Bidirectional Teaching between Lightweight Multi-view Network for Intestine Segmentation from CT Volume

Qin An[a], Hirohisa Oda[b], Yuichiro Hayashi[a], Takayuki Kitasaka[c], Aitaro Takimoto[d], Akinari Hinoki[d], Hiroo Uchida[d], Kojiro Suzuki[e], Masahiro Oda[f,a], Kensaku Mori[a,f,g]

[a]Graduate School of Informatics, Nagoya University, Nagoya, Japan
[b]School of Management and Information, University of Shizuoka, Shizuoka, Japan
[c]School of Information Science, Aichi Institute of Technology, Toyota, Japan
[d]Nagoya University Graduate School of Medicine, Nagoya, Japan
[e]Department of Radiology, Aichi Medical University, Nagakute, Japan
[f]Information Technology Center, Nagoya University, Nagoya, Japan
[g]Research Center for Medical Bigdata, National Institute of Informatics, Tokyo, Japan

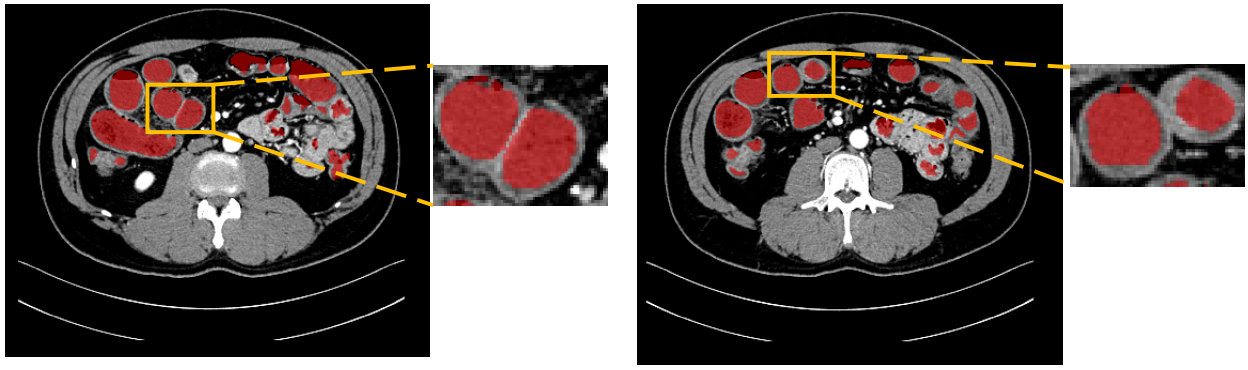
Abstract. Purpose: This paper proposes a novel semi-supervised method for intestine segmentation to help clinicians diagnose intestine diseases. Accurate intestine segmentation is critical to developing a treatment plan for intestine diseases like intestinal obstruction. Although full-supervision learning achieves good results with sufficient labeled data, the intestine has limited labeled data due to labeling time-consuming caused by its complex spatial structure. We propose a 3D segmentation network incorporating a bidirectional teaching strategy to improve the accuracy of intestine segmentation with the limited labeled dataset.
Method: The proposed method is based on semi-supervised learning, segmenting the intestine from computed tomography (CT) volumes. The model incorporates bidirectional teaching, where two backbones with different initial weights are simultaneously trained for intestine segmentation and generate pseudo-labels for using unlabeled data to reduce the influence of limited labeled data. In addition to the difficulty of limited labeled data, intestine segmentation is difficult due to complex spatial features. Thus, we propose a lightweight multi-view symmetric network, called LMVS-Net, which uses stacks of small-sized convolutional kernels instead of large ones to reduce the parameters and capture multi-scale features from various perceptual fields to improve the learning ability.
Results: We evaluated the proposed method with 59 CT volumes and repeated all experiments five times. Experimental results showed that the average Dice of the proposed method was 80.45%, the average precision was 84.12%, and the average recall was 78.84%.
Conclusions: The proposed method can effectively utilize large-scale unlabeled data with pseudo-labels, which is crucial in reducing the effect of limited labeled data in medical image segmentation. Furthermore, we assign different weights to the pseudo-labels to improve their reliability. From the result, we can see that the method produced competitive performance compared to previous methods.

Keywords: Intestine segmentation; semi-supervision; computer-aided diagnosis; pseudo-label.

***Qin An, qinan@mori.m.is.nagoya-u.ac.jp**

1 Introduction

Intestine obstruction[1–3] is a severe disease involving mechanical or functional intestine blockage, often resulting in intense abdominal pain, vomiting, and distention. Computed tomography (CT) is an important imaging method that can help clinicians diagnose diseases by providing more detailed imaging of the intestines. However, it is time-consuming for clinicians to check a patient's



CT slice 1 CT slice 2

Fig 1 The example of CT slices. We enlarge some regions in the yellow boxes and see some contact in the intestine wall, while just segmenting lumen regions can avoid this contact. In the enlarged parts, the red and light regions denote the lumen and wall, respectively.

CT volume. Computer-assisted diagnosis systems can assist clinicians in accurately identifying obstructions within the CT volume. However, the intestine is long and folded in the abdomen, which will cause most of the contact in the intestine wall. Therefore, we aim to develop a segmentation method that extracts the intestine lumen, including the small and large intestine lumens, to avoid most contact and help clinicians better understand the intestine's structure. In the paper, the word 'intestine segmentation' denotes the small and large bowel lumen segmentation. The CT slices in Fig. 1 show that just segment lumen regions can avoid most contact.

For organ segmentation, numerous methods[4–6] have been explored by researchers. Most of these methods rely on pixel-level labeled data. However, medical image annotations at the pixel-level are more difficult than those of natural images because clinicians with specialized anatomical knowledge need to do the former. Given the limited labeled dataset, the effective utilization of unlabeled data becomes crucial for medical image segmentation. Based on that, the segmentation of the intestine presents more challenges due to its elongated, tubular morphology and intricate spatial configuration, setting it apart from larger organs such as the liver. At present, several methods have been employed for intestine segmentation. Rajamani et al.,[7] proposed using a region

growing method to segment the intestine regions from CT volumes. Zhang et al.[8] also employed a region growing combined with the pre-trained probability map to segment the colon. Frimmel et al.[9] proposed using the centerline to segment the colon. Alberto et al.[10] proposed an adaptive 3D region-growing algorithm to segment the colon. The three previous methods segment the colon from CT volumes based on the region-growing method. Shin et al.[11] proposed adding cylindrical topological constraint into 3D U-Net to segment small bowel. Oda et al.[12] employed an improved 3D U-Net to segment intestines (including small and large bowel) from CT volumes. The two methods employed the fully-supervised method for the intestine segmentation task. The intestine has a complex spatial structure and low contract between the surrounding tissue, which increases the difficulty of segmenting the intestines from CT. Compared with the region-growing method, fully-supervised methods can extract comprehensive semantic features of the intestine from CT volumes to achieve better segmentation results. However, fully supervised methods require a large amount of labeled data, which is time-consuming and difficult. Therefore, we propose using an improved semi-supervised method to segment the intestine from CT volumes. Those methods just rely on full-supervised learning or hand-crafted features. Furthermore, the intestine has a complex spatial structure and low contract between the surrounding tissue, which increases the difficulty of segmenting the intestines from CT.

Semi-supervised segmentation methods offer a pragmatic compromise between supervised and unsupervised learning by harnessing both labeled and unlabeled data. Their key advantage over purely supervised learning lies in their ability to leverage a larger pool of unlabeled data, often more readily available, to enhance model training. A noteworthy example of such a semi-supervised segmentation method is cross pseudo-supervision (CPS).[13] CPS utilizes a combination of labeled and unlabeled data, using labeled data to generate pseudo-labels for unlabeled data. By incorporating

information from both labeled and unlabeled data, CPS enhances the model's adaptability and robustness, ultimately leading to more accurate and reliable segmentation results. In recent research, several semi-supervision segmentation methods[14,15] have attracted researchers' attention and applied them to organ segmentation tasks. Pseudo-labels,[16] mean teacher models,[17] and entropy minimization,[18] all of these methods based on semi-supervision, have achieved good performance in some organ segmentation tasks. However, it is notable that most existing intestine segmentation methods have yet to use the advantages of semi-supervised segmentation techniques, which effectively utilize both labeled and unlabeled data. Therefore, this work introduces an innovative approach designed to harness the potential of unlabeled data for the specific task of intestine segmentation.

We propose a novel segmentation network called a lightweight multi-view symmetrical network (LMVS-Net) to learn more intestine information by effectively extracting multi-scale semantic information, which can improve the model's learnability and solve the challenge caused by complex structure and low contrast. Furthermore, inspired by the CPS,[13] we incorporate the bidirectional teaching strategy with the new backbone, LMVS-Net, to utilize the unlabeled data to decrease the influence of limited labeled data that is common in organ segmentation tasks. The proposed method was developed based on our SPIE conference's method (MVS-Net).[19] Based on MVS-Net, we employ the stacks of small-sized convolutional kernels rather than large convolutional kernels to reduce the number of model parameters while maintaining segmentation accuracy. And utilize distance map weight in the loss function. In addition to the optimization of the methods, we also increase experiments, such as using different numbers of labeled data to validate the method and the ablation study about loss functions to make the research more comprehensive.

The framework trains two networks with the same structure but different initial weights simul-

taneously, taking labeled and unlabeled data as input. The labeled data are used to calculate the supervision loss, which guides the training process. The unlabeled data are utilized to calculate the loss value with pseudo-labels generated by the predictions of the two sub-networks, respectively. This is a strategy that makes the network use large-scale unlabeled datasets by generating pseudo-labels in semi-supervision methods. However, the pseudo-labels may be unreliable due to relying on the prediction of the network, especially in the early stages of training. In general, the prediction of the central parts, which are not adjacent to the background of the organ, is reliable. The peripheral parts, which are adjacent to the background, are unreliable due to the peripheral parts having complex situations, such as contact with other tissues or organs. Thus, we assume that the predictions of the central parts are more reliable than those of the peripheral parts, and this reliability is represented in terms of the closest distance between the closest background and each foreground part. To decrease the influence of the unreliability of the pseudo-labels during loss calculation, we generate a distance map according to the Euclidean distance transform of pseudo-labels and assign different weights to different regions calculated by the distance between the closest background regions to each foreground. Furthermore, the intestine is a tubular-shaped organ. For such organs, centerline-based information can be utilized as their shape-based feature. We employ not only the conventional Dice[20] but also the clDice loss,[21] which can use centerline information to guide the training.

This paper proposes that the LMVS-Net combined with bidirectional teaching be used to segment the intestine from CT volumes. The contributions of this paper are summarized as:

1) We present the lightweight multi-view model, LMVS-Net, for the intestines segmentation task.

2) We generate a distance weight map from the pseudo-labels to reduce the influence of potentially unreliable exist in pseudo-labels.

2 Method

2.1 Overview

The proposed method segments the intestine regions from a CT volume. In the method, we simultaneously train two LMVS-Nets having different initial weights. The training data consists of limited labeled data and large-scale unlabeled data. Consequently, two loss functions are defined, one for labeled data called supervision loss and the other for unlabeled data called unsupervised loss. In detail, the loss value for labeled data is calculated to update the models' parameters by supervision loss. The bidirectional teaching strategy generates pseudo-labels for unlabeled data for updating the parameters by unsupervised loss.

In the training, we crop the CT patches with size $256 \times 256 \times 16$ voxels and stride as $128 \times 128 \times 8$ voxels from CT volumes and randomly choose these patches as mini-batches to train the segmentation model. We apply flipping and cut-out[22] operations as data augmentation before training. In the testing, we also crop the CT patches from CT volumes by overlapped sliding windows with size $256 \times 256 \times 16$ voxels and the stride $128 \times 128 \times 8$ voxels to infer the segmentation result by a trained segmentation model. Finally, we merge the results of those patches with the same size as the CT volume to generate the output of the testing. We use the average of predicted probability for the overlapped regions as the final result. Figure 2 shows the flowchart of the intestine segmentation method.

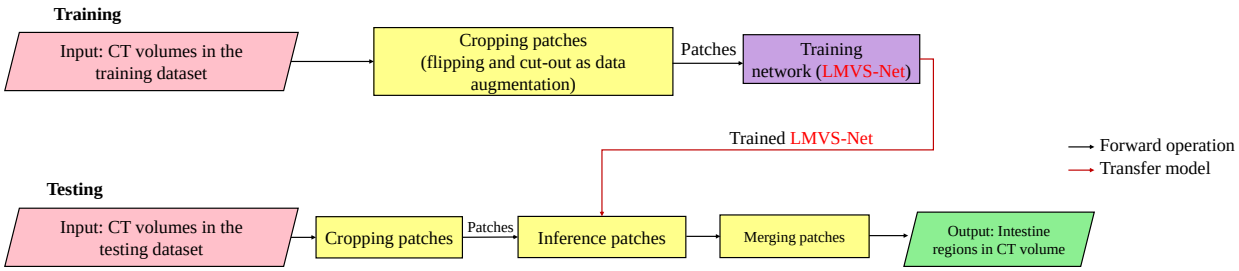


Fig 2 The flowchart of the intestine segmentation method in training and testing. In the training, we use CT volumes as input with some data augmentation to train a segmentation model. In the testing, we also crop the CT patches and then use the trained model to infer the patches. Finally, we merge them as the final output.

2.2 Lightweight Multi-view Symmetric Network (LMVS-Net)

The LMVS-Net enhances its learning capability by employing various sizes of convolutional kernels to capture multi-view features from various perceptual fields. Therefore, the model is called multi-view from the feature level. Additionally, it effectively reduces the number of network parameters by stacking small-sized convolutional kernels instead of using large ones.

Szegedy, et al[23] introduce Inception-v2 using convolutional kernels of different sizes to capture multi-scale features enabling the network to fully utilize features from multi-view, which helps enhance the network's learning ability. Furthermore, Inception-v2 uses small-size convolutional kernels to reduce the number of parameters, making the network more feasible for training. We integrate concepts from Inception-v2 into the 3D U-Net, leveraging multi-scale features to enhance network performance. Considering that CT volumes are 3D images requiring significant computational resources, we also employ stacked small-size convolutional kernels to achieve the effect of large-size convolutional kernels, which reduce computational costs while preserving a symmetric structure. Specifically, we use two $3 \times 3 \times 3$ convolution kernels to realize the convolution effect of one $5 \times 5 \times 5$, three $3 \times 3 \times 3$ convolution kernels to realize the convolution effect of one $7 \times 7 \times 7$. The $1 \times 1 \times 1$ convolution kernel in the multi-view convolutional block is designed to

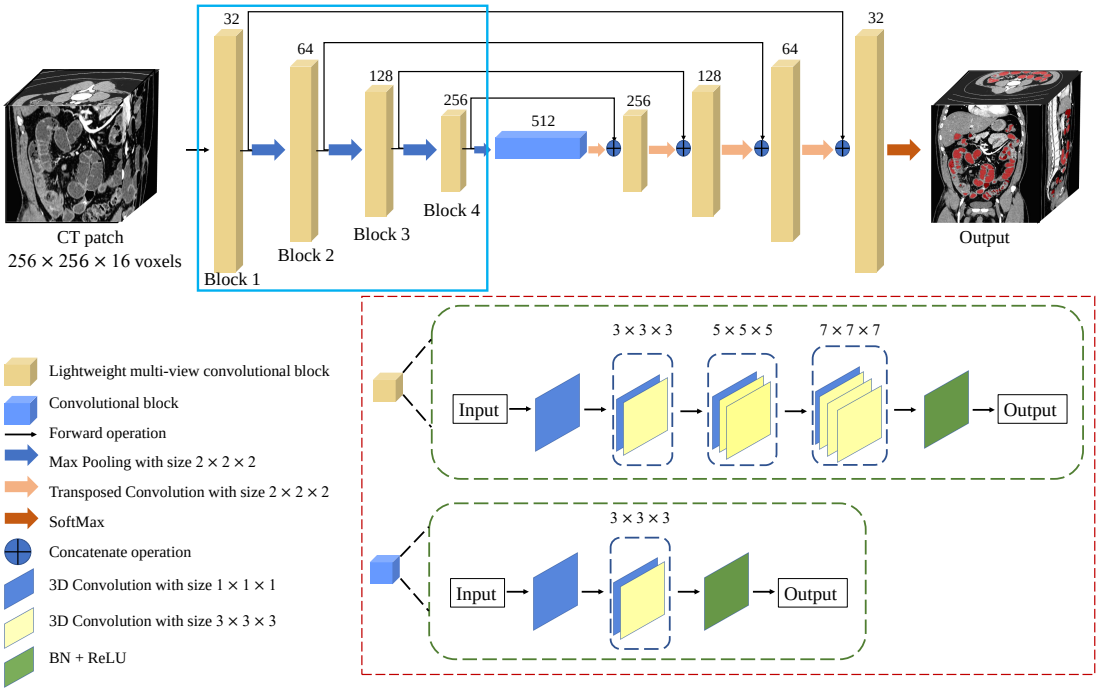


Fig 3 The structure of LMVS-Net. BN denotes batch normalization, and ReLU denotes the ReLU activation function. The numbers above the lightweight multi-view convolutional blocks indicate the number of kernels in each convolutional block. The detailed structure of the lightweight multi-view convolutional block and the convolutional block are shown in the red box. The encoder is shown in the blue box, and we show the internal feature map extracted by Block 1-4 in Fig. 4.

change the number of channels and decrease the parameters of the model. The detailed structure of the LMVS-Net is shown in Fig. 3. Firstly, we get the mean of the feature maps according to the number of channels and output the final feature maps. Second, we normalized the gray of feature maps to (0-255). Finally, we use the color mapping method[24] in OpenCV to map the feature maps' gray values to color values from blue to red. In our research, we used blue to red. The visualization in Fig. 4 shows that the intestine regions are more obvious as the deeper of the convolutional block.

2.3 Bidirectional Teaching

We train two LMVS-Nets having different initial weights using CT patches $\mathbf{X}^l$ with their ground-truth $\mathbf{G}$, and unlabeled CT patches $\mathbf{X}^u$ as input. The overview of bidirectional teaching between

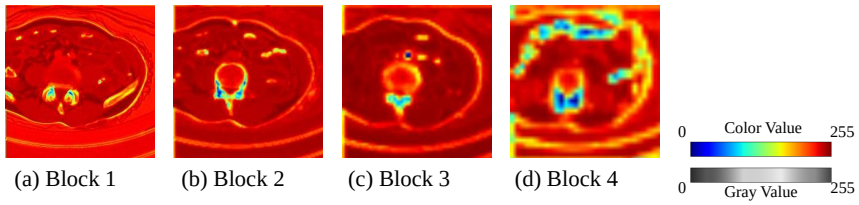


Fig 4 The internal feature maps visualization extracted by the encoder of LMVS-Net. (a-d) are feature maps extracted by each convolutional block. Regions with large feature values are colored in red, and lower values in yellow or blue. Intestinal regions tend to be colored in yellow or blue. Block 1-4 means the convolution block in the encoder part.

two LMVS-Nets is shown in Fig. 5.

The input consists of a pair of labeled and unlabeled data, prompting the proposed framework to generate two corresponding predictions, $\hat{\mathbf{P}}_i^l = f_i\left(\mathbf{X}^l\right)$ and $\hat{\mathbf{P}}_i^u = f_i\left(\mathbf{X}^u\right)$ from one LMVS-Net, where $i \in 1, 2$ denotes the *i*-th LMVS-Nets (LMVS-Net1, LMVS-Net2), as shown in Fig. 5, $f_i(\cdot)$ represents *i*-th LMVS-Net. $\hat{\mathbf{P}}_i^l$ and $\hat{\mathbf{P}}_i^u$ denote the predictions from the labeled and the unlabeled CT patches, respectively. We obtain the pseudo-labels by binarizing the predictions using argmax operation. The LMVS-Net1 generates the pseudo-label for the prediction of unlabeled CT patches from LMVS-Net2, and vice versa.

$$\mathbf{P}_i^* = \operatorname{argmax}\left(\mathbf{P}_i^u\right), \tag{1}$$

where $\mathbf{P}_i^*$ represents the pseudo-labels. The two LMVS-Nets supervise each other until training is complete, called bidirectional teaching.

2.4 Loss Function with Distance Weight

The overall loss is a compound loss with two parts, a supervised loss, L_{sup}, for the labeled data, and an unsupervised loss, L_{un}, for the unlabeled data. The supervised loss consists of two loss

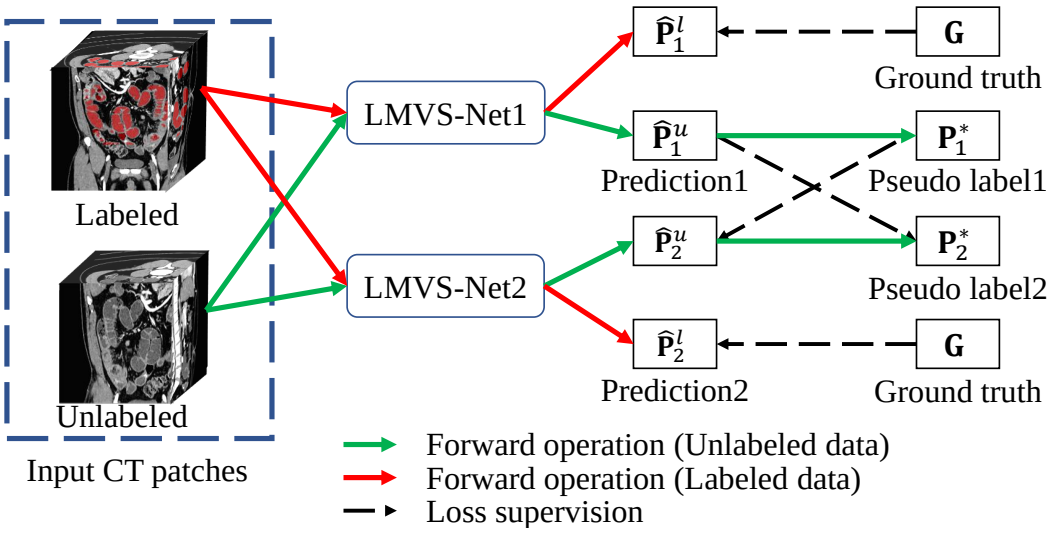


Fig 5 The overview of the bidirectional teaching between two LMVS-Nets. It is worth noting that the input for each iteration is a pair of labeled and unlabeled patches. Pseudo-labels are generated for unlabeled data, enabling the utilization of large-scale unlabeled data during training.

functions,

$$L_{sup}\left(\hat{\mathbf{P}}_i^l, \mathbf{G}\right) = \alpha L_{dice}\left(\hat{\mathbf{P}}_i^l, \mathbf{G}\right) + (1-\alpha) L_{cldice}\left(\hat{\mathbf{P}}_i^l, \mathbf{G}\right), \tag{2}$$

where L_{dice} denotes Dice loss, which is commonly used in medical image segmentation. L_{cldice} denotes clDice loss,[21] which uses the intersection between the prediction's centerline and the ground truth and the intersection of the prediction and the ground truth's centerline to calculate the loss value. $\hat{\mathbf{P}}_i^l$ denotes the prediction of labeled CT patches in i-th LVSM-Net $(i \in 1, 2)$, and $\mathbf{G}$ denotes the ground truth. α denotes the weight of the loss function, we experimentally set the value of α = 0.5.

In the unsupervised loss function, pseudo-labels are assigned to unlabeled data and calculate the loss. However, these pseudo-labels are derived from the model's predictions and may be unreliable. It is crucial to handle the pseudo-labels carefully to reduce the influence of the unreliable. As we all know, the segmentation in the boundary part is a common challenge in medical image segmentation tasks. In this work, we calculate the mean of the distance maps for each patch and

use it as a weight in the loss function. This allows us to assign higher weights to patches that contain fewer boundary regions. Specifically, we generate a distance map by Euclidean distance transformation[25] to calculate the distance map. Then we normalize the distance map to $[0, 1]$ by the min-max normalization. Finally, we use the mean of the normalized distance map as a weight, called distance weight, and assign the weight to the unsupervised loss. Consequently, the unsupervised loss is defined as follows,

$$L_{un}\left(\hat{\mathbf{P}}_i^u, \mathbf{P}_j^*, \mathbf{D}_{X,Y,Z}^j\right) = \frac{1}{XYZ}\sum_{x=0}^{X-1}\sum_{y=0}^{Y-1}\sum_{z=0}^{Z-1} L_{dice}\left(\hat{\mathbf{P}}_i^u, \mathbf{P}_j^*\right)\mathbf{D}_{x,y,z}^j, \tag{3}$$

where $\hat{\mathbf{P}}_i^u$ denotes the prediction of unlabeled CT patches in i-th LMVS-Net, $\mathbf{P}_j^*$ denotes pseudo-labels from j-th LMVS-Net. Note that $i \in (1, 2), j \in (1, 2)$, and $i \neq j$. $\mathbf{D}_{x,y,z}^j$ represents the distance weight, and X, Y, Z denote the number of voxels in each of the three dimensions of $\mathbf{D}_{x,y,z}^j$, $x \in 0, ..., X-1, y \in 0, ..., Y-1$, and $z \in 0, ..., Z-1$ denote the position of each voxel. The final loss L is defined as,

$$L = L_{sup}\left(\hat{\mathbf{P}}_1^l, \mathbf{G}\right) + L_{sup}\left(\hat{\mathbf{P}}_2^l, \mathbf{G}\right) + \gamma L_{un}\left(\hat{\mathbf{P}}_1^u, \mathbf{P}_2^*, \mathbf{D}_2\right) + \gamma L_{un}\left(\hat{\mathbf{P}}_2^u, \mathbf{P}_1^*, \mathbf{D}_1\right), \tag{4}$$

where γ is a weight factor, which is defined by the training iteration, $\gamma(t) = 0.1e^{\omega}$, $\omega = -5(1 - \frac{t}{T})$, $t = (1, 2, ..., T)$ denotes the current iteration, and T represents the total iteration number.

3 Experiments and Results

3.1 Dataset and Evaluation Metrics

In this work, we used a dataset consisting of 171 cases of ileus patients' CT volumes and introduce the detailed information of CT volumes in Table 1. For the ground truth preparation, one technical researcher expert in medical imaging, a medical student studying intestinal diseases, and an experienced pediatric surgeon labeled the data first. Then, the experienced pediatric surgeon checked all the ground truths.

We used 85 CT volumes for training, including 13 densely labeled and 72 unlabeled CT volumes. Then, 27 CT volumes with sparsely labeled were used for validation, 59 CT volumes including 58 with sparsely labeled, and 1 with densely labeled for testing. Here, the densely labeled CT volume means the clinicians labeled intestines in hundreds of continuous slices but did not label intestine regions in every slice. The sparsely labeled CT volume means the clinicians labeled intestines in some of the discontinuous slices.

Before training, we cropped the 3D CT patches with $256 \times 256 \times 16$ voxels to trade off the computational efficiency. To enhance the generalization ability of the segmentation model, we employed data augmentation in our experiments, including flipping and cut-out. We quantitatively evaluated the segmentation results using three accuracy metrics: 1) Dice, 2) recall, and 3) precision rates and one distance-based metric: normalized surface distance (NSD). Almost all of the testing cases were sparsely labeled, for one sparsely labeled CT volume, the percentage of labeled slices in each CT volume ranged from 1.00% to 5.31%, with the number of labeled slices varying between 6 to 29. In the evaluation, we infer CT volumes and get 3D segmentation. However, we extracted the 2D labeled CT slice to calculate the validation metrics. The total number of 2D CT slices was

Table 1 Detailed information on CT volumes after interpolation.Information of CT volumes after interpolation. We present the slice size, slice number, pixel spacing, and slice thickness of our intestines dataset.

	Original	Interpolation Result
Slice size (pixels)	512×512	(281 - 463)×(281 - 463)
Slice number (slices)	198 - 546	396 - 762
Resolution (mm)	(0.549 - 0.904)×(0.549 - 0.904)×(1.000 - 2.000)	1.000×1.000×1.000

1202. Furthermore, we repeated every experiment five times with different random seeds, which can prove that our method with good performance even with different initializations. We calculated the standard deviation (SD) of the results of the five experiments.

3.2 Implementation details

In this work, our proposed method was implemented using PyTorch and executed on an NVIDIA A100 80G GPU. The model was trained using the SGD optimizer with a batch size of 8. The initial learning rate was set to 0.01, and the poly learning rate strategy was used to adjust the learning rate. Additionally, the number of iterations was set to 30000, and the model obtained from the final iteration was used for the prediction.

3.3 Results

To validate the proposed method's effectiveness, we conducted a comparison with different methods, 3D U-Net,[26] Cross Pseudo-Supervision (CPS),[13] Entropy Minimization (EM),[18] and Mean Teacher (MT).[17] All these methods use 3D U-Net as the backbone. Table 2 shows the segmentation performances of these methods on the intestine segmentation dataset. It can be seen that the CPS method achieved a Dice score of 76.00%, precision of 84.64%, and recall of 73.97%. The proposed method achieved the best Dice score of 80.45%, precision of 84.12%, and recall of

78.84%. For the distance-based metrics, NSD, the proposed method is smaller than other methods when using 13 or 6 labeled cases. We calculated the p-value to assess the validity of our method using the Wilcoxon signed-rank test on the Dice scores. We calculated the p-value to assess the validity of our method using the Wilcoxon signed-rank test on the Dice scores. The evaluation was conducted when the training dataset contained 13 densely labeled cases and 72 unlabeled cases. In the test, we obtained p-values (0.000,0.002,0.000,0.004) that were all < 0.05 between the Dice scores of the proposed method and four existing methods (3D U-Net, CPS, EM, and MT) based on the Wilcoxon signed-rank test. The result indicated the validity of our method. In addition, the proposed method uses stacks of small-sized convolutional kernels instead of large convolutional kernels to reduce the model's parameter. The proposed model, using stacks of small-sized convolutional kernels instead of large convolutional kernels, has 14M parameters (Dice score 80.45%). In comparison, the model[19] employing only large convolutional kernels has 17M parameters (Dice score 78.86%). 3D U-Net[26] has 19M parameters. Compared to the 3D U-Net, the proposed method also has fewer parameters.

Figure 6 shows a boxplot according to the Dice scores of different methods trained with 6 and 13 densely labeled cases. The 3D segmentation results of these methods are illustrated in Fig. 7. In Fig 7, the regions with red color indicate true positives, the regions with green color indicate false positives, and the regions with blue color indicate false negatives. As we use a densely labeled case to present the 3D visualization result, intestine regions lack labels in certain slices. However, these methods can effectively segment unlabeled intestine regions, depicted in gray. We use gray to indicate these unlabeled intestine regions. Figure 8 shows the 2D intestine segmentation results in axial, sagittal, and coronal three planes by different methods. The segmentation results of the proposed method from three different planes are shown in Fig. 9.

Table 2 We compared the quantitative results of the proposed method with four previous methods, including the classical fully supervised method (3D U-Net) and three semi-supervision methods (CPS, EM, and MT) when the model was trained using 6 cases and 13 cases as densely labeled samples. Given an evaluation metric, the null hypothesis of our test is that there is no significant difference in our method outperforming other methods in terms of the metric. We set the significance level at 0.05. We conducted the Wilcoxon signed-rank test on the Dice scores when the model was trained using 13 cases as densely labeled data and 72 cases as unlabeled data. In the table, $\star$ denotes a p-value < 0.05, indicating that our method outperformed the other methods. The p-values were calculated between the proposed method and four previous methods (3D U-Net, CPS, EM, and MT) on the Dice score. We present the SD and highlight the best performance of each evaluation term with a bold font.

Labeled	Methods	Dice (%)	Precision (%)	Recall (%)	NSD
6 Cases	3D U-Net[26]	46.20±7.10	59.28±5.47	44.57±10.77	37.74
	CPS[13]	72.99±1.00	84.26±1.22	67.63±2.13	30.14
	EM[18]	69.45±2.05	84.39±1.06	62.50±3.38	28.45
	MT[17]	69.73±6.26	82.48±1.21	63.78±8.46	30.40
	LMVS-Net (Proposed)	**77.18**±1.87	**84.42**±1.33	**73.19**±3.39	**19.33**
13 Cases	3D U-Net	46.80*±13.96	80.75±3.77	37.80±17.88	28.44
	CPS	76.00*±3.84	84.64±0.69	71.79±0.516	18.81
	EM	76.10*±2.84	85.40±1.33	71.19±4.96	22.18
	MT	74.05*±6.82	**86.17**±1.03	68.20±9.63	23.34
	LMVS-Net (Proposed)	**80.45**±0.71	84.12±0.98	**78.84**±1.76	**18.28**

We also compared the proposed with two previous methods for intestine segmentation (M U-Net, MVS-Net) and show the result in Table 3. M U-Net[27] is an improved 3D U-Net segmentation model based on full-supervision learning, which can utilize the multi-dimensional features from CT volumes. MVS-Net[19] is a semi-supervised method similar to the proposed method. We can see from the result that our method achieves the highest Dice value and the lowest NSD, and M U-Net, an improved 3D U-Net segmentation model based on fully supervised learning, achieves the lowest Dice value and the highest NSD.

In addition, we compare our method with the TotalSegmentator,[28] the quantitative results shown in Table 7, and the qualitative results shown in Fig 10 (3D visualization) and Fig 11 (2D visualization).

Table 3 Results of three intestine segmentation methods. We highlight the best performance of each evaluation term with a bold font, respectively.

Methods	Dice (%)	Precision (%)	Recall (%)	NSD
M U-Net[27]	73.22±4.94	79.89±6.79	70.61±6.65	25.28
MVS-Net[19]	78.50±8.06	**85.88**±8.34	75.06±8.46	22.81
LMVS-Net (Proposed)	**80.45**±0.71	84.12±0.98	**78.84**±1.76	**18.28**

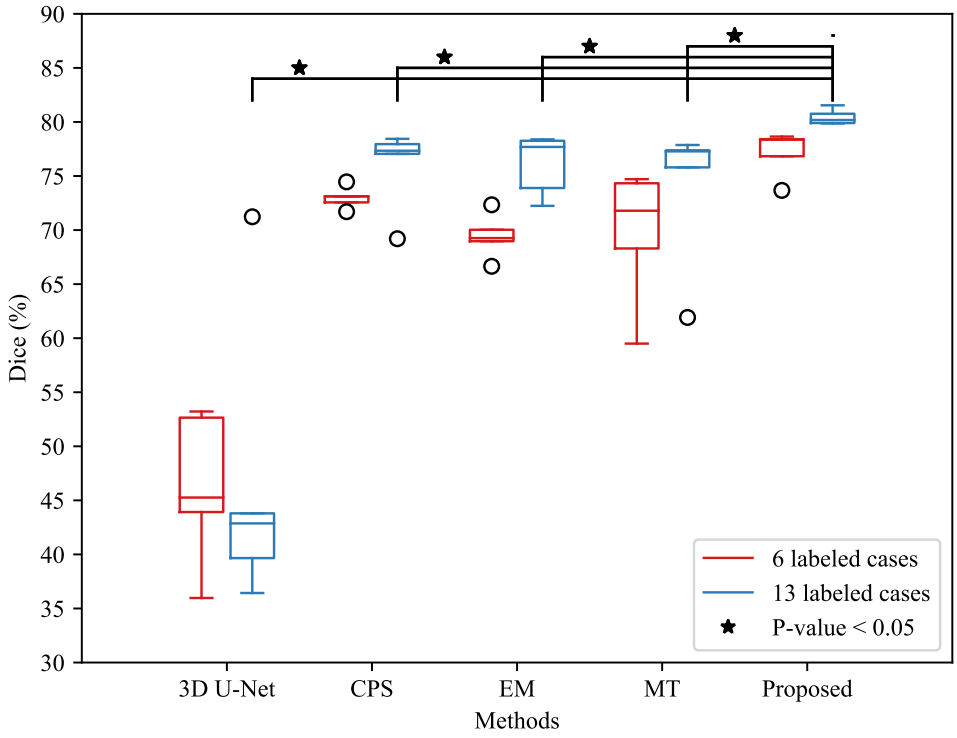


Fig 6 Boxplot of Dice scores of different methods that were trained on datasets containing 6 and 13 densely labeled cases and 72 unlabeled cases. We calculated the p-value using the Wilcoxon signed-rank test on Dice scores when the model was trained on datasets containing 13 densely labeled cases and 72 unlabeled cases, and $\star$ denotes the p-value < 0.05. The red colors indicate the results when we trained the model using 6 labeled cases and 72 unlabeled cases. The blue colors indicate the results when we trained the model using 13 labeled cases and 72 unlabeled cases. The circle indicates outliers.

3.4 Ablation study

Table 4 presents the ablation study result of LMVS-Net. We combine different architectures with bidirectional teaching. The results showed that compared with the 3D U-Net and attention U-Net,

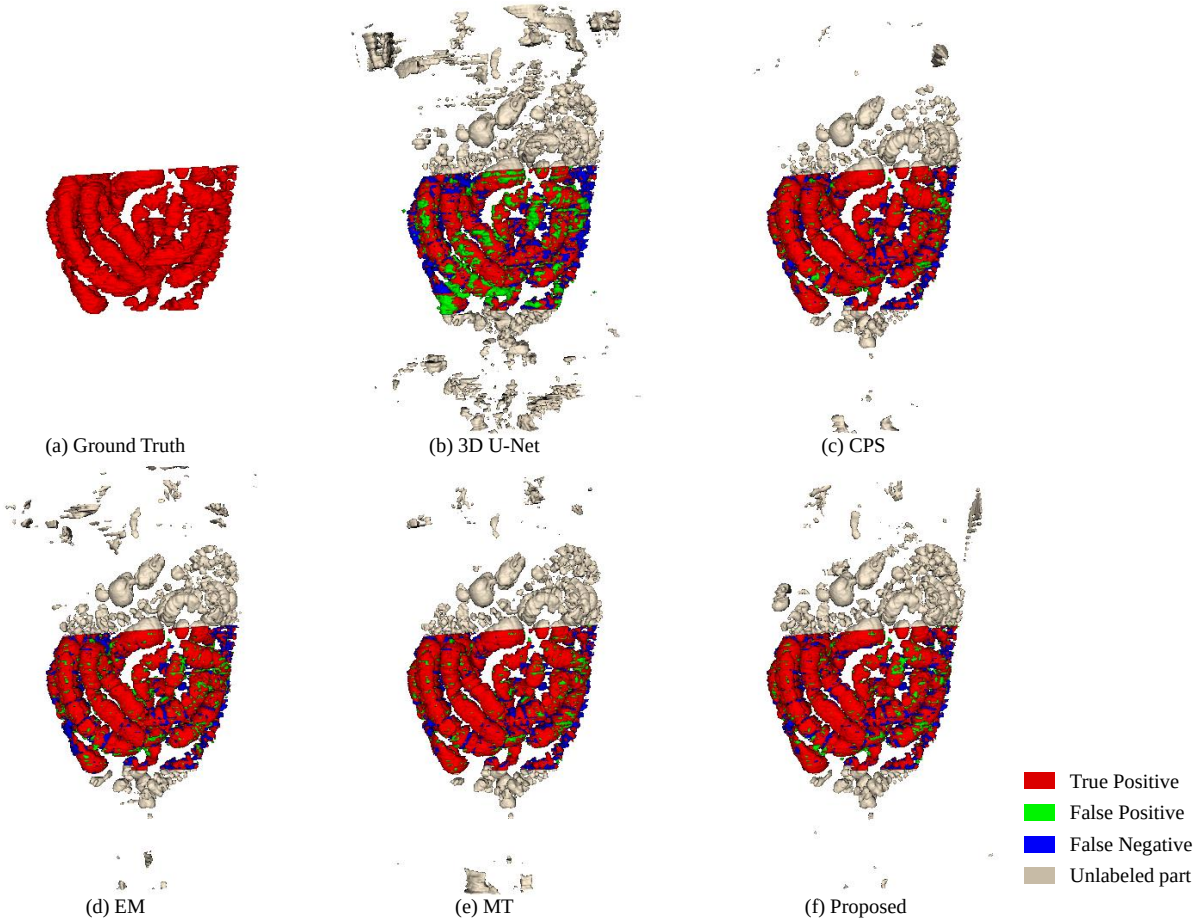


Fig 7 3D intestine segmentation results from different methods. (b)-(e) are the results of four previous methods; (f) is the result of the proposed method.

the LMVS-Net achieved the best performance when combined with bidirectional teaching. The result of the ablation study of the loss function is shown in Table 5. We used the LMVS-Net with different loss functions to train the segmentation model. It can be found that compared with only using Dice loss, clDice loss performs not very well. However, combining the two loss functions performed well in the intestine segmentation task. Since the Dice loss focuses on overall segmentation accuracy, the clDice loss can use critical centerline information to improve performance. We evaluated the influence of weight α in the supervision loss function 2 and the result is shown in Fig 12. We can find that the best segmentation accuracy is achieved when $\alpha = 0.5$. We evaluate

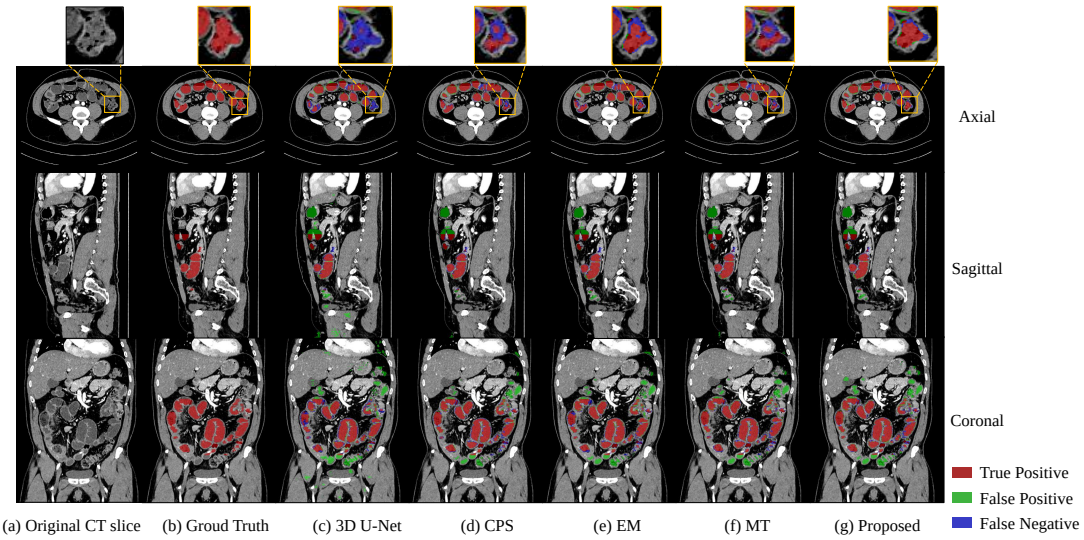


Fig 8 Intestine segmentation results from different methods on axial, sagittal, and coronal three 2D planes, respectively. (a) is the original CT slice, (b) is ground truth, (c) is the result of 3D U-Net, (d) is the result of CPS; (e) is the result of EM; (f) is the result of MT; (g) is the result of the proposed method.

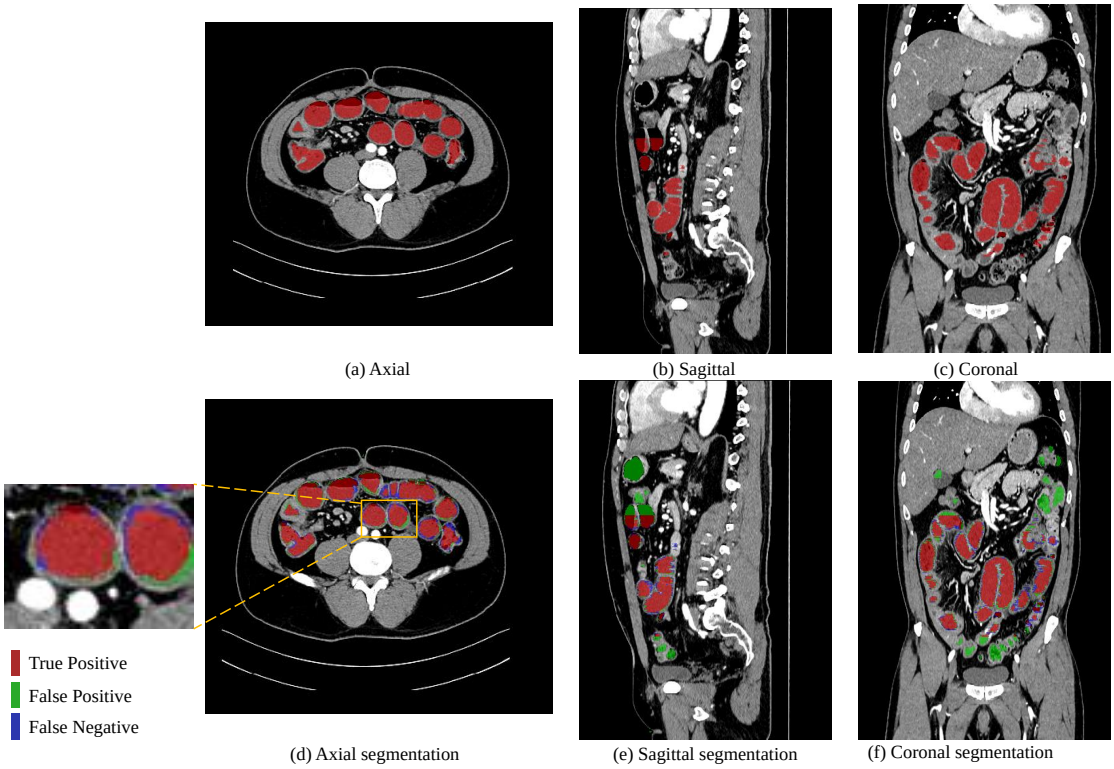
Table 4 To validate the effectiveness of the LVMSNet, we use different models with bidirectional teaching (BT). We highlight the best performance of each evaluation term with a bold font.

Method	Dice (%)	Precision (%)	Recall (%)
3D U-Net with BT	76.00±3.84	84.64±0.69	71.79±5.16
3D U-Net and ATT U-Net with BT	75.80±2.67	85.25±1.48	71.02±4.73
ATT U-Net with BT	71.55±1.87	**85.80**±0.56	64.62±2.57
LMVS-Net with BT	**78.11**±1.30	85.63±1.11	**73.66**±2.59

the effectiveness of the distance weight map and show the result in Table 6. We conducted the experiments with and without distance weight and showed the quantitative results. It can be found that the proposed distance weight achieved higher segmentation accuracy, improved 1.50% Dice score, and 3.70 % in precision.

4 Discussion

Our proposed method segmented most of the intestine regions when we trained it with labeled and unlabeled data. Table 2 shows the results of four previous methods and the proposed method



(a) Axial (b) Sagittal (c) Coronal

(d) Axial segmentation (e) Sagittal segmentation (f) Coronal segmentation

Fig 9 The proposed method's intestine segmentation results are displayed on three planes. The first row represents the ground truth, and the second row displays the corresponding segmentation results.

Table 5 Results of the ablation studies for the loss functions. We use different loss functions in the proposed method. We highlight the best performance of each evaluation term with a bold font.

Method	Dice (%)	Precision (%)	Recall (%)
LMVS-Net with Dice	79.14±1.79	84.52±0.88	76.25±3.09
LMVS-Net with clDice	78.33±1.99	85.05±1.01	74.56±2.59
LMVS-Net with CE and Dice	78.11±1.30	**85.63**±1.11	73.66±2.59
LMVS-Net with clDice and Dice	**79.42**±1.10	84.78±1.43	**76.58**±2.68

on our dataset. Compared with the full-supervision method 3D U-Net, the proposed method and other semi-supervision methods can use the information from unlabeled data and achieve better segmentation results, which denoted that the use of unlabeled data helps reduce the influence of limited labeled data in the intestine segmentation task. Compared with EM and MT the two semi-supervision methods, CPS and the proposed method achieved better performances. EM focuses on making the model more certain about its predictions of unlabeled data and MT focuses on

Table 6 The ablation study of the distance weight (DW). We highlight the best performance of each evaluation term with a bold font.

Method	Dice (%)	Precision (%)	Recall (%)
LMVS-Net without DW	79.66±2.11	**85.17**±1.24	76.78±4.58
LMVS-Net with DW	**80.45**±0.71	84.12±0.98	**78.84**±1.76

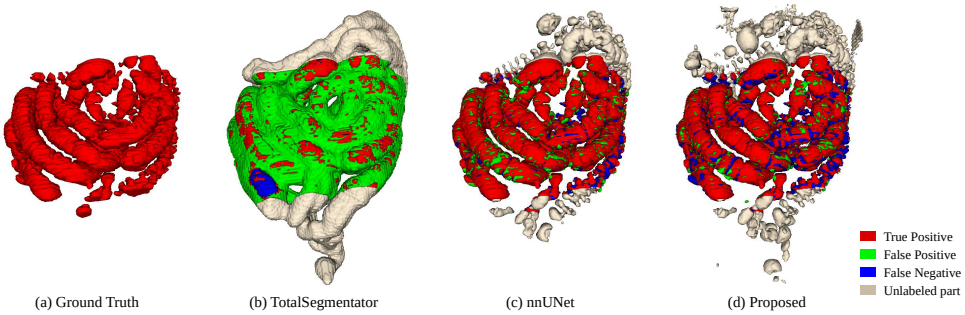


Fig 10 Results of TotalSegmentator, nnUNet, and our method.

enhancing the model's consistency and reducing uncertainty in its predictions of unlabeled data. However, the CPS and the proposed method use the bidirectional teaching strategy and assign pseudo-labels to the unlabeled data to improve model performance. We also investigated the performance between CPS and the proposed method. The CPS method first proposes bidirectional teaching between two 3D U-Nets. The proposed method optimizes the original CPS's backbone, replacing the 3D U-Net with LMVS-Net, and uses bidirectional teaching between two LMVS-Nets. The LMVS-Net achieves multi-scale semantic information from various perceptual fields, which effectively improves the model's segmentation performance.

Furthermore, we explore the model's performance when the model was trained using 6 cases as labeled data and 72 unlabeled cases as unlabeled data, as well as 13 cases as labeled and the same 72 unlabeled cases as unlabeled data. It can be found that the Dice score decreases when we reduce the number of labeled data in the training dataset. We repeated the experiment five times for each method. Figure 6 shows the distribution of five repeated experiments' Dice scores for each

Table 7 Results of TotalSegmentator, nnUNet, and the proposed method. We highlight the best performance of each evaluation term with a bold font, respectively.

Methods	Dice (%)	Precision (%)	Recall (%)	NSD
TotalSegmentator[28]	56.56	50.03	78.55	19.74
nnUNet[29]	81.32	**84.51**	79.82	**15.20**
LMVS-Net (Proposed)	80.45	84.12	78.84	18.28

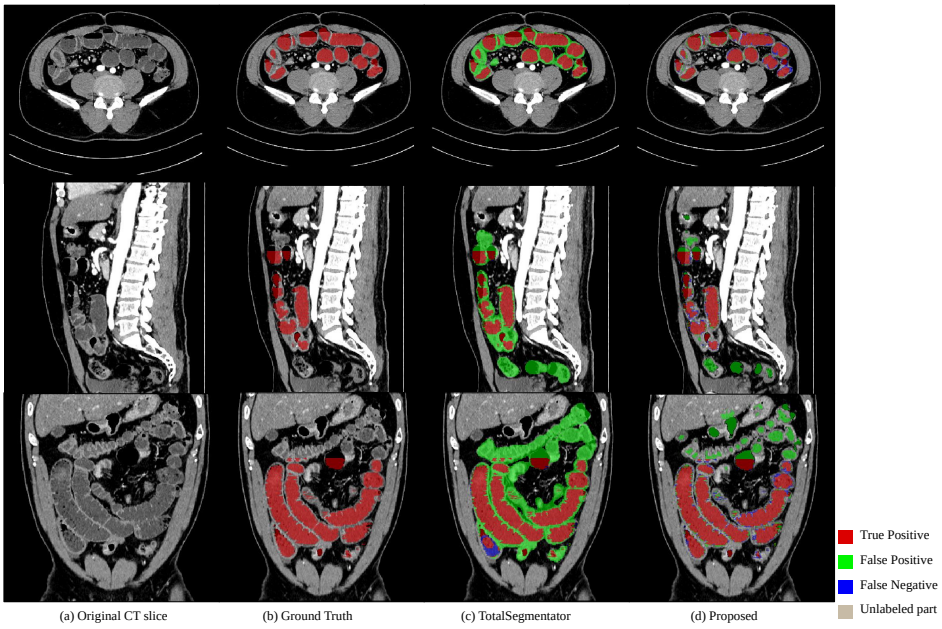


Fig 11 The 2D visualization segmentation result of TotalSegmentator and the proposed method.

method. In the figure, we can find that the proposed method achieves the best Dice score.

Figure 7 shows the qualitative results between various methods when the training dataset contains 13 labeled CT volumes and 72 unlabeled CT volumes. The proposed method can segment more intestine regions with fewer mis-segmentation regions than other methods, especially for 3D U-Net. We can see more detailed results in Fig. 8. The figure shows the intestine segmentation results of different methods in axial, sagittal, and coronal planes, respectively. It can be found that the proposed method can segment more intestine regions with fewer false positives such as the regions in yellow boxes. Figure 9 shows the visualization results of the proposed method in three

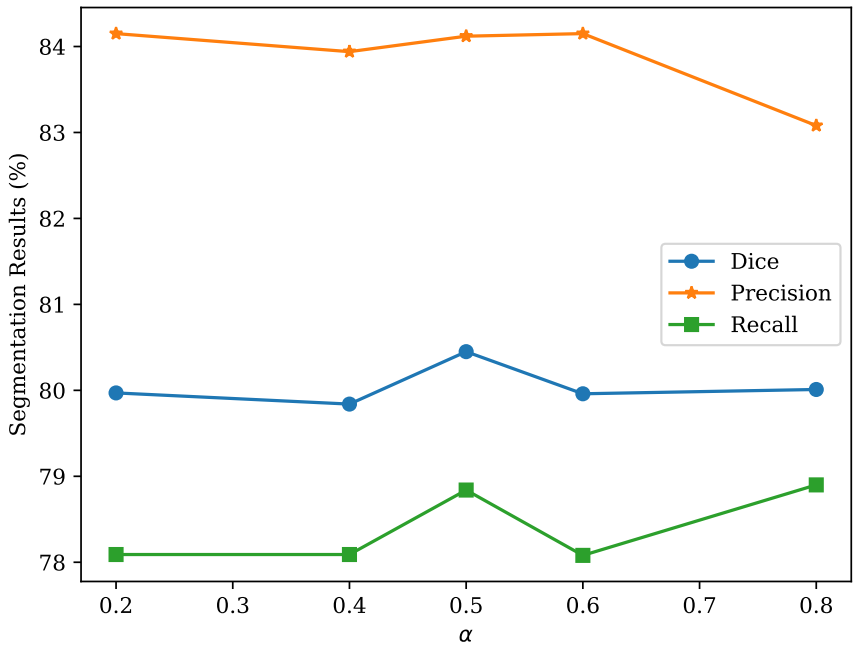


Fig 12 The line chart of different weights (α) in the supervision loss functions, equation 2. We can see that the best Dice score is achieved when $\alpha = 0.5$.

planes. The figure shows that there is still some mis-segmentation in the boundary regions. The reason may be that the intestine exhibits significant shape variations and low contrast compared to surrounding tissues, which presents more challenges.

Compared to TotalSegmentator, our method achieves higher accuracy in intestine segmentation. TotalSegmentator is a pre-trained model designed to segment various organs, including the entire intestinal region, not just the intestinal lumen. As a result, it exhibited lower precision when applied to our testing dataset. The experimental results demonstrate that when the method trained based on a large dataset is transferred to our task, the results are not very good due to the domain gap problem. Our method is more competitive for the intestine segmentation task. nnUNet employs various performance optimization strategies, including hyperparameter tuning and data augmentation, that automatically adjust numerous parameters to select the optimal settings for segmentation tasks. In contrast, our method uses fixed settings. In terms of model size, nnUNet

(88M parameters) is significantly larger than our proposed method (14M parameters). Regarding accuracy, we hypothesized that there would be no significant difference between nnUNet and our method and set the significance level at 0.05. Using the Wilcoxon signed-rank test, we calculated a p-value > 0.05, indicating no significant difference in performance between the two methods. Although our method is less accurate compared to nnUNet, it uses less than a quarter of the parameters. Thus, we can conclude that achieving over 80% Dice score with fewer parameters is a noteworthy success for our approach.

5 Conclusion

This paper proposed an intestine segmentation method, which is based on the LMVS-Net incorporated with bidirectional teaching for intestine segmentation from CT volumes. The proposed method aims to effectively utilize large-scale unlabeled data and important centerline information of tubular organs. Furthermore, the method decreases the influence of pseudo-labels' unreliability by generating weights based on their distance maps. The proposed method demonstrates good segmentation performance and offers competitive results compared to the baseline and previous semi-supervision methods. We compared the segmentation results of several previous methods, including 3D U-Net, Entropy Minimization (EM), and Mean Teacher (MT). In these methods, the 3D U-Net is a full-supervision method that relies on a substantial amount of labeled data, making it not very well compared with the semi-supervision methods for intestine segmentation tasks.

Although significant progress has been made in intestinal segmentation for the proposed method, there is still some space to improve the segmentation accuracy in the boundary part. In future research, we aim to utilize advanced clinical experiments as prior knowledge to reduce annotation costs, particularly for intestine segmentation. The task often involves complicated spatial struc-

tures and belongs to the dense-label-based tasks. By effectively utilizing such prior knowledge, we can enhance the accuracy and efficiency of the segmentation. Furthermore, solving the domain gap problem to transfer the model trained by a large dataset to the intestine segmentation tasks and integrating optimization techniques similar to nnUNet in our method to produce higher performance with semi-supervision also deserves investigation.

Acknowledgments

Thanks for the help and advice from Mori laboratory. Parts of this work were supported by Hori Sciences and Arts Foundation, MEXT/JSPS KAKENHI (17H00867, 22H03703), the JSPS Bilateral International Collaboration Grants, and the JST CREST (JPMJCR20D5). This work was also financially supported by the JST SPRING, Grant Number JPMJSP2125. In addition, the dataset mentioned in the paper are private. The IRB Approval Number of the dataset is Nagoya University School of Medicine IRB Approval Number 2017-0103.

Conflict of Interest

The authors declare that there is no conflict of interest in this paper.

References

1 F. Sinicrope, "Ileus and bowel obstruction," *Holland-Frei Cancer Medicine. 6th edition. Hamilton BC Decker* (2003).

2 K. L. Bower, D. I. Lollar, S. L. Williams, *et al.*, "Small bowel obstruction," *Surgical Clinics* **98**(5), 945–971 (2018).

3 A. Bogusevicius, J. Pundzius, A. Maleckas, *et al.*, "Computer-aided diagnosis of the character of bowel obstruction," *International Surgery* **84**(3), 225—228 (1999).

4 Y. Zhou, L. Xie, E. K. Fishman, *et al.*, "Deep supervision for Pancreatic cyst Segmentation in Abdominal CT Scans," in *Medical Image Computing and Computer Assisted Intervention-MICCAI 2017: 20th International Conference, LNCS 10435*, 222–230, Springer (2017).

5 C. Angermann and M. Haltmeier, "Random 2.5 D U-Net for Fully 3D Segmentation," in *Machine Learning and Medical Engineering for Cardiovascular Health and Intravascular Imaging and Computer Assisted Stenting: First International Workshop, MLMECH 2019, and 8th Joint International Workshop, CVII-STENT 2019*, 158–166, Springer (2019).

6 X. Han, "Automatic liver lesion segmentation using a deep convolutional neural network method," *arXiv preprint arXiv:1704.07239* (2017).

7 K. Rajamani *et al.*, "Segmentation of colon and removal of opacified fluid for virtual colonoscopy," *Pattern Analysis and Applications* **21**(1), 205–219 (2018).

8 W. Zhang and H. M. Kim, "Fully automatic colon segmentation in computed tomography colonography," in *2016 IEEE International Conference on Signal and Image Processing (ICSIP)*, 51–55, IEEE (2016).

9 H. Frimmel, J. Näppi, and H. Yoshida, "Centerline-based colon segmentation for CT colonography," *Medical Physics* **32**(8), 2665–2672 (2005).

10 A. Bert, I. Dmitriev, S. Agliozzo, *et al.*, "An automatic method for colon segmentation in ct colonography," *Computerized Medical Imaging and Graphics* **33**(4), 325–331 (2009).

11 S. Y. Shin, S. Lee, D. Elton, *et al.*, "Deep Small Bowel Segmentation with Cylindrical Topological Constraints," in *Medical Image Computing and Computer Assisted Intervention- MICCAI 2020: 23th International Conference, LNCS 12264*, 207–215, Springer (2020).

12 H. Oda, K. Nishio, T. Kitasaka, *et al.*, "Visualizing intestines for diagnostic assistance of

ileus based on intestinal region segmentation from 3D CT images," in *SPIE Medical Imaging 2020: Computer-Aided Diagnosis*, **11314**, 728–735 (2020).

13 X. Chen, Y. Yuan, G. Zeng, *et al.*, "Semi-supervised semantic segmentation with cross pseudo supervision," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2613–2622 (2021).

14 I. B. Senkyire and Z. Liu, "Supervised and semi-supervised methods for abdominal organ segmentation: A review," *International Journal of Automation and Computing* **18**(6), 887–914 (2021).

15 A. Chebli, A. Djebbar, and H. F. Marouani, "Semi-supervised learning for medical application: A survey," in *2018 International Conference on Applied Smart Systems (ICASS)*, 1–9, IEEE (2018).

16 Y. Zou, Z. Zhang, H. Zhang, *et al.*, "PseudoSeg: Designing pseudo labels for semantic segmentation," *arXiv preprint arXiv:2010.09713* (2020).

17 A. Tarvainen and H. Valpola, "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results," *Advances in Neural Information Processing Systems* **30** (2017).

18 T.-H. Vu, H. Jain, M. Bucher, *et al.*, "ADVENT: Adversarial entropy minimization for domain adaptation in semantic segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2517–2526 (2019).

19 A. Qin, H. Oda, Y. Hayashi, *et al.*, "Intestine Segmentation from CT Volume based on Bidirectional Teaching," in *SPIE Medical Imaging 2024: Image Processing*, (accepted).

20 F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully convolutional neural networks for

volumetric medical image segmentation," in *2016 fourth International Conference on 3D Vision (3DV)*, 565–571, IEEE (2016).

21 S. Shit, J. C. Paetzold, A. Sekuboyina, *et al.*, "clDice-a novel topology-preserving loss function for tubular structure segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 16560–16569 (2021).

22 T. Devries and Graham W. Taylor, "Improved Regularization of Convolutional Neural Networks with Cutout," *CoRR* **abs/1708.04552** (2017).

23 C. Szegedy, V. Vanhoucke, S. Ioffe, *et al.*, "Rethinking the Inception Architecture for Computer Vision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2818–2826 (2016).

24 P. Fadia, J. Shreasiya, and J. Pareek, "Image Colorization," *International Association of Biologicals and Computational Digest* **1**(2), 231–243 (2022).

25 D. G. Bailey, "An Efficient Euclidean Distance Transform," in *Combinatorial Image Analysis*, 394–408, Springer Berlin Heidelberg (2005).

26 Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, *et al.*, "3D U-Net: Learning dense volumetric segmentation from sparse annotation," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2016, LNCS 9901*, 424–432, Springer International Publishing, (Cham) (2016).

27 Q. An, H. Oda, Y. Hayashi, *et al.*, "M U-Net: Intestine Segmentation Using Multi-dimensional Features for Ileus Diagnosis Assistance," in *International Workshop on Applications of Medical AI*, 135–144, Springer (2023).

28 J. Wasserthal, H.-C. Breit, M. T. Meyer, *et al.*, "TotalSegmentator: Robust Segmentation of 104 Anatomic Structures in CT Images," *Radiology: Artificial Intelligence* **5**(5) (2023).

29 F. Isensee, P. F. Jaeger, S. A. Kohl, *et al.*, "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation," *Nature methods* **18**(2), 203–211 (2021).

List of Figures

4 The internal feature maps visualization extracted by the encoder of LMVS-Net. (a-d) are feature maps extracted by each convolutional block. Regions with large feature values are colored in red, and lower values in yellow or blue. Intestinal regions tend to be colored in yellow or blue. Block 1-4 means the convolution block in the encoder part.

5 The overview of the bidirectional teaching between two LMVS-Nets. It is worth noting that the input for each iteration is a pair of labeled and unlabeled patches. Pseudo-labels are generated for unlabeled data, enabling the utilization of large-scale unlabeled data during training.

6 Boxplot of Dice scores of different methods that were trained on datasets containing 6 and 13 densely labeled cases and 72 unlabeled cases. We calculated the p-value using the Wilcoxon signed-rank test on Dice scores when the model was trained on datasets containing 13 densely labeled cases and 72 unlabeled cases, and $\star$ denotes the p-value < 0.05. The red colors indicate the results when we trained the model using 6 labeled cases and 72 unlabeled cases. The blue colors indicate the results when we trained the model using 13 labeled cases and 72 unlabeled cases. The circle indicates outliers.

7 3D intestine segmentation results from different methods. (b)-(e) are the results of four previous methods; (f) is the result of the proposed method.

8 Intestine segmentation results from different methods on axial, sagittal, and coronal three 2D planes, respectively. (a) is the original CT slice, (b) is ground truth, (c) is the result of 3D U-Net, (d) is the result of CPS; (e) is the result of EM; (f) is the result of MT; (g) is the result of the proposed method.

List of Tables

2 We compared the quantitative results of the proposed method with four previous methods, including the classical fully supervised method (3D U-Net) and three semi-supervision methods (CPS, EM, and MT) when the model was trained using 6 cases and 13 cases as densely labeled samples. Given an evaluation metric, the null hypothesis of our test is that there is no significant difference in our method outperforming other methods in terms of the metric. We set the significance level at 0.05. We conducted the Wilcoxon signed-rank test on the Dice scores when the model was trained using 13 cases as densely labeled data and 72 cases as unlabeled data. In the table, $\star$ denotes a p-value < 0.05, indicating that our method outperformed the other methods. The p-values were calculated between the proposed method and four previous methods (3D U-Net, CPS, EM, and MT) on the Dice score. We present the SD and highlight the best performance of each evaluation term with a bold font.

3 Results of three intestine segmentation methods. We highlight the best performance of each evaluation term with a bold font, respectively.

4 To validate the effectiveness of the LVMSNet, we use different models with bidirectional teaching (BT). We highlight the best performance of each evaluation term with a bold font.

5 Results of the ablation studies for the loss functions. We use different loss functions in the proposed method. We highlight the best performance of each evaluation term with a bold font.

6 The ablation study of the distance weight (DW). We highlight the best performance of each evaluation term with a bold font.

7 Results of TotalSegmentator, nnUNet, and the proposed method. We highlight the best performance of each evaluation term with a bold font, respectively.