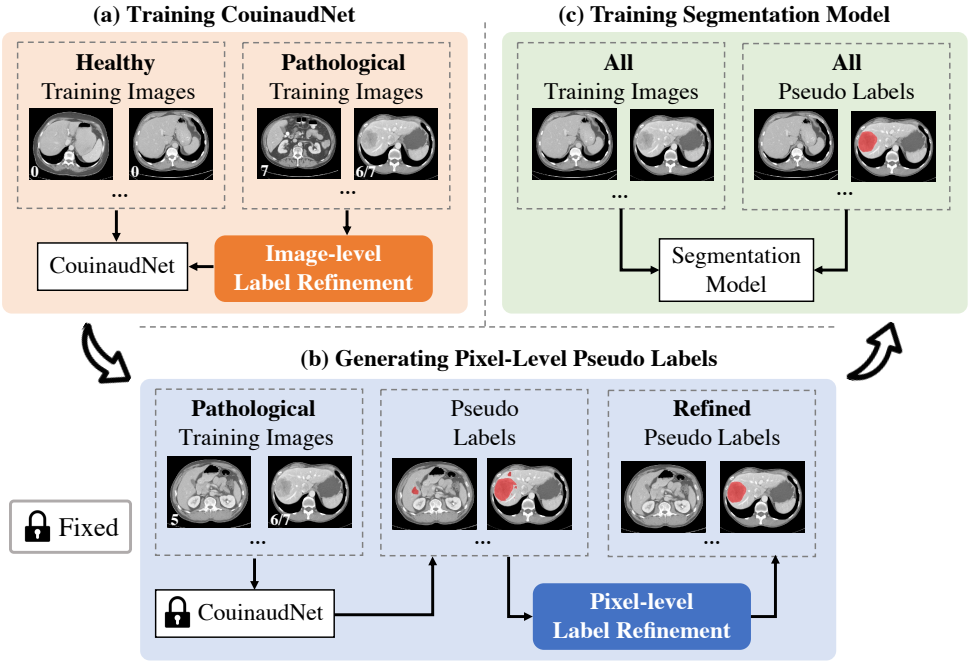


Figure 2: Overview of our proposed method for liver tumor segmentation with volume-level labels. Volume-level labels are treated as noisy image-level labels, and two label refinement strategies based on anatomical priors are proposed to reduce the label noise.

we introduce image-level label refinement (ILR) and pixel-level label refinement (PLR) to reduce the training noise. First, we identify CT slices with incorrect image-level labels and rectify their pathological labels (1∼8) to healthy labels (0). Second, we propose a continuity index to re-weight the pixel-level prediction probability, aiming to reduce the false positives in the pixel-level pseudo labels. We evaluate our method on two public datasets for liver tumor segmentation, which only provide the volume-level labels for training. The experiment results demonstrate that our method achieves competitive results while requiring significantly less annotation effort.

Our main contributions are summarized as follows: (1) We propose to train deep neural networks for liver tumor segmentation using volume-level labels, which significantly reduces human annotation effort. (2) We treat the volume-level labels as noisy image-level labels, and propose two label refinement strategies based on anatomical priors to reduce the training noise and improve the performance. (3) The experiment results quantitatively demonstrate the effectiveness of our proposed method.

2 Method

Figure 2 illustrates the proposed method for liver tumor segmentation with volume-level labels. Our method roots in CouinaudNet which can estimate pixel-level pseudo labels from image-level Couinaud segment labels [11], and proposes two label refinement strategies

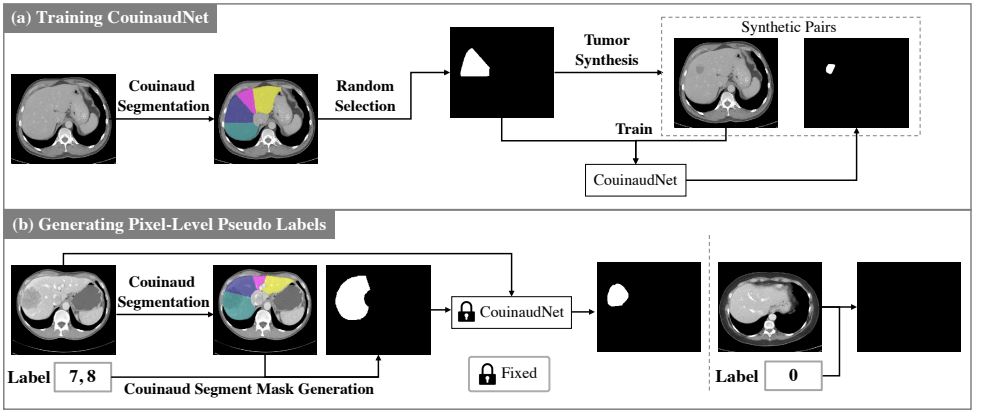


Figure 3: Illustration of CouinaudNet. (a) Training CouinaudNet. CouinaudNet is first trained only with healthy training images. A Couinaud segment is randomly selected to generate the Couinaud segment mask, and a synthesized tumor is injected to the healthy image in the selected Couinaud segment region. The synthetic image and the derived Couinaud segment mask are used as inputs to CouinaudNet. (b) Generating Pixel-Level Pseudo Labels. The trained CouinaudNet is used to generate pseudo tumor masks for the pathological training image (label: 1∼8). A healthy training image (label: 0) has no tumors, and its tumor mask is denoted as a black mask. In this way, pixel-level pseudo labels are available for all training images.

based on anatomical priors to reduce label noise.

We first briefly illustrate CouinaudNet in Figure 3. Image-level labels are required for CouinaudNet, which can be divided into two parts: healthy training images (label: 0) and pathological training image (label: 1∼8). During training, only healthy training images are used and synthetic tumors are injected to a random Couinaud segment. The synthetic image and the derived Couinaud segment mask are used as inputs to CouinaudNet, and the output is the liver tumor mask. With the trained CouinaudNet, pseudo tumor masks are estimated for those pathological training images. Healthy training images have no tumors, and their tumor masks are denoted as black masks.

2.1 Liver Tumor Segmentation with Volume-level Labels

All the slices in one CT volume share the same image-level labels based on the provided volume-level labels, such that volume-level labels can be treated as noisy image-level labels. CouinaudNet is proposed to train deep neural networks with image-level labels based on Couinaud scheme in [10], and we adopt it as the baseline model.

Training CouinaudNet. Only healthy training images are used for training CouinaudNet, which aims at learning a deep model for estimating pixel-level pseudo labels from image-level labels. Pathological training images are not used for training CouinaudNet, but part of them with incorrect labels can be added to original healthy training images for training CouinaudNet after image-level label refinement .

Generating Pixel-level Pseudo Labels. $\mathcal{D}_h = \{x_{h_i}, m_{h_i}\}_{i=1}^{M}$ denotes the set of healthy training images and their pseudo labels, and the value of their pseudo labels is always zero

since no tumors exist. CouinaudNet has been well trained with healthy training images, and it can be used to estimate pixel-level pseudo labels for pathological training images. The set of pathological training images and their pixel-level pseudo labels is described as $\mathcal{D}_t = \{x_{t_i}, m_{t_i}\}_{i=1}^{N}$. These pseudo labels can be further improved after pixel-level label refinement.

Training Segmentation Model. Provided all the training images and their corresponding pixel-level pseudo labels $\mathcal{D} = \{\mathcal{D}_t, \mathcal{D}_h\}$, we can train a fully supervised liver tumor segmentation model f^s. f^s predicts a tumor segmentation map p_i for training image x_i, and we employ cross entropy loss to train the model, which is defined as follows:

$$\mathcal{L}_{seg} = -\sum_i m_i \cdot \log(f^s(x_i)), \tag{1}$$

where m_i is pixel-level pseudo label for x_i. Various deep models have been proposed for liver tumor segmentation with pixel-level supervision, and we adopt nnU-Net [3] which has achieved impressive performance in various biomedical image segmentation challenges.

2.2 Image-level Label Refinement

Volume-level labels are propagated to image-level labels, and training images are labeled as healthy only when no tumors exist in the whole CT volume. In this way, only a small amount of training images are healthy, while a large amount of training images are falsely labeled as pathological. To this end, we aim to filter out these training images with incorrect labels and rectify their pathological labels to healthy labels. Motivated by the observation that all eight Couinaud segments of the liver can not exist on the same CT slice simultaneously, we can identify CT slices that do not contain the Couinaud segments indicated in the pathological labels, and rectify their labels.

Given an input image x_i, its Couinaud scheme can be obtained automatically from the Couinaud constructor in [10], where each pixel has value $pv \in S, S \subset \{1,2...,8\}$. y_j is the image-level label derived from the volume-level label for input image x_j, and the refined label $\hat{y}_j$ is defined as follows:

$$\hat{y}_j = \begin{cases} 0, & \text{if } y_j \notin S \\ y_j, & otherwise \end{cases} \tag{2}$$

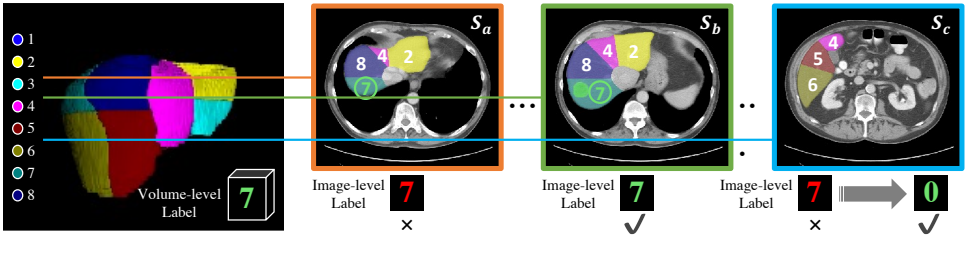


Figure 4: Illustration of the proposed image-level label refinement strategy. S_a, S_b and S_c share the same label, but S_a and S_c are falsely labeled as pathological because they do not have tumors. After image-level label refinement, the image-level label of S_c is rectified to the correct label. Best viewed in color.

Figure 3 illustrates the proposed image-level label refinement strategy. Three slices from the CT volume are shown in the figure, and they share the same image-level label as the volume-level label. However, S_a and S_c are falsely labeled as pathological since they do not have tumors. S_c contains Couinaud segment $S = \{4,5,6\}$, and its label is $y_c = 7$. Considering $y_c \notin S$, its label can be adjusted to $\hat{y}_c = 0$. In this way, some CT slices with incorrect labels can be rectified.

2.3 Pixel-level Label Refinement

After image-level label refinement, labels of some training images can be corrected. However, a significant number of images still suffer from incorrect labels. In this regard, we propose a continuity index to re-weight the pixel-level prediction probability, aiming to reduce the false positives in the pseudo labels generated from the trained CouinaudNet. Considering that tumors should be continuous across adjacent slices, our idea is to reduce the prediction probability of those isolated tumors.

Given the CouinaudNet f^c trained using healthy training images, each pathological training image x_k can obtain the predictive output o_k, *i.e.*, $o_k = f^c(x_k)$. The pseudo tumor mask m_k can be generated from o_k after binarization, which is denoted as:

$$m_k = \mathbb{1}(o_k \geq \tau), \quad \tau \in (0,1) \tag{3}$$

where $\mathbb{1}(.)$ is the indicator function, and τ is the threshold for binarization.

Given binary tumor masks A and B, the Dice score can be calculated as:

$$DICE(A,B) = \frac{2|A \cap B|}{|A| + |B|} \tag{4}$$

We define the continuity index CI_k of x_k as follows:

$$CI_k = c \cdot max\{DICE(m_k, m_{k-1}), DICE(m_k, m_{k+1})\} \tag{5}$$

where m_{k-1} and m_{k+1} are the pseudo tumor masks of neighbouring slices, and c is the scaling coefficient for calculating the continuity index.

After CI_k is obtained for x_k, its refined pixel-level pseudo label $\hat{m}_k$ can be generated as follows:

$$\hat{m}_k = \mathbb{1}((o_k \cdot CI_k) \geq \tau), \quad \tau \in (0,1) \tag{6}$$

where CI_k is used to re-weight the pixel-level prediction probability. Isolated prediction results are unlikely to be true tumors, and their probability is reduced after re-weighting.

3 Experiments

3.1 Experiment Setup

3.1.1 Dataset and Evaluation Metrics.

We evaluate our method on two public datasets for liver tumor segmentation. The 3DIRCADb dataset [16], which contains 15 CT volumes with liver tumors and 5 CT volumes without tumors. Among 15 CT volumes with tumors, 10 of them provide the location of tumors based on Couinaud scheme. 10 CT volumes with location labels and 5 CT volumes without

ILR	PLR	Dice per case(%)	Dice global(%)
		5.9	8.1
✓		21.9	17.9
	✓	12.7	17.2
✓	✓	35.6	32.2

Table 1: Evaluation of each component on 3DIRCADb test set. (**ILR**: image-level label refinement; **PLR**: pixel-level label refinement.)

tumors are used as training datasets in our experiment, while the left 5 CT volumes with tumors are used as test datasets. Details of the dataset are provided in the Supplementary Materials. Another dataset is Medical Segmentation Decathlon Task08 (MSD08), where the volume-level labels are automatically generated based on the overlap between tumor masks and groundtruth Couinaud scheme. Following [10], we use 161 volumes for training, and 142 volumes for testing. For evaluation, we employ two commonly used metrics for liver tumor segmentation, including Dice per case score and Dice global score. The Dice per case score is the average score across all cases, while the Dice global score combines all cases into one case to obtain the global dice score.

3.1.2 Implementation Details

In the pre-processing step, we rescale the CT intensity range to [0,255] using the CT window [-200, 250]. Similar to [10], we only keep the liver region and remove irrelevant information to reduce computation cost. Unless otherwise notes, all hyperparameters are as in [10] for CouinaudNet. Provided all training images and their tumor masks, *i.e.*, healthy training images with healthy masks (black masks with no tumors) and pathological training images with estimated pseudo tumor masks using the trained CouinaudNet, we can train a fully supervised tumor segmentation model implemented by nnU-net [3]. The threshold τ for binarization is set to 0.5, and the scaling coefficient c for calculating the continuity index is set to 5 in our method.

3.2 Analysis of Our Method

3.2.1 Image-level Label Refinement.

To validate the effectiveness of image-level label refinement (ILR), we analyze the performance of the image-level labels derived from volume-level labels on 3DIRCADb. As shown in Figure 5(a), a large amount of training images without tumors are falsely labeled as pathological before ILR. After ILR, the number of training images with incorrect labels is reduced from 854 to 465, which demonstrates that ILR can effectively improve the quality of the propagated image-level labels. In this way, more healthy images can be used to train CouinaudNet, and a better trained CouinaudNet can estimate pixel-level pseudo labels for pathological training images with higher quality.

3.2.2 Pixel-level Label Refinement.

To analyze the contribution of pixel-level label refinement (PLR), we calculate the Dice scores between the generated pseudo labels and manually annotated labels before and after

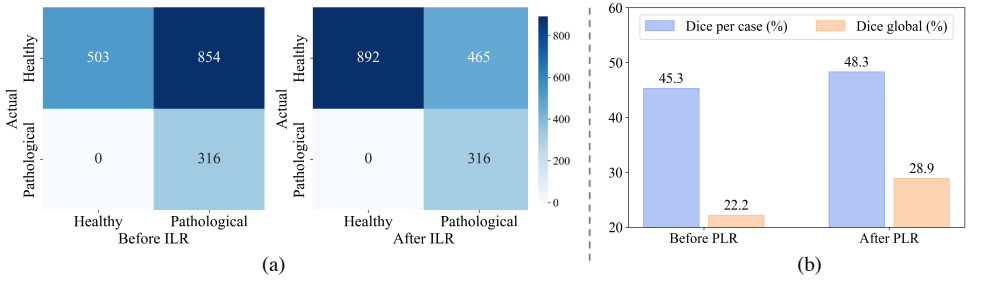


Figure 5: (a) Confusion matrix of the image-level labels derived from volume-level labels before and after image-level label refinement (ILR) on 3DIRCADb. (b) The Dice scores between the generated pixel-level pseudo labels and manually annotated labels before and after pixel-level label refinement (PLR) on 3DIRCADb.

Supervision	Methods	Dice per case(%)	Dice global(%)
Pixel-level	nnU-Net [3]	49.8	48.9
Image-level	CouinaudNet [10]	34.3	33.3
Volume-level	CouinaudNet [10]	5.9	8.1
	Ours	35.6	32.2

Table 2: Comparison of liver tumor segmentation results on 3DIRCADb test set.

PLR on 3DIRCADb. Figure 5(b) shows that the quality of pseudo labels shows an obvious improvement after PLR. We also try to apply PLR as a post-processing step, but we find the performance is not ideal. The possible reason is that errors caused by PLR would directly impact the performance in post-processing, while pseudo labels have a certain tolerance of such errors during model training.

3.2.3 Ablation Study.

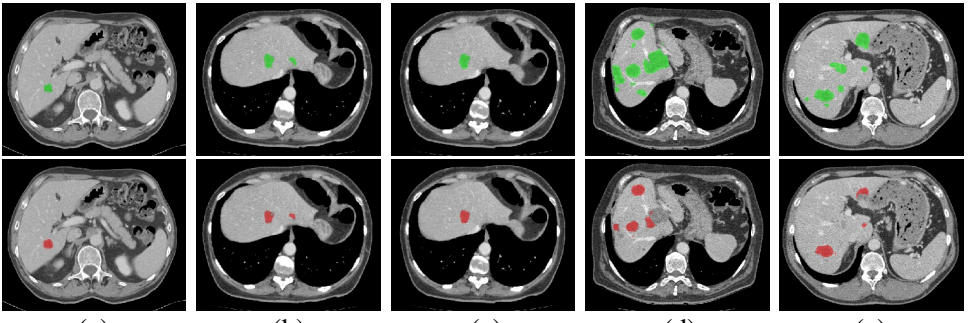
We conduct ablation study to evaluate the two label refinement strategies. As shown in Table 1, it shows poor performance of liver tumor segmentation without any label refinement. After adding ILR, the dice per case score is increased by 16.0% and the dice global score is increased by 9.8%. When further combined with PLR, the dice per case score is increased to 35.6% and the dice global score is increased to 32.2%. Therefore, we can conclude that both label refinement strategies cam improve the tumor segmentation performance over the vanilla CouinaudNet with volume-level labels.

3.3 Comparison with State-of-the-Arts

Table 2 displays the liver tumor segmentation results with different levels of supervision on 3DIRCADb test set. nnU-Net [3] trained with pixel-level labels can be regarded as the upper bound, but both the dice per case score and the dice global score are lower than 50%. The average number of tumors in the training dataset is 2, but the numbers of tumors in the test dataset are 7, 20, 8, 20 and 46 respectively, **which makes it a challenging task essentially**. Using image-level labels shows degraded performance since its supervision is weakened.

Supervision	Methods	Dice per case(%)	Dice global(%)
Pixel-level	nnU-Net [3]	70.5	86.7
Image-level	CouinaudNet [10]	62.2	74.0
Volume-level	CouinaudNet [10]	23.1	28.5
	Ours	58.9	72.3

Table 3: Comparison of liver tumor segmentation results on MSD08 test set.



(a) (b) (c) (d) (e)

Figure 6: Examples of liver tumor segmentation results by models trained with volume-level labels using our proposed method on the 3DIRCADb test dataset. The green ones denote the ground truths, while the red ones denote the segmentation results using our method. Best viewed in color.

Directly applying CouinaudNet for training deep neural networks with volume-level images performs poorly, while our method with two label refinement strategies shows obvious improvement over the vanilla CouinaudNet and can achieve comparable or even better performance than the method using image-level labels, which demonstrates the effectiveness of our proposed method.

3DIRCADb faces a great challenge due to its small dataset size, and the task difficulty is also an important factor for affecting the results. To further validate the effectiveness of our method, we also conduct experiments on the MSD08 dataset, as shown in Table 3. MSD08 has a larger number of test samples than 3DIRCADb. Based on the experiment results, we can get the same conclusion that our method can achieve competitive results compared to other methods with stronger supervision.

Figure 6 shows some examples of liver tumor segmentation results by models trained with volume-level labels using our proposed method on the 3DIRCADb test dataset. We can observe that tumors can be well segmented in Figure 5 (a-c). Our method also has some limitations, and we show some failure cases in Figure 5 (d-e). We can observe that the segmentation performance is not good when segmenting tumors with complicated textures. This is a common problem in medical image segmentation, especially when the supervision information of volume-level labels is substantially weakened. We would like to explore solutions to improve the performance for these hard cases in the future work.

4 Conclusion

In this paper, we present a weakly supervised method to train deep neural networks for liver tumor segmentation with volume-level labels. We treat volume-level labels as noisy image-level labels, and propose two label refinement strategies based on anatomical priors to reduce the label noise. Our method achieves impressive results compared to competing methods but requires less annotation effort. The volume-level label studied in our method is often used by radiologists for recording liver pathologies in radiology reports, and our method provides a promising solution for fully leveraging the wealth of plenty radiology reports which are stored in hospital archiving and communication systems.

References

[1] Reuben Dorent, Samuel Joutard, Jonathan Shapey, Aaron Kujawa, Marc Modat, Sébastien Ourselin, and Tom Vercauteren. Inter extreme points geodesics for end-to-end weakly supervised image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 615–624. Springer, 2021.

[2] Junlin Hou, Jilan Xu, Rui Feng, Yuejie Zhang, Fei Shan, and Weiya Shi. Cmc-cov19d: Contrastive mixup classification for covid-19 diagnosis. In *International Conference on Computer Vision*, pages 454–461, 2021.

[3] Fabian Isensee, Paul F Jaeger, Simon AA Kohl, Jens Petersen, and Klaus H Maier-Hein. nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*, 18(2):203–211, 2021.

[4] Zhanghexuan Ji, Yan Shen, Chunwei Ma, and Mingchen Gao. Scribble-based hierarchical weakly supervised learning for brain tumor segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 175–183. Springer, 2019.

[5] Marie-Ange Lebre, Antoine Vacavant, Manuel Grand-Brochier, Hugo Rositi, Armand Abergel, Pascal Chabrot, and Benoit Magnin. Automatic segmentation methods for liver and hepatic vessels from ct and mri volumes, applied to the couinaud scheme. *Computers in Biology and Medicine*, 110:42–51, 2019.

[6] Marvin Lerousseau, Maria Vakalopoulou, Marion Classe, Julien Adam, Enzo Battistella, Alexandre Carré, Théo Estienne, Théophraste Henry, Eric Deutsch, and Nikos Paragios. Weakly supervised multiple instance learning histopathological tumor segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 470–479. Springer, 2020.

[7] Xiaomeng Li, Hao Chen, Xiaojuan Qi, Qi Dou, Chi-Wing Fu, and Pheng-Ann Heng. H-denseunet: hybrid densely connected unet for liver and tumor segmentation from ct volumes. *IEEE Transactions on Medical Imaging*, 37(12):2663–2674, 2018.

[8] Xiaoming Liu, Quan Yuan, Yaozong Gao, Kelei He, Shuo Wang, Xiao Tang, Jinshan Tang, and Dinggang Shen. Weakly supervised segmentation of covid19 infection with scribble annotation on ct images. *Pattern Recognition*, 122:108341, 2022.

[9] Zhenqiu Liu, Yanfeng Jiang, Huangbo Yuan, Qiwen Fang, Ning Cai, Chen Suo, Li Jin, Tiejun Zhang, and Xingdong Chen. The trends in incidence of primary liver cancer caused by specific etiologies: results from the global burden of disease study 2016 and implications for liver cancer prevention. *Journal of hepatology*, 70(4):674–683, 2019.

[10] Fei Lyu, Andy J. Ma, Terry Cheuk-Fung Yip, Grace Lai-Hung Wong, and Pong C. Yuen. Weakly supervised liver tumor segmentation using couinaud segment annotation. *IEEE Transactions on Medical Imaging*, 41(5):1138–1149, 2022. doi: 10.1109/TMI.2021.3132905.

[11] Pietro Majno, Gilles Mentha, Christian Toso, Philippe Morel, Heinz O Peitgen, and Jean HD Fasel. Anatomy of the liver: an outline with three levels of complexity–a further step towards tailored territorial liver resections. *Journal of Hepatology*, 60(3):654–662, 2014.

[12] Hui Qu, Pengxiang Wu, Qiaoying Huang, Jingru Yi, Zhennan Yan, Kang Li, Gregory M Riedlinger, Subhajyoti De, Shaoting Zhang, and Dimitris N Metaxas. Weakly supervised deep nuclei segmentation using partial points annotation in histopathology images. *IEEE Transactions on Medical Imaging*, 39(11):3655–3666, 2020.

[13] Gregory A Roth, Degu Abate, Kalkidan Hassen Abate, Solomon M Abay, Cristiana Abbafati, Nooshin Abbasi, Hedayat Abbastabar, Foad Abd-Allah, Jemal Abdela, Ahmed Abdelalim, et al. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: a systematic analysis for the global burden of disease study 2017. *Lancet*, 392(10159):1736–1788, 2018.

[14] Shiv K Sarin, Manoj Kumar, Mohammed Eslam, Jacob George, Mamun Al Mahtab, Sheikh M Fazle Akbar, Jidong Jia, Qiuju Tian, Rakesh Aggarwal, David H Muljono, et al. Liver diseases in the asia-pacific region: a lancet gastroenterology & hepatology commission. *The lancet Gastroenterology & hepatology*, 5(2):167–228, 2020.

[15] Hyunseok Seo, Charles Huang, Maxime Bassenne, Ruoxiu Xiao, and Lei Xing. Modified u-net (mu-net) with incorporation of object-dependent high level features for improved liver and liver-tumor segmentation in ct images. *IEEE Transactions on Medical Imaging*, 39(5):1316–1325, 2019.

[16] L Soler, A Hostettler, V Agnus, A Charnoz, J Fasquel, J Moreau, A Osswald, M Bouhadjar, and J Marescaux. 3d image reconstruction for comparison of algorithm database: A patient specific anatomical and medical image database. *IRCAD, Strasbourg, France, Tech. Rep*, 2010.

[17] Lei Song, Haoqian Wang, and Z Jane Wang. Bridging the gap between 2d and 3d contexts in ct volume for liver and tumor segmentation. *IEEE Journal of Biomedical and Health Informatics*, 2021.

[18] Youbao Tang, Yuxing Tang, Yingying Zhu, Jing Xiao, and Ronald M Summers. E^2net: An edge enhanced network for accurate liver and tumor segmentation on ct scans. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 512–522. Springer, 2020.

[19] Jiang Tian, Li Liu, Zhongchao Shi, and Feiyu Xu. Automatic couinaud segmentation from ct volumes on liver using glc-unet. In *International Workshop on Machine Learning in Medical Imaging*, pages 274–282. Springer, 2019.

[20] Patricia C Valery, Mathieu Laversanne, Paul J Clark, Jessica L Petrick, Katherine A McGlynn, and Freddie Bray. Projections of primary liver cancer to 2030 in 30 countries worldwide. *Hepatology*, 67(2):600–611, 2018.

[21] Gabriele Valvano, Andrea Leo, and Sotirios A Tsaftaris. Learning to segment from scribbles using multi-scale adversarial attention gates. *IEEE Transactions on Medical Imaging*, 2021.

[22] Juan Wang and Bin Xia. Bounding box tightness prior for weakly supervised image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 526–536. Springer, 2021.

[23] Xi Wang, Fangyao Tang, Hao Chen, Luyang Luo, Ziqi Tang, An-Ran Ran, Carol Y Cheung, and Pheng-Ann Heng. Ud-mil: uncertainty-driven deep multiple instance learning for oct image classification. *IEEE Journal of Biomedical and Health Informatics*, 24(12):3431–3442, 2020.

[24] Zhanwei Xu, Yukun Cao, Cheng Jin, Guozhu Shao, Xiaoqing Liu, Jie Zhou, Heshui Shi, and Jianjiang Feng. Gasnet: Weakly-supervised framework for covid-19 lesion segmentation. *arXiv preprint arXiv:2010.09456*, 2020.

[25] Qingsen Yan, Bo Wang, Dong Gong, Dingwen Zhang, Yang Yang, Zheng You, Yanning Zhang, and Javen Qinfeng Shi. A comprehensive ct dataset for liver computer assisted diagnosis. In *British Machine Vision Conference*, 2021.

[26] Yue Zhang, Chengtao Peng, Liying Peng, Huimin Huang, Ruofeng Tong, Lanfen Lin, Jingsong Li, Yen-Wei Chen, Qingqing Chen, Hongjie Hu, et al. Multi-phase liver tumor segmentation with spatial aggregation and uncertain region inpainting. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 68–77. Springer, 2021.

[27] Hao Zheng, Zhiguo Zhuang, Yulei Qin, Yun Gu, Jie Yang, and Guang-Zhong Yang. Weakly supervised deep learning for breast cancer segmentation with coarse annotations. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 450–459. Springer, 2020.