

Dual consistency assisted multi-confident learning for the hepatic vessel segmentation using noisy labels

Nam Phuong Nguyen*,¹

namng1210@gmail.com

Tuan Van Vo*,¹

vovantuan1999a@gmail.com

Soan T. M. Duong^{1,2}

soanduong@lqdtu.edu.vn

Chanh D. Tr. Nguyen^{1,3}

v.chanhng@vinbrain.net

Trung Bui

bhtrung@yahoo.com

Steven Q. H. Truong^{1,3}

v.brain01@vinbrain.net

¹ VinBrain JSC, Vietnam

² Computer Science Department,
Le Quy Don Technical University,
Vietnam

³ College of Engineering
and Computer Science,
VinUniversity, Vietnam

Abstract

Noisy hepatic vessel labels from Computer Tomography (CT) are popular due to vessels' low-contrast and complex morphology. This is challenging for automatic hepatic vessel segmentation, which is essential to many hepatic surgeries such as liver resection and transplantation. To exploit the noisy labeled data, we proposed a novel semi-supervised framework called dual consistency assisted multi-confident learning (DC-Multi-CL) for automatic hepatic vessel segmentation. The proposed framework contains a dual consistency architecture that learns not only the high-quality annotation data but also the low-quality data by boosting the prediction consistency on low-quality labeled data robustly. Furthermore, we also present a multi-confident learning component to exploit the capability of global context information from multi-level network features and eradicate the human efforts on refining the low-quality data. Combining these ideas, we believe that it raises a potentially valuable approach to handle segmentation task, especially when the annotation data are noisy, e.g. unlabeled and mislabeled voxel-wise. Extensive experiments on two public datasets, i.e. 3DIRCADb and MSD8, demonstrate the effectiveness of each component and the superiority of the proposed method to other state-of-the-art methods in hepatic vessel segmentation and semi-supervised segmentation. The implementation of DC-Multi-CL is available at: https://github.com/VinBrainJSC/DualConsistency_Mutil-CL.git.

1 Introduction

Hepatic vessel segmentation is essential to identify the liver segments toward helpful guidance for liver resection and transplantation. Recently, a deep learning approach to automatically segment hepatic vessels from computer tomography has been promising [5, 9, 20]. However, a good deep learning segmentation model often requires a great number of CT data and their ground-truth, i.e. high-quality (HQ) voxel-wise annotations. Low-quality (LQ) annotations lead to undesirable performance degradation in deep learning. Meanwhile, labeling the hepatic vessels in CT is difficult and time-consuming due to expertise-demanding, complex morphology, noise, pathological variations, and poor contrast, resulting in a limited number of HQ vessel annotations. Therefore, it has been of interest for researchers to train deep models for hepatic vessel segmentation given only a few HQ-annotated data and a popular LQ-annotated data of hepatic vessel data. Semi-supervised learning (SSL) is the most appropriate approach to explore the auxiliary image information from additional datasets and regularize the learner.

Related Work. To tackle the shortage of labeled data, many researchers have approached semi-supervised learning to leverage large amounts of unlabeled data toward improving the performance of supervised learning over the small labeled dataset. Advanced semi-supervised learning techniques in medical image segmentation often are based on adversarial training, pseudo-labeling, and consistency regularization [2, 9, 6, 12, 13, 18, 19].

Consistency regularization with the perturbation-based methods [12, 18] has been most widely used in semi-supervised segmentation. These methods enforce the consistency of the predictions/intermediate features by adding small perturbations to unlabeled samples, in which the decision boundary should lie in low-density regions. Mean Teacher [18] methods impose consistency over perturbed inputs or augmented images, encouraging the model to produce similar output/distributions for the perturbed inputs. Meanwhile, Cross-Consistent Training (CCT) [12] enforces consistency between the main decoder predictions and those of the auxiliary decoders in which the main decoder is fed by the encoder’s output and the auxiliary decoders are fed by the encoder’s output with perturbations. Thus, CCT is able to improve the representation of the encoder.

PseudoSeg [25] explores pseudo segmentation for SSL by using the pseudo-segmentation of a weakly augmented image to supervise the segmentation of a strongly augmented image based on a single segmentation network. Cross Pseudo Supervision [2] adopts two identically and independently initialized segmentation networks with the same input and uses each network’s pseudo segmentation maps to supervise the other network. These existing methods, though, have shown being able to leverage knowledge from unlabeled data for learning but not utilize the potentially useful information of the LQ label.

In addition, to fully exploit the noisy labeled data, recent works have proposed to decrease the negative effects brought by the noisy labels, such as assigning lower weights to the noisy labeled samples [14, 24], modeling the label corrupting process [3] and confident learning [11]. However, these studies were initially proposed for the classification problem, while the localization of pixel-wise noises is necessary for segmentation. Recently, Xu et al. [21] firstly proposed an SSL framework, which uses the mean teacher-assisted confident learning to take advantage of the noise from the low-quality labeled data towards superior vessel segmentation performance. In our method, we leverage the confident learning module performance by utilizing more meaningful contextual information from all scales to produce the maximum confident guidance for segmentation networks.

To summarize, our the contributions of this study are highlighted as follows:

- We proposed a dual consistency to encourage scale consistency for an input from LQ labeled data. It leverages both low- and high-quality annotations by the interpolation consistency mechanism to further encourage prediction consistency on low-quality labeled data robustly. Besides, the dual consistency also enforces the consistency of the LQ labeled data by the cross pseudo supervision as inspired by [10].
- We introduce the interpolation consistency training (ICT), which was initially proposed for the classification problem, into the vessel segmentation, aiming to encourage the consistency among the predictions of the unlabeled data via the interpolation instead of adding small perturbation [10]. Furthermore, we adopt the multi-scale consistency at the interpolation sample to further exploit the prediction discrepancy of the segmentation model.
- We introduce Multi-Confident Learning (Multi-CL) as the upgrade version of confident learning with the guidance maps aggregated from multiple scales instead of the one from the last scale. Multi-CL, thus, utilizes the capability of global context information from the multi-level network features and avoids information lost during the upsampling process. With the help of interpolating consistency and incorporating uncertainty as weights across different scales, Multi-CL can also benefit from better exploiting the prediction discrepancy of segmentation models at multiple scales.
- We conduct extensive experiments on two public liver vessel datasets, i.e. 3DIRCADb and MSD8, with five-fold cross-validation. The proposed DC-Multi-CL is state-of-the-art compared to other existing SSL methods on the hepatic vessel segmentation.

2 Methods

2.1 Datasets

In this study, we used two public datasets, i.e. 3DIRCADb [10] and MSD8 [10], which are referred as high-quality and low-quality datasets, respectively (see the visualization in Fig. 1). The 3DIRCADb dataset consists of 20 CT scans of the enhanced portal venous phase with corresponding HQ vessel annotation. Each scan consists of around 74 to 260 slices with the thickness ranging from 1 mm to 4 mm. All slices in the scans are axial with a size of 520×520 pixels, and the pixel spacing is in the range of $0.56 \div 0.86$ mm. The MSD8 dataset consists of 443 CT hepatic scans with LQ vessel annotations, i.e. approximately 65.5% vessel is unlabeled and around 8.5% labeled vessel is misclassified [8]. Each scan comprises axial slices with a slice thickness from 0.8 to 8 mm and pixel spacing from $0.57 \div 0.98$ mm.

2.2 Dual consistency assisted multi-confident learning framework

2.2.1 Overview of the proposed method

Our proposed DC-Multi-CL framework architecture consists of two parallel segmentation networks $F_1(\cdot)$ and $F_2(\cdot)$, as manifested in Fig. 2. Specifically, the aim of $F_1(\cdot)$ network is to learn the main features from the input data, while the $F_2(\cdot)$ acts as the auxiliary network

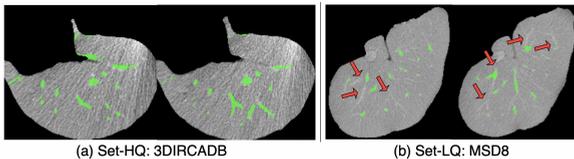


Figure 1: 2D visualization of (a) 3DIRCADb dataset [10] (Set-HQ), and (b) MSD8 dataset [11] (Set-LQ). Green patterns represent the labeled vessels and red arrows point at unlabeled pixels.

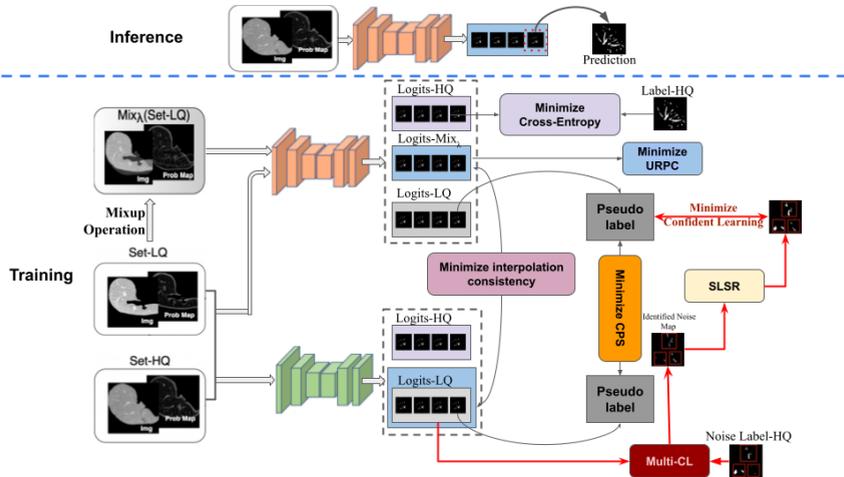


Figure 2: Illustration of the proposed dual consistency assisted multi-confident learning framework for automatic hepatic vessel segmentation.

to provide potentially useful information from noisy labels during the training process. Note that our proposed method only uses the prediction of $F_1(\cdot)$ for the inference process. We experimentally find that the noisy labels effectively significantly enhance the prediction quality of the entire segmentation model.

In detail, we select the backbone segmentation network as U-net [12], which can produce pyramid predictions at different scales [13]. These two network sharing have the same structure and are initialized differently. They produce a set of multi-scale predictions $[p_{10}, p_{11}, \dots, p_{1s}, \dots, p_{1(S-1)}]$ and $[p_{20}, p_{21}, \dots, p_{2s}, \dots, p_{2(S-1)}]$ for an input image x_i , respectively, where the p_{*s} is the prediction at scale s , and a smaller s corresponds to a higher resolution in the decoder. S is the number of scales in the pyramid prediction.

2.2.2 Dual consistency

Let \mathcal{D}_{HQ} and \mathcal{D}_{LQ} be the high-quality labeled data and noisy labeled (low-quality) data. We denoted a labeled image pair as $(x_i, y_i) \in \mathcal{D}_{HQ}$, where y_i is ground truth and a noisy labeled image pair as $(\tilde{x}_i, \tilde{y}_i)$, where noisy samples \tilde{x}_i with its label is \tilde{y}_i . The supervised branch exploits HQ-labeled data by calculating \mathcal{L}_s , the combination of the standard cross-entropy

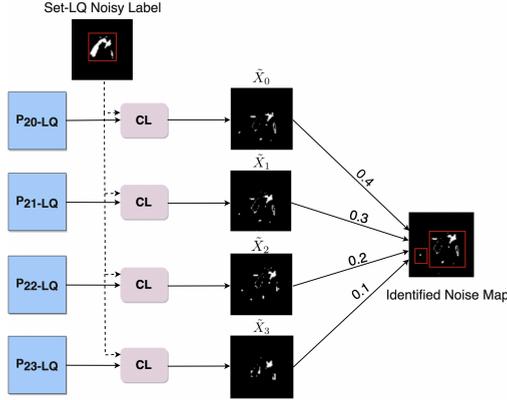


Figure 3: Illustration of the proposed multi-confident learning.

loss and dice loss at multiple scales:

$$\mathcal{L}_s = \frac{1}{S} \sum_{s=0}^{S-1} \frac{1}{|D_{HQ}|} \sum_{x_i \in D_{HQ}} \frac{1}{2} \left(\mathcal{L}_{dce}(p_{1s}(x_i), y_i) + \mathcal{L}_{dce}(p_{2s}(x_i), y_i) \right), \quad (1)$$

where y_i , \mathcal{L}_{dce} , denote the ground truth of input x_i and the combination of dice loss and the cross-entropy loss.

In the unsupervised branch, we further enhance consistency on two segmentation networks by combining interpolation consistency on multi-scale predictions. This not only forces the $F_1(\cdot)$ network to learn better features but also allows the $F_2(\cdot)$ network to take advantage of confident learning to produce the potential high quality refined data.

1. In order to learn from images of set LQ with pseudo labels, we use the pixel-wise one-hot label map Y_{10} output from one network $F_1(\cdot)$ to supervise the pixel-wise confidence map p_{20} of the other network $F_2(\cdot)$ and vice versa. The cross pseudo supervision loss at the highest resolution scale p_{10} and p_{20} in the decoder on the Set-LQ is written as:

$$\mathcal{L}_{cps} = \frac{1}{|D_{LQ}|} \sum_{x_i \in D_{LQ}} \left(\mathcal{L}_{ce}(p_{10}(\mathbf{x}_i), Y_{20}) + \mathcal{L}_{ce}(p_{20}(\mathbf{x}_i), Y_{10}) \right). \quad (2)$$

2. Following a mixup [23] operation: $\text{Mix}_\lambda(a, b) = \lambda a + (1 - \lambda)b$. We denote a noisy labeled image as $\mathbf{x}_i, \mathbf{x}_k \in D_{LQ}$. To get more conservative consistent regularization, we use the ICT [14] for each different scales of decoder to encourage the model $F_1(\cdot)$ to predict the fake label $\text{Mix}_\lambda(F_2(\cdot, \mathbf{x}_i), F_2(\cdot, \mathbf{x}_k))$ at interpolation location $\text{Mix}_\lambda(\mathbf{x}_i, \mathbf{x}_k)$. Following [23], on each update, we sample a random λ from $\beta(\alpha, \alpha)$. The unsupervised interpolation consistency loss L_{us} at multi scales is written as:

$$\mathcal{L}_{us} = \frac{1}{S} \sum_{s=0}^{S-1} \sum_{n_i, n_k \in D_{LQ}} \text{MSE} \left(p_{1s}(\mathbf{x}_m), \text{Mix}_\lambda(p_{2s}(\mathbf{x}_i), p_{2s}(\mathbf{x}_k)) \right), \quad (3)$$

where $\mathbf{x}_m = \text{Mix}_\lambda(\mathbf{x}_i, \mathbf{x}_k)$ and MSE is the mean squared error.

3. Inspired by previous works [14], to better exploit the prediction discrepancy of segmentation model at interpolation location $\text{Mix}_\lambda(\mathbf{x}_i, \mathbf{x}_k)$ with multiple scale, we adapt a novel

uncertainty rectified pyramid consistency loss for the predictions at the interpolation location \mathbf{x}_m . First, we denote the average $F_1(\cdot)$ network prediction at the interpolation \mathbf{x}_m across these scales as $p_c = \frac{1}{S} \sum_{s=0}^{S-1} p_{1s}(\mathbf{x}_m)$. Then, we formulate an efficient uncertainty estimation based on pyramid predictions at the interpolation \mathbf{x}_m .

Specifically, we use the KL-divergence between the average prediction and the prediction at scales as the uncertainty measurement $D_s = \sum_{j=0}^C p_{1s}^j(\mathbf{x}_m) \cdot \log \frac{p_c}{p_{1s}^j(\mathbf{x}_m)}$, where $p_{1s}^j(\mathbf{x}_m)$ is the j^{th} channel of $p_{1s}(\mathbf{x}_m)$, and C is the class number. The approximated uncertainty shows the difference between the $p_{1s}(\mathbf{x}_m)$ and p_c . Note that for a given pixel in D_s , a larger value indicates the prediction for that pixel at scale s is far from the other scales. As result, we obtain a set of an uncertainty maps D_0, D_1, \dots, D_{S-1} where D_s corresponds to uncertainty of $p_{1s}(\mathbf{x}_m)$. Finally, the uncertainty rectified pyramid consistency (URPC) loss at each scale of location \mathbf{x}_m is written as:

$$\mathcal{L}_{urpc} = \frac{1}{S} \frac{\sum_{s=0}^{S-1} \sum_v (p_{1s}^v(\mathbf{x}_m) - p_c)^2 \cdot w_s^v}{\sum_{s=0}^{S-1} \sum_v w_s^v} + \frac{1}{S} \sum_{s=0}^{S-1} \|D_s\|_2, \quad (4)$$

where p_{1s}^v and D_s^v are the corresponding prediction and uncertainty values for pixel v . We use a pixel- and scale-wise weight w_s^v to automatically rectify the MSE loss [10]. The weight for a pixel v at scale s is defined as: $w_s^v = e^{-D_s^v}$, it corresponds to pixel-wise exponential operation for $-D_s$.

2.2.3 Self denoised label aggregated by Multi-Confident Learning (Multi-CL)

As the confident learning technique method was first proposed for pruning mislabeled samples and improved the training by estimating the joint distribution between the noisy labels \tilde{y} and the true (latent) labels y^* in image-level classification. Therefore with its assistance, we want to adapt this method and even enhance its contribution to a higher level by jointly considering the additional guidance from different scale predictions of the decoder. We describe this method as multi-confident learning (Multi-CL).

Given a training set $\mathbf{X} = (\mathbf{x}, \tilde{y})^n$, containing n noisy samples \mathbf{x} with its label is \tilde{y} , the binary predicted probability map could be obtained through the $F_2(\cdot)$ segmentation network. Simply taking predictions from argmax operation as an indication of guidance for noisy labels leads to unsuccessfully counting errors for class imbalance properly or may guess over-confident for a certain class than others. Thus, we select the predictions by threshold t_j as an alternative. The prediction by thresholding is calculated as the expected predicted probabilities $\hat{p}_j(\mathbf{x})$ of all examples labeled with $\tilde{y} = j$: $t_j = \frac{1}{|\mathbf{X}_{\tilde{y}=j}|} \sum_{\mathbf{x} \in \mathbf{X}_{\tilde{y}=j}} \hat{p}_j(\mathbf{x})$.

We flatten the vessel annotation and predictions to 1D vectors, then feed them together with inputs into the confident learning component to make the algorithm treat each pixel like an individual image. Then, the confusion matrix can be formed by counting $\mathbf{C}_{\tilde{y}, y^*}$ where each element $\mathbf{C}_{\tilde{y}, y^*}[i][j]$ indicates the number of samples \mathbf{x} (observed label $\tilde{y} = i$) may belong to the true latent label $y^* = j$:

$$\mathbf{C}_{\tilde{y}, y^*}[i][j] = |\hat{\mathbf{X}}_{\tilde{y}=i, y^*=j}|, \text{ where} \quad (5)$$

$$\hat{\mathbf{X}} = \{\mathbf{x} \in \mathbf{X}_{\tilde{y}=i} : \hat{p}_j(\mathbf{x}) \geq t_j, j = \arg \max_{k \in \mathbf{M}: \hat{p}_k(\mathbf{x}) \geq t_k} \hat{p}_k(\mathbf{x})\}.$$

After obtaining a confusion matrix $\mathbf{C}_{\tilde{y},y^*}$, the joint distribution $m \times m$ $\mathbf{Q}_{\tilde{y},y^*}$ between the true labels and the noisy labels could be computed as:

$$\mathbf{Q}_{\tilde{y},y^*}[i][j] = \frac{\mathbf{C}_{\tilde{y},y^*}[i][j]}{\sum_{j \in M} \mathbf{C}_{\tilde{y},y^*}[i][j]} \cdot |\mathbf{X}_{\tilde{y}=i}|}{\sum_{i \in M, j \in M} \left(\frac{\mathbf{C}_{\tilde{y},y^*}[i][j]}{\sum_{j \in M} \mathbf{C}_{\tilde{y},y^*}[i][j]} \cdot |\mathbf{X}_{\tilde{y}=i}| \right)}. \quad (6)$$

Then, Prune By Class (PBC) is the approach we choose to clean data. Particularly, for each class $i \in M$, the $n \cdot \sum_{j \in M, j \neq i} (\mathbf{Q}_{\tilde{y},y^*}[i][j])$ samples with lowest self-confidence $\hat{p}(\tilde{y} = i; \mathbf{x} \in \mathbf{X}_i)$ are selected as the wrong-labeled samples. The obtained result \tilde{X} indicates the binary noise identification mask where "1" denotes the potential wrong label pixel and "0" denotes the potential correct one. Similarly, this method performs exactly the same on multi-scale predictions from lower to upper layers. We choose the number of scales in the pyramid prediction $S = 4$. Therefore, as shown in Fig. 3, the binary noise identification masks $\tilde{X}_0, \tilde{X}_1, \tilde{X}_2, \tilde{X}_3$ are obtained by putting the noisy label mask with the predictions at its corresponding scales as input for CL module. Then, the four identification masks are aggregated as:

$$\tilde{X}_{mcl} = 0.4\tilde{X}_0 + 0.3\tilde{X}_1 + 0.2\tilde{X}_2 + 0.1\tilde{X}_3, \quad (7)$$

$$\tilde{X}_{mcl} = \begin{cases} 1, & \tilde{X}_{mcl} \geq t_{mcl}, \\ 0, & \tilde{X}_{mcl} < t_{mcl}. \end{cases} \quad (8)$$

As the predictions have different spatial resolutions, even when at the last step, they can be re-sampled to the same shape as the input, the output can also face problems such as model collapse or loss of fine details due to the different spatial frequencies. Therefore, we assign the weight of each component to be decreased from high resolution to low resolution. Moreover, we set the value of threshold t_{mcl} to 0.5 since \tilde{X}_{mcl} still needs a threshold to define it back to the original binary noise identification mask. After having \tilde{X}_{mcl} , rather than directly imposing it, Self Label Smoothing Regularization (SLSR) can be formulated as:

$$\hat{y}(\mathbf{x}) = \tilde{y}(\mathbf{x}) + \mathbf{1}(\mathbf{x} \in \tilde{X}_{mcl}) \cdot (-1)^{\tilde{y}} \cdot \rho, \quad (9)$$

where $\mathbf{1}(\cdot)$ is the indicator function and ρ is the smooth factor. Since CL may have uncertainties, soft correction is prioritized and ρ is set as 0.8 empirically. With the obtained soft-corrected labels, we use the auxiliary self-denoised CL loss \mathcal{L}_{cl} , a combination of cross-entropy loss and focal loss:

$$\mathcal{L}_{cl} = \frac{1}{|D_{LQ}|} \sum_{\mathbf{x} \in D_{LQ}} \frac{\mathcal{L}_{dice}(p_{10}(\mathbf{x}), \hat{y}(\mathbf{x})) + \mathcal{L}_{focal}(p_{10}(\mathbf{x}), \hat{y}(\mathbf{x}))}{2}. \quad (10)$$

The proposed DC-Multi-CL framework learns from both HQ labeled data and LQ labeled data by minimizing the following the total loss function:

$$\mathcal{L}_{total} = \mathcal{L}_s + \lambda_c \cdot (\mathcal{L}_{cps} + \mathcal{L}_{us} + \mathcal{L}_{urpc}) + \lambda_{cl} \cdot \mathcal{L}_{cl}, \quad (11)$$

where λ_c, λ_{cl} are two ramp-up weighting functions for balancing different losses.

3 Experiments

3.1 Experiment settings

We followed the experiment setup MTCL [21] to conduct the report on 3DIRCADb. In this setup, the vessel probability map based on the Sato tubeness filter [16] was used as an auxiliary modality input. By doing that, the network could have more robust vessel signals instead of relying entirely on the processed images. To further increase the reliability, we conducted 5-fold cross-validation for all experiments, which is a more accurate method than picking random samples [21] as a validation dataset. Specifically, within each fold in semi-supervised learning methods, sixteen HQ cases and the whole LQ dataset were used for training, and the remaining four HQ cases were for evaluation. In contrast, for fully supervised learning methods, only the HQ dataset was used for training and validating due to the previous study [8, 20, 22] shows that mixing both these datasets could make degrade model performance.

We adopted Dice Score (DSC) and 95% Hausdorff Distance (HD95) to evaluate the segmentation performance. The experiments were implemented in PyTorch and trained on NVIDIA Tesla A100 with 80GB RAM. We utilized SGD optimizer ($momentum = 0.9$, $weight_decay = 0.0001$) to train the whole network for 40000 iterations with batch size of 4. The base learning rate was set as 0.01 and decay by a factor of $(1 - \frac{iter}{max_iter})^{0.9}$.

3.2 Performance and Comparison

A comprehensive comparison with existing methods was conducted under the same experimental environment and dataset settings to make the results more convincing, as shown in Table 1. This section compares our proposed DC-Multi-CL framework with both SOTA semi-supervised segmentation approaches and fully supervised segmentation methods for hepatic vessel segmentation.

Table 1: Comparison on 3DIRCADb for different methods.

Learning approaches	Methods	DSC \uparrow	HD \downarrow
Supervised learning	U-net [8]	64.01 \pm 5.11	9.50 \pm 2.03
	Swin-Unet [20]	63.95 \pm 3.61	9.99 \pm 1.46
Semi-supervised learning	MT [18]	65.78 \pm 4.17	8.55 \pm 1.36
	URPC [10]	66.47 \pm 4.94	8.49 \pm 1.23
	CCT [12]	66.10 \pm 4.47	8.85 \pm 1.25
	CPS [9]	66.55 \pm 4.19	8.85 \pm 1.63
Semi-supervised learning with LQ and HQ data	MTCL [21]	66.29 \pm 4.87	8.75 \pm 1.38
	Proposed DC-Multi-CL	67.41\pm4.68	8.38\pm1.56

In addition, some methods were evaluated on manually refined annotations [8], or specific data augmentation strategies like filters [9], leading to poor transparency results, so we do not consider these results with ours. All methods were reimplemented in 2D and using U-net [8] as the backbone for a fair comparison. Then, all methods were evaluated on Set-HQ. From the quantitative comparison of Table 1, our proposed method outperforms all the compared ones. Specifically, compared with the semi-supervised approach MTCL [21], the

average DSC increases from 66.29% to 67.41%. In terms of HD-95, our proposed method also achieves remarkable improvement. As demonstrated in Fig. 4, our method has fewer false negatives (green color) and false positives (yellow color) than other methods.

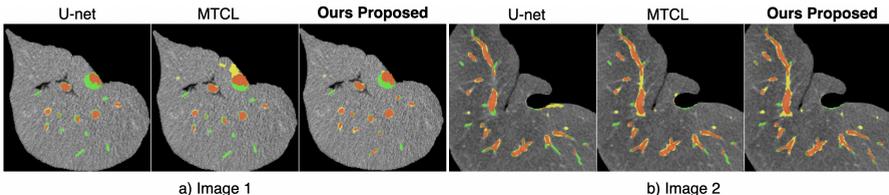


Figure 4: Visual comparison results on 3DIRCADb dataset. Orange indicates the correct segmented area, green the false negative and yellow the false positive.

4 Ablation Study

Effectiveness to each component. We conduct experiments to analyze the contributions of our proposed method. Table 2 shows that whether using Multi-CL alone or ICT alone, the performance of our proposed method always gains significant improvement, and the accomplished results could be even better when combining those components. This demonstrates that each proposed component plays a crucial role in our framework.

Table 2: Ablation study of the proposed framework on 3DIRCADb with 5-fold cross-validation. **DC-Multi-CL**: the proposed dual consistency assisted multi-confident learning framework, CPS: Cross pseudo supervision, ICT: Interpolation consistency training, Multi-CL: Multi confident learning, URPC: Uncertainty rectified pyramid consistency.

Method	DSC \uparrow	HD \downarrow
CPS [18]	66.55 \pm 4.19	8.85 \pm 1.63
CPS+ICT	66.53 \pm 4.54	8.51 \pm 1.51
CPS+Multi-CL	66.62 \pm 4.60	8.62 \pm 1.33
CPS+ICT+Multi-CL	66.71 \pm 4.62	8.73 \pm 1.43
CPS+URPC	66.53 \pm 4.15	8.53 \pm 1.25
CPS+ICT+URPC	66.87 \pm 3.90	8.58 \pm 0.94
CPS+Multi-CL+URPC	66.60 \pm 3.69	8.76 \pm 1.12
Proposed DC-Multi-CL	67.41\pm4.68	8.38\pm1.56

Note that in our experiment, we also compare the results produced by the URPC component alone without other integrated solutions. Based on Table 2, the experiment result does not show a significant improvement if used individually.

Can DC-Multi-CL be robust with other semi-supervised architecture? To prove the robustness, we further adapted our methods with commonly used semi-supervised architectures such as MT [18]. Regarding the 5-fold cross-validation set performance (as shown in Table 3), our proposed framework for the MT improve 1.48% in DSC and 0.21 in HD-95 compared to MT [18].

Table 3: Ablation study of the proposed framework on 3DIRCADb of Mean Teacher semi-supervised methods with 5-fold cross-validation.

Method	DSC \uparrow	HD \downarrow
MT [18]	65.78 \pm 4.17	8.55 \pm 1.36
MTCL [21]	66.29 \pm 4.87	8.75 \pm 1.38
MT+URPC	66.65 \pm 4.07	8.63 \pm 1.26
MT+ICT+URPC	67.05 \pm 4.42	8.31 \pm 1.28
MT+Multi-CL+URPC	66.58 \pm 4.21	8.45 \pm 1.33
(Proposed) MT+DC-Multi-CL	67.26 \pm 4.38	8.34 \pm 1.25

Robustness to multi-confident learning. To verify the guiding role of multi-confident learning against confident learning [21], we present a visual comparison in Fig. 5. It is observed that multi-confident learning yields better blood vessel structure with clear and precise boundaries. Our method not only retains all the important information in the CL denoised label but also yields better blood vessel structure with clear and precise boundaries based on the amount of information aggregated from multi scales. Thus, multi-confident learning would be able to detect details that CL may have missed and produce more similar results to the ground truth.

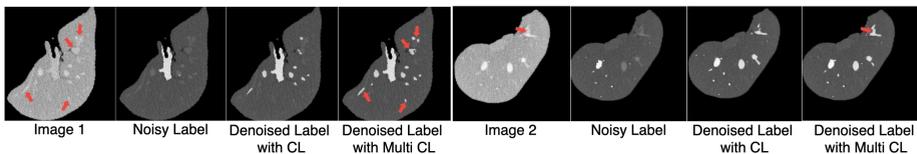


Figure 5: Illustration of the multi-confident learning performance for the MSD8 dataset.

5 Conclusion

This paper introduces a dual consistency assisted multi-confident learning framework for dealing with a shortage of high-quality data in the challenging hepatic vessel segmentation task. The proposed framework is trained with a small amount of high-quality labeled data and many noisy labeled data. With the effort of applying dual consistency training to generate consistent predictions and the refined noisy data produced by the multi-confident learning module, our proposed method significantly exceeds the previous state-of-the-art in vessel segmentation tasks. The intensive experiments also demonstrate the extraordinary promise of the proposed approach with a consistent segmentation improvement when integrating these components into different semi-supervised methods.

References

- [1] 3DIRCADb Dataset. URL <https://www.ircad.fr/research/data-sets/liver-segmentation-3d-ircadb-01/>.
- [2] Xiaokang Chen, Yuhui Yuan, Gang Zeng, and Jingdong Wang. Semi-Supervised Semantic Segmentation with Cross Pseudo Supervision. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1–10, 6 2021.
- [3] Jacob Goldberger and Ehud Ben-Reuven. Training deep neural-networks using a noise adaptation layer. In *Proc. International Conference on Learning Representations*, 2016.
- [4] Yves Grandvalet and Yoshua Bengio. Semi-supervised Learning by Entropy Minimization. In *Proc. Advances in Neural Information Processing Systems*, volume 17, pages 1–8, 2004.
- [5] Qing Huang, Jinfeng Sun, Hui Ding, Xiaodong Wang, and Guangzhi Wang. Robust liver vessel extraction using 3D U-Net with variant dice loss function. *IEEE Trans. Computers in Biology and Medicine*, 101:153–162, 10 2018. ISSN 18790534.
- [6] Dong hyun Lee. Pseudo-Label: The Simple and Efficient Semi-Supervised Learning Method for Deep Neural Networks. In *Workshop on Challenges in Representation Learning*, volume 93, page 2, 2013.
- [7] Titinunt Kitrungrotsakul, Xian-Hua Han, Yutaro Iwamoto, Lanfen Lin, Amir Hossein Foruzan, Wei Xiong, and Yen-Wei Chen. Vesselnet: A deep convolutional neural network with multi pathways for robust hepatic vessel segmentation. *IEEE Trans. Computerized Medical Imaging and Graphics*, 75:74–83, 2019. ISSN 0895-6111.
- [8] Li Liu, Jiang Tian, Cheng Zhong, Zhongchao Shi, and Feiyu Xu. Robust hepatic vessels segmentation model based on noisy dataset. In *Proc. Medical Imaging 2020: Computer-Aided Diagnosis*, volume 11314, pages 122 – 128, 2020.
- [9] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. *arXiv preprint arXiv:2103.14030*, 2021.
- [10] Xiangde Luo, Wenjun Liao, Jieneng Chen, Tao Song, Yinan Chen, Shichuan Zhang, Nianyong Chen, Guotai Wang, and Shaoting Zhang. Efficient Semi-Supervised Gross Target Volume of Nasopharyngeal Carcinoma Segmentation via Uncertainty Rectified Pyramid Consistency. In *Proc. Medical Image Computing and Computer Assisted Intervention*, pages 318–329, 12 2021.
- [11] Curtis G. Northcutt, Lu Jiang, and Isaac L. Chuang. Confident Learning: Estimating Uncertainty in Dataset Labels. *Journal of Artificial Intelligence Research*, 70:1373–1411, 10 2019.
- [12] Yassine Ouali, Céline Hudelot, and Myriam Tami. Semi-Supervised Semantic Segmentation with Cross-Consistency Training. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1–11, 2020.

- [13] Siyuan Qiao, Wei Shen, Zhishuai Zhang, Bo Wang, and Alan Yuille. Semi-supervised Learning by Entropy Minimization. In *Proc. European Conference on Computer Vision*, pages 135–152, 2018.
- [14] Mengye Ren, Wenyuan Zeng, Bin Yang, and Raquel Urtasun. Learning to reweight examples for robust deep learning. In *Proc. International Conference on Machine Learning*, pages 4334–4343, 2018.
- [15] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Proc. Medical Image Computing and Computer Assisted Intervention*, pages 234–241, 5 2015.
- [16] Yoshinobu Sato, Shin Nakajima, Nobuyuki Shiraga, Hideki Atsumi, Shigeyuki Yoshida, Thomas Koller, Guido Gerig, and Ron Kikinis. Three-dimensional multi-scale line filter for segmentation and visualization of curvilinear structures in medical images. *IEEE Trans. Medical Image Analysis*, 2(2):143–168, 1998. ISSN 1361-8415.
- [17] Amber L Simpson, Michela Antonelli, Spyridon Bakas, Michel Bilello, Keyvan Farahani, Bram Van Ginneken, Annette Kopp-Schneider, Bennett A Landman, Geert Litjens, Bjoern Menze, et al. A large annotated medical image dataset for the development and evaluation of segmentation algorithms. *arXiv preprint arXiv:1902.09063*, 2019.
- [18] Antti Tarvainen and Harri Valpola. Mean teachers are better role models: weight-averaged consistency targets improve semi-supervised deep learning results. In *Proc. Neural Information Processing Systems*, pages 1–10, 2017.
- [19] Vikas Verma, Kenji Kawaguchi, Alex Lamb, Juho Kannala, Yoshua Bengio, and David Lopez-Paz. Interpolation Consistency Training for Semi-Supervised Learning. *IEEE Trans. Neural Networks*, 145:90–106, 3 2022.
- [20] Mian Wu, Yinling Qian, Xiangyun Liao, Qiong Wang, and Pheng-Ann Heng. Hepatic vessel segmentation based on 3d swin-transformer with inductive biased multi-head self-attention. *arXiv preprint arXiv:2111.03368*, 11 2021.
- [21] Zhe Xu, Donghuan Lu, Yixin Wang, Jie Luo, Jayender Jagadeesan, Kai Ma, Yefeng Zheng, and Xiu Li. Noisy Labels are Treasure: Mean-Teacher-Assisted Confident Learning for Hepatic Vessel Segmentation. In *Proc. Medical Image Computing and Computer Assisted Intervention*, pages 3–13, 6 2021.
- [22] Jinzhu Yang, Meihan Fu, and Ying Hu. Liver vessel segmentation based on inter-scale v-net. *IEEE Trans. Mathematical Biosciences and Engineering*, 18:4327–4340, 2021. ISSN 15510018.
- [23] Hongyi Zhang, Moustapha Cisse, Yann N. Dauphin, and David Lopez-Paz. mixup: Beyond Empirical Risk Minimization. *arXiv preprint arXiv:1710.09412*, 10 2017.
- [24] Haidong Zhu, Jialin Shi, and Ji Wu. Pick-and-learn: Automatic quality evaluation for noisy-labeled image segmentation. In *Proc. International Conference on Medical Image Computing and Computer Assisted Intervention*, pages 576–584, 2019.
- [25] Yuliang Zou, Zizhao Zhang, Han Zhang, Chun-Liang Li, Xiao Bian, Jia-Bin Huang, and Tomas Pfister. Pseudoseg: designing pseudo labels for semantic segmentation. In *Proc. International Conference on Learning Representations*, pages 1–18, 10 2021.