Dual Uncertainty-Guided Mixing Consistency for Semi-Supervised 3D Medical Image Segmentation

Chenchu Xu, Yuan Yang, Zhiqiang Xia, Boyan Wang, Dong Zhang, Yanping Zhang, and Shu Zhao

***Abstract*—3D semi-supervised medical image segmentation is extremely essential in computer-aided diagnosis, which can reduce the time-consuming task of performing annotation. The challenges with current 3D semi-supervised segmentation algorithms includes the methods, limited attention to volume-wise context information, their inability to generate accurate pseudo labels and a failure to capture important details during data augmentation. This article proposes a dual uncertainty-guided mixing consistency network for accurate 3D semi-supervised segmentation, which can solve the above challenges. The proposed network consists of a Contrastive Training Module which improves the quality of augmented images by retaining the invariance of data augmentation between original data and their augmentations. The Dual Uncertainty Strategy calculates dual uncertainty between two different models to select a more confident area for subsequent segmentation. The Mixing Volume Consistency Module that guides the consistency between mixing before and after segmentation for final segmentation, uses dual uncertainty and can fully learn volume-wise context information. Results from evaluative experiments on brain tumor and left atrial segmentation shows that the proposed method outperforms state-of-the-art 3D semi-supervised methods as confirmed by quantitative and qualitative analysis on datasets. This effectively demonstrates that this study has the potential to become a medical tool for accurate segmentation. Code is available at: https://github.com/yang6277/DUMC.**

***Index Terms*—3D semi-supervised segmentation, contrastive training, dual uncertainty strategy, mixing volume consistency.**

Manuscript received 28 October 2022; revised 18 February 2023; accepted 6 March 2023. Date of publication 17 March 2023; date of current version 11 July 2023. This work was supported in part by the National Natural Science Foundation of China under Grants 62106001 and U1908211, in part by the University Synergy Innovation Program of Anhui Province under Grant GXXT-2021-007, and in part by the Anhui Provincial Natural Science Foundation under Grant 2208085Y19. (Chenchu Xu and Yuan Yang are co-first authors.) Recommended for acceptance by G. SONG. *(Corresponding authors: Zhiqiang Xia and Boyan Wang.)*

Chenchu Xu is with the School of Computer Science and Technology, Key Laboratory of Intelligent Computing and Signal Processing of Ministry of Education, Anhui University, Hefei, Anhui 230093, China, and also with the Institute of Artificial Intelligence, Hefei Comprehensive National Science Center, Hefei 230088, China (e-mail: cxu332@gmail.com).

Yuan Yang, Yanping Zhang, and Shu Zhao are with the School of Computer Science and Technology, Key Laboratory of Intelligent Computing and Signal Processing of Ministry of Education, Anhui University, Hefei, Anhui 230093, China (e-mail: e21201053@stu.ahu.edu.cn; zhangyp2@gmail.com; zhaoshuzs2002@hotmail.com).

Zhiqiang Xia is with the Anhui Provincial Hospital, University of Science and Technology of China, Hefei 230026, China (e-mail: ahslyyxzq@163.com).

Dong Zhang is with the Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC V6T 1Z4, Canada (e-mail: zhangdong9612@gmail.com).

Boyan Wang is with the Department of Computer Science and Technology, Tsinghua University, Beijing 100190, China (e-mail: wby000000@mail.tsinghua.edu.cn).

Digital Object Identifier 10.1109/TBDATA.2023.3258643

I. Introduction

3D semi-supervised medical image segmentation in medical image analysis has been a subject of continuous interest to numerous researchers. This interest has been earned by 3D segmentation abilities in facilitating the planning and preparation of treatment for physicians by providing quantitative analysis of tissue volume, localization of pathologically modified tissues and visualization of anatomical structures amongst many other tasks. Applying 3D semi-supervised segmentation to the large amount of unlabeled data available in the medical field would reduce the time-consuming and laborious work of manually delineate the target area of medical images in clinical practice. Superior to 2D segmentation, 3D methods [1] provide rich details of organs, tissues and lesions, thereby revealing more disease characteristics which enhances the effectiveness of diagnosis.

3D semi-supervised medical image segmentation techniques can be categorized into three broad types. First, the Consistency Regularization, which on the basis of supervised segmentation and consistent regularization-based approaches, adds distinct disturbances to the same images and enforces similar predictions. These techniques generate segmentation model that are used to unlabeled datasets, which can achieve semi-supervised segmentation [2], [3], [4]. [5] proposed a consistency regularization-based method, which advocated interpolation consistency to promote the segmentation of interpolation of two unlabeled data sets to be consistent with the interpolation of segmentation maps of those data sets. Second, the Pseudo label-based method, is a 3D segmentation method which employs a rough segmentation model trained on labelled datasets usually having a few numbers of datasets [6], [7], [8]. This model enables the model to generate fake labels (called pseudo labels) to the unlabeled datasets, thereby facilitating the training of all datasets. Tarvainen and Valpola [9] suggested a Pseudo label-based student-teacher paradigm in which the teacher model would monitor the student model, while a cross pseudo supervision study was performed to encourage one segmentation network to supervise the other segmentation network and vice versa [10]. Lastly, the generative model-based method [11], [12], [13] [14] trains generative adversarial networks to generate augmented labelled datasets with a comparable distribution to the original labelled datasets. These augmented datasets increase the number of labelled datasets required for segmentation model training. Bowles et al. [15] developed a novel method for extracting more information from a dataset by creating synthetic samples that resemble real images.

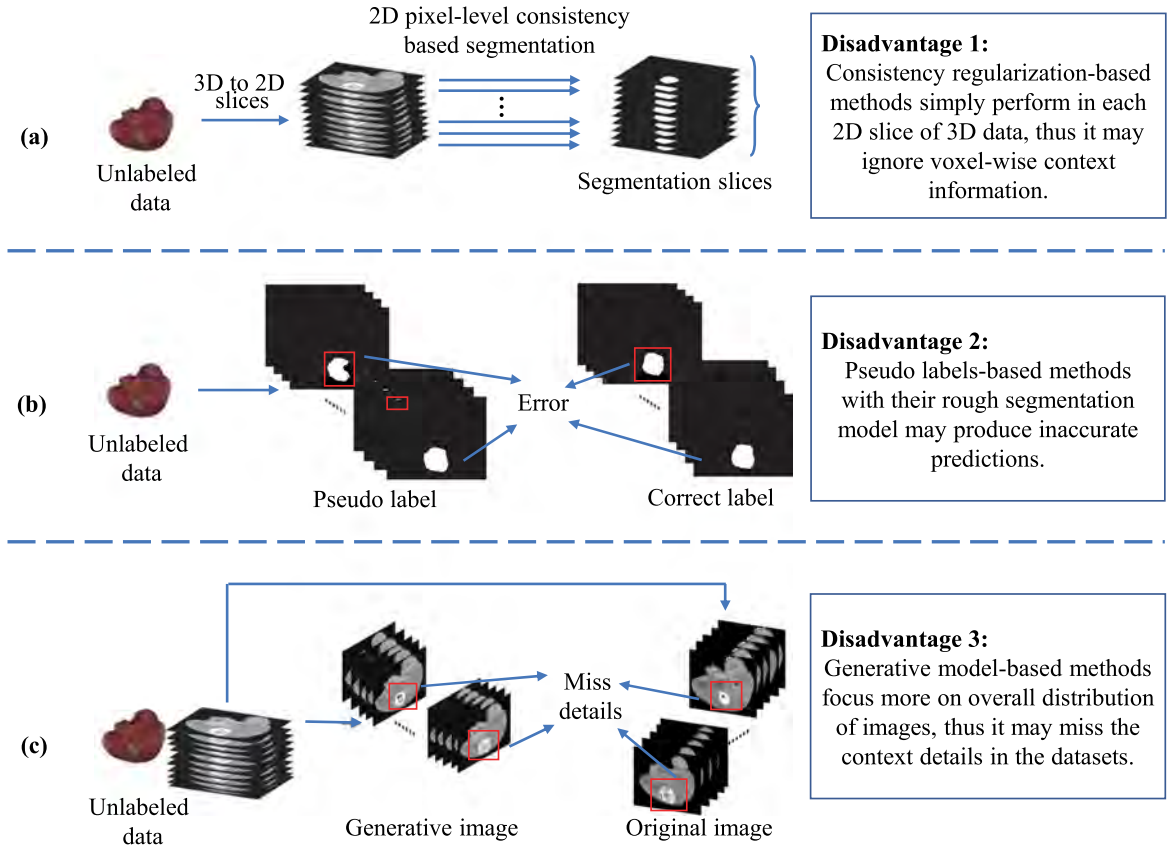


Fig. 1. Current 3D semi-supervised segmentation methods suffer from three disadvantages: (a) Consistency regularization ignores the volume-wise context information; (b) Pseudo labels-based method may be inaccurate; (c) Generative model-based method may miss some details.

The aforementioned 3D segmentation strategies have contributed significantly to the advancement of 3D semi-supervised image segmentation. However, semi-supervised 3D medical image segmentation is still a challenging task since existing regularization-based algorithms simply operate on 2D slices of a 3D data only. As illustrated in Fig. 1(a), it may neglect volume-wise context information and cause the model to produce boundary error in segmentation and gap between slices in 3D modelling. Similarly, as shown in Fig. 1(b), label-based approaches with their rough segmentation model may make it challenging to guarantee the accuracy of pseudo labels for unlabeled data. Continuous propagation of erroneous pseudo labels across the model may cause inaccurate segmentations, resulting in poor model performance. Generative model-based methods focus more on the overall distribution of augmented datasets to be the same as the original datasets as depicted in Fig. 1(c), but may underestimate context details (e.g., edge information or some complex area) in the datasets, which could increase the difficulty to accurately learn pathological tissues, such as tumors and lesions.

In order to address these challenges, we propose a Dual Uncertainty-guided Mixing Consistency (DUMC) network for reliable semi-supervised 3D image segmentation to eliminate the need for manual annotations. DUMC performs data augmentations on the basis of supervised learning with labelled data by employing contrastive similarity to ensure high similarity between the augmented data and original data in order to maintain the invariance of data augmentation. DUMC then generates a robust disturbance that mixes the volume of unlabeled 3D data to enhance the representation of diverse features. In the meantime, DUMC uses two distinct models (a network with transpose convolution decoder and a network with trilinear interpolation decoder) to calculate the uncertainty and uses low uncertain areas to restrain the consistency between this strong disturbance of two unlabeled data in order to enhance the network's robustness and generalizability. DUMC consists of a Contrastive Training Module (CTM) which translates various augmented 3D data into 1D feature vectors, then calculates their similarity to minimize the disparity between augmented and original data, thereby preserving the invariance of various augmentations. This module can make it easy to identify pathological tissues. The Dual Uncertainty Strategy (DUS) calculates uncertainty based on two slightly distinct models in order to select the confident parts, which then guides the subsequent segmentation. The Mixing Volume Consistency Module (MVCM) is the last connected module which promotes the consistency of sub-volume (a part of the entire image) mixing before and after segmentation and learns volume-wise context information.

In summary, the main contributions of this paper are as follows:

- This study presents a novel method for improving 3D semi-supervised segmentation by reducing model error, utilizing unlabeled data and learning volume-wise context information more effectively. The method achieved a new state-of-the-art level of performance on the segmentation of two human organs. In addition, the proposed methods' code and dataset have been made publicly available so that the experiment can be repeated.
- Dual uncertainty technique is proposed to minimize the propagation of error in segmentation. The technique generates two segmentation uncertainty based on two types of up-sampling modes and constrains them in order to select the optimal. Consequently, it allows for flexibility when calculating uncertainty, ensuring the stability of the certainty regions.
- Mixing volume consistency is presented to enable volume-wise context information to be incorporated into consistent regularization and to exploit unlabeled data in an effective manner. The consistency produces volume-wise perturbations across and within unlabeled data, allowing the segmentation network to increase its performance by maintaining the consistency between perturbations through both global similarity and local variation while dealing with unlabeled data.

II. Related Work

A. Uncertainty-Based Method for 3D Semi-Supervised Segmentation

The semi-supervised method has a significant impact on the segmentation of 3D medical images, which has largely solved the need for mutual laborious annotations in diverse tissue segmentation applications. To the precise, the uncertainty-based semi-supervised approach has become one of the most popular domains, because of its promising effect of limiting data error by computing uncertainty based on the mean value of multiple forward propagations. The uncertainty-based semi-supervised technique limits the propagation of error throughout the network by selecting less uncertain (i.e., more confident) regions to guide subsequent segmentation. However, most current methods derive uncertainty from a single model according to Luo et al. Yu et al. and etc. [8], [16], [17], [18], [19], [20], [21], [21], [22], [23] which may disregard the noise produced by the model. Nevertheless, since the model may not be sufficiently stable at the beginning of the training stage to produce accurate guidance, the proposed DUMC method employs dual uncertainty from two networks for improved guidance, which can form a mutual restriction relationship to guide mixing volume consistency (Section II-B) to select more reliable parts to guide consistency and prevent error propagations through the model.

B. Consistency Regularization in Semi-Supervised Segmentation

The method of consistency regularization method is commonly employed in image segmentation. However, these techniques may ignore some factors, such as intra-volume context information, which could be experienced using 2D slices instead of 3D images to generate segmentation [24], [25]. Numerous approaches employ the use of a single image to enforce the consistency of their disturbances in accordance to Chen et al. [10], [26], [27], which may lead to inaccurate segmentation outcome because of its insufficiency in capturing inter-volume context information, a factor that also limits consistency regularization. The ignored integral volume context information is an additional factor that limits consistency regularization. Studies by Basak et al. [5] have considered the utilization of random interpolation based on a voxel-by-voxel basis which might disrupt inter-voxel relationships during training (ICT). Therefore, this study suggests mixing volume consistency, which ensures mixing consistency before and after segmentation. The proposed mixing approach is based on blending of two unlabeled data, this provides the advantage of capturing and retaining single-data intra-volume context information. Second, mixing two unlabeled data provides batch-size inter-volume context information. Finally, the sub-volume of one original data is replaced with the identical volumetric location of another, preserving the context information for the volume as a whole.

C. Contrastive Learning in Semi-Supervised Segmentation

Contrastive learning [28], [29], [30], [31] has emerged as one of the most promising semi-supervised segmentation techniques. Hu et al. [24] utilized a self-supervised global contrastive learning step and a supervised local contrastive learning step to capture image-level and pixel-level features, while Zhao et al. [32] propose a cross-level contrastive learning strategy to improve local feature representation in semi-supervised medical image segmentation. Contrastive learning can cause comparable samples to move closer together and different samples to move farther apart, which can can help maintain the invariance of features or representations. However, its usage in 3D image segmentation is rare. The proposed CTM in our study adds a 3D average pooling layer to transform the 3D features into $1 \times 1 \times 1$, allowing them to be utilized effectively in 3D image segmentation. CTM also employs contrastive similarity (SimCLR [33]) of augmented data and original data within the DUMC approach to maintain the invariance of augmentation from the same 3D images, enabling it accurately to detect challenging pathological features (e.g., lesions and tumors).

III. Methodology

DUMC is a semi-supervised segmentation method. DUMC takes labeled images x^l and unlabeled images x^u (x_1^u and x_2^u) as input and output their segmentations. As shown in Fig. 2, DUMC contains three parts which includes the Contrastive Training Module (CTM, Section III-A). CTM calculates the similarity between original images and augmented images to keep the invariance of augmented data. It takes original images x and their augmentations $\tilde{x}$ as input and outputs new augmented images. The second part is Dual Uncertainty Strategy (DUS, Section III-B). DUS uses uncertainty from two different networks and chooses more confident areas to guide the following segmentation task. It takes labeled images x^l and mixed unlabeled images

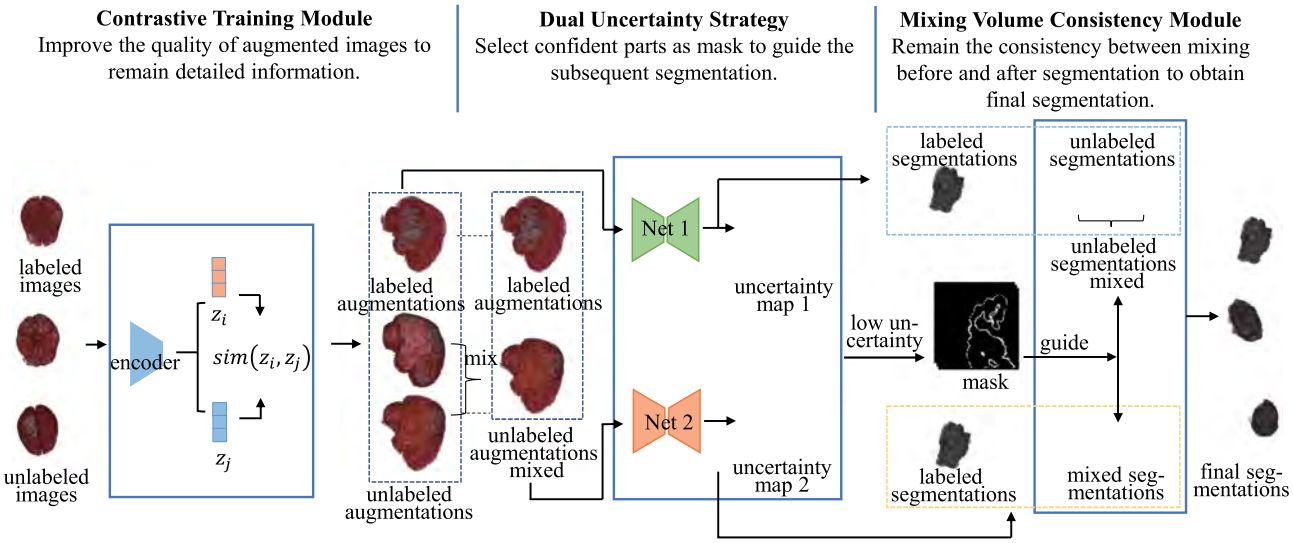


Fig. 2. The proposed Dual Uncertainty-guided Mixing Consistency (DUMC) network consists of Contrastive Training Module (CTM) that can improve the quality of augmented images by keeping invariance of data augmentation. The Dual Uncertainty Strategy (DUS) can select more confident areas by calculating dual uncertainty from two different models. The Mix Volume Consistency Module (MVCM) also, can retain volume-wise information by using dual uncertainty to guide the consistency between mixing after segmentation and before segmentation for final segmentation. Note that Net 1 (main network) is the UNet/VNet with the up-sampling of trilinear interpolation. And Net 2 (auxiliary network) is the UNet/VNet with up-sampling of transpose convolution.

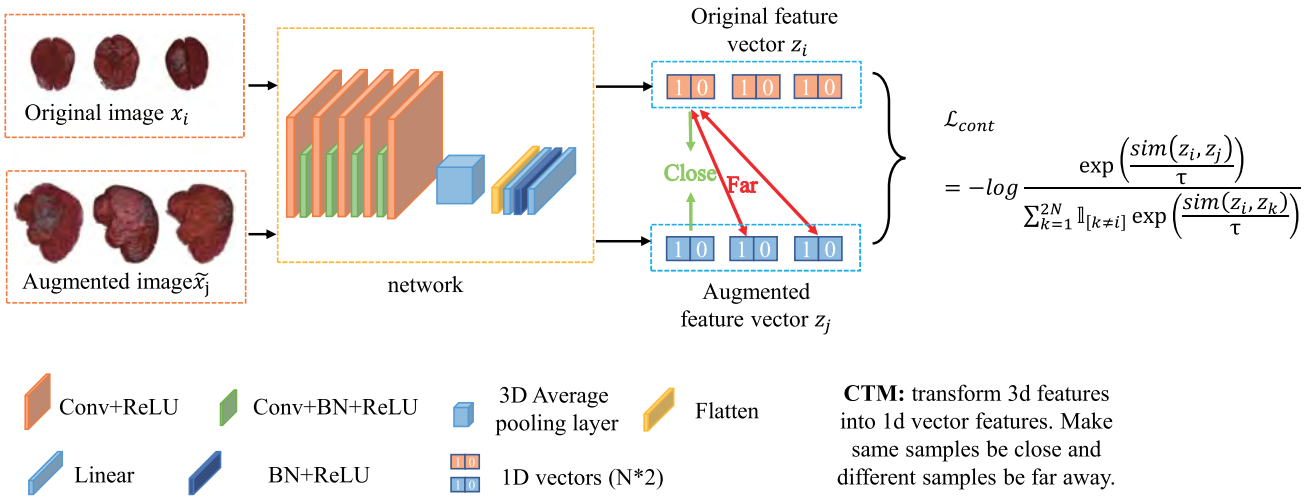


Fig. 3. CTM transforms 3d features into 1d vector features. Clusters similar samples away from dissimilar samples.

$x^{u'}$ (mixing x_1^u and x_2^u to form $x^{u'}$) as input and outputs their masks (i.e., more confident parts) and segmentations. The third part is Mixing Volume Consistency Module (MVCM, Section III-C). MVCM encourages the mixing consistency between before and after segmentation to produce better segmentation results. It takes masks and segmentations as input and uses masks to guide the mixing consistency, thereby causing accurate output of the final segmentation of all images.

A. Contrastive Training Module (CTM) to Ensure Invariance of Data Augmentation

The proposed CTM calculates the contrastive similarity between original data and their augmentations that forces similar samples to lie closer and less similar samples to lie away, thereby ensuring the accuracy of augmented data. CTM has positive pairs of x_i (including labeled data and unlabeled data) and its according augmented data $\tilde{x}_j$ when $i = j$. CTM can obtain low-resolution 3D features after four 3×3 convolution layers and four 2×2 convolution layers. Then CTM passes the low-resolution 3D features through 3D average pooling, flatten, and MLP layers to get $z_i = g(x_i)$ and $z_j = g(\tilde{x}_j)$, which can project the features into potential space, where $g(\cdot)$ denotes the all above network layers. Its purpose is to pull the pairs of voxels that come from two augmentations of the same image closer, while the voxels from different images are pushed away, more details are shown in Fig. 3. In order to achieve this, CTM uses a contrastive loss from SimCLR [33] which can be expressed

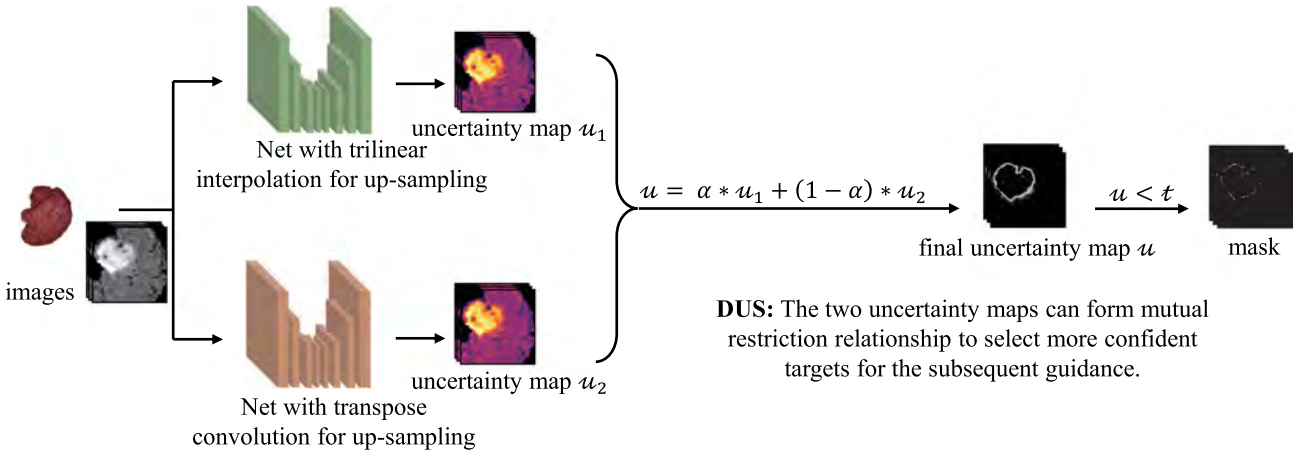


Fig. 4. The proposed DUS for the two uncertainty maps which can form a mutual restriction relationship to select more confident targets for the subsequent guidance.

mathematically as:

$$L_{cont} = -\log \frac{exp(sim(z_i, z_j)/\tau)}{\sum_{k=1}^{2N} I_{[k \neq i]} exp(sim(z_i, z_k)/\tau)} \tag{1}$$

where $I_{[k \neq i]} \in (0, 1)$ denote an indicator function evaluating to 1 if $k \neq i$ and τ denote a temperature parameter; $sim(u, v) = u^T v/||u||||v||$ denote the dot product between l_2 normalized u and v (i.e., cosine similarity).

CTM uses a 3D average pooling layer and MLP to transform 3D features into 1D feature vectors and calculate the similarity between original images and augmented images in order to keep similar images moving closer together and other images moving farther apart. This allows CTM capture more pathological features by maintaining good invariance of augmented data.

B. Dual Uncertainty Strategy (DUS) to Avoid Error Propagation and Guide Subsequent Segmentation Task

The proposed DUS calculates the dual uncertainty from two networks with different decoders to find less uncertain (i.e., high confidence) parts, thereby guiding the subsequent segmentation task. This DUS enables the two models to gradually learn from a more confident target, which can prevent the noise or error propagation of pseudo labels. Specifically, as shown in Fig. 4, DUS performs N stochastic forward passes with random dropout. N is 8 and means the 8 forward without gradient. In addition, The proposed DUS can respectively get two sets of probability vectors $P_{t\,t=1}^{T}$ and $Q_{t\,t=1}^{T}$ from the two models $F1(\cdot)$ and $F2(\cdot)$. The proposed DUS chooses the predictive entropy as the metric to estimate the uncertainty and uses the two uncertainties from different networks to calculate the dual uncertainty, which can be summarized as the following equations:

$$\mu_c = \frac{1}{T} \sum_t P_t^c \tag{2}$$

$$\nu_c = \frac{1}{T} \sum_t Q_t^c \tag{3}$$

$$u_j = \alpha \left(-\sum_c \mu_c \log \mu_c \right) + (1-\alpha) \left(-\sum_c \nu_c \log \nu_c \right), \tag{4}$$

where $P_t^c(Q_t^c)$ is the probability of the cth class in the tth time prediction; α is a hyperparameter balancing two uncertainties from two different networks to get mutual restriction relations.

DUS selects certain parts of predictions as masks by the guidance of the dual uncertainty, which can be summarized as:

$$Mask = I(u_j < t) \tag{5}$$

where t is a threshold to select certain predictions. I denotes the indicator function. And u_j can be acquired from (4).

In summary, DUS calculates two uncertainties and forms a mutual restriction relation of two networks to select more confident areas of predictions, which is not limited to effectively guiding mixing volume consistency in Section III-C but also guide the original consistency. This DUS from two different networks can produce more accurate segmentations than from the same or single network.

C. Mixing Volume Consistency Module (MVCM) to Fully Use Voxel-Aware Information

MVCM randomly mixes two unlabeled data before segmentation and after segmentation and enforces their consistency to enhance the acquisition of volume-based context information for better segmentation. Specifically, as shown in Fig. 5, MVCM randomly copies one sub-volume from one unlabeled image x_i^u and pastes the sub-volume directly into another unlabeled image x_j^u to get a new mixed image $x^{u\prime}$, where x_i^u and x_j^u belong to unlabeled images in the same batch size. To achieve this, MVCM designs a volume ν, which is a binary image of the same size as the two original unlabeled images. MVCM can generate a mixed unlabeled image through the following equation:

$$x^{u\prime} = V \odot x_i^u + (1 - V) \odot x_j^u \tag{6}$$

where $\odot$ denotes the element-wise product. Then MVCM enforced the segmentation of a mix of two unlabeled data to be

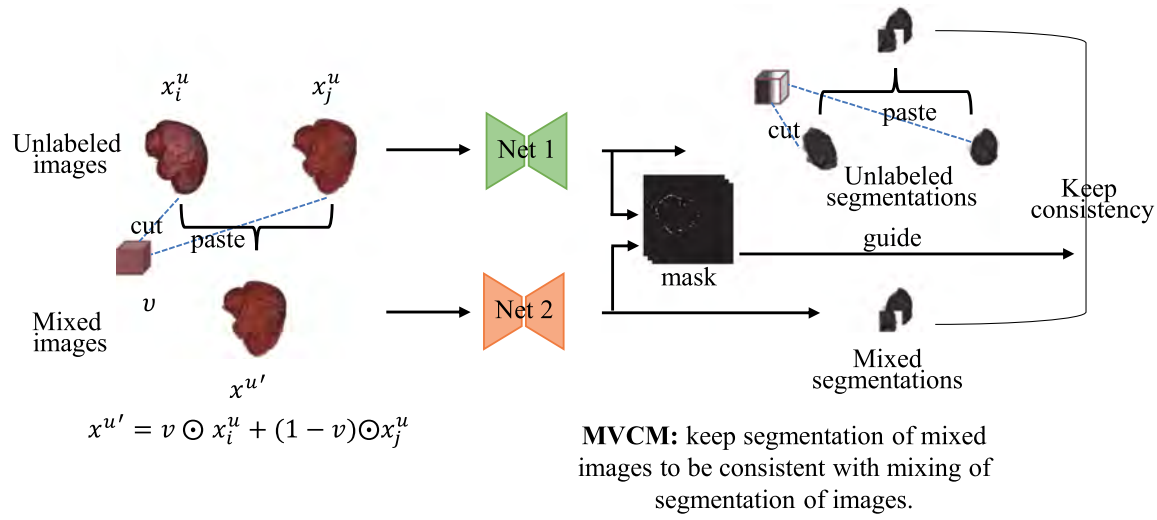


Fig. 5. MVCM keeps segmentation of mixed images to be consistent with mixing of segmentation of images to learning more voxel-aware context information for more accurate result.

consistent with the mix of segmentation of those data. Therefore, the MVCM can be summarized as:

$$F1(x^{u\prime}) \sim V \odot F2(x_i^u) + (1-V) \odot F2(x_j^u) \tag{7}$$

$$F2(x^u) = V \odot F2(x_i^u) + (1-V) \odot F2(x_j^u) \tag{8}$$

where $F1(\cdot)$ denotes the auxiliary network with a decoder using transposed convolution for up-sampling, and $F2(\cdot)$ denotes the main network with a decoder using trilinear interpolation for up-sampling. MVCM finally creates the mixing volume consistency loss between mixing before segmentation $F1(x^{u\prime})$ and after segmentation $F2(x^u)$ denoted in equation from (7) and (8) respectively, the created mixed volume consistency loss is then represented as:

$$L_{con1} = L_2(F1(x^{u\prime}), F2(x^u)), \tag{9}$$

where L_2 denotes the Mean Squared Error (MSE) loss. On the basis of (9) and (10), DUS-guide consistency loss L_{con} could be expressed as:

$$L_c = \beta L_{con1} + (1-\beta) L_2(F1(x^u), F2(x^u)), \tag{10}$$

$$L_{con} = \frac{I\left(\sum_v Mask * L_c\right)}{I\left(\sum_v Mask\right)} \tag{11}$$

where β is a hyperparameter that balances the mixing consistency and original consistency to get more features of unlabeled images; $(\cdot)$ denotes the indicator function. I is indicator function. Finally, the main network (trilinear interpolation for up-sampling) is used to produce the final segmentation. In addition, MVCM creates a new mixed unlabeled image to get more feature information in the process of segmentation, which is derived from relations of nearby voxels between a single unlabeled image and two unlabeled images.

MVCM builds three-fold context information for 3D images. The first is the intra-volume context information, which is built by the saved intra-volume context information when copying sub-volumes, because it preserves adjacent voxels relationships. The second is the inter-volume context information. This is built by the process of mixing unlabeled images which keeps the inter-volume context information of different unlabeled images in those mixed images. Hence increasing the detection of pathological features. The third is the integral volume context information. The mixed images can remain the integral volume of context information due to the fact that they have the same size as the original images. Furthermore, the original consistency from two network branches can ensure that mixing images would not miss to capture all important details.

D. Overall Training Objective

The proposed overall training objective is to minimize supervised losses and unsupervised losses. On the labeled dataset, supervised losses L_{sup} include the linear composition of cross-entropy loss and dice loss. On the unlabeled dataset, unsupervised losses include consistency loss L_{con} and contrastive loss L_{cont}. So, the overall loss function is:

$$L = L_{sup} + \lambda L_{con} + \sigma L_{cont} \tag{12}$$

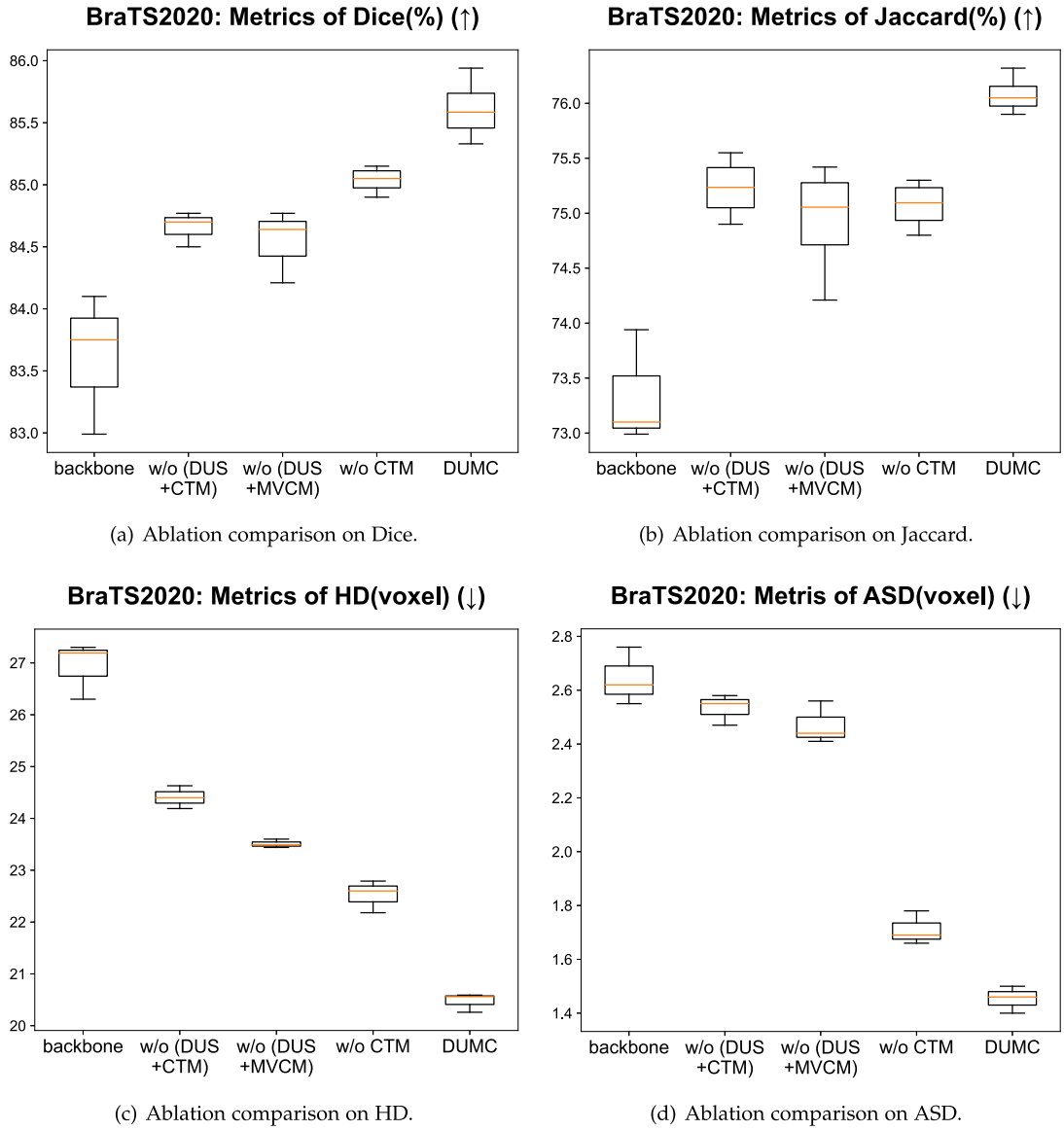
where λ, σ are hyperparameters that balance each term.

IV. Experimental Settings

A. Dataset and Metrics

Two publicly available 3D MRI datasets were used to evaluate the performance of the proposed method. The Brain Tumor Segmentation 2020 (BraTS2020) dataset [38] consist of T1, T2, T1 contrasted-enhanced (T1ce) and Flair brain MR images from 396 patients provided for the 2020 brain tumor segmentation challenge. 300 volume of dimension $96 \times 96 \times 96$ were utilized for training, whereas 69 were used for testing [39], [40]. The 2018 Left Atrial Segmentation [1] (LA2018) database had 100 MRI images, 80 images of size $112 \times 112 \times 80$ were used for training, 20 were used testing. The volumes were preprocessed

[1]http://atriaseg2018.cardiacatlas.org



(a) Ablation comparison on Dice.

(b) Ablation comparison on Jaccard.

(c) Ablation comparison on HD.

(d) Ablation comparison on ASD.

Fig. 6. The ablation comparison result on BraTS2020 shows that our CTM and DUS and DUMC can improve the performance of segmentation respectively. The combination of the three modules achieves the best performance. (a) Ablation comparison on Dice. (b) Ablation comparison on Jaccard. (c) Ablation comparison on HD. (d) Ablation comparison on ASD.

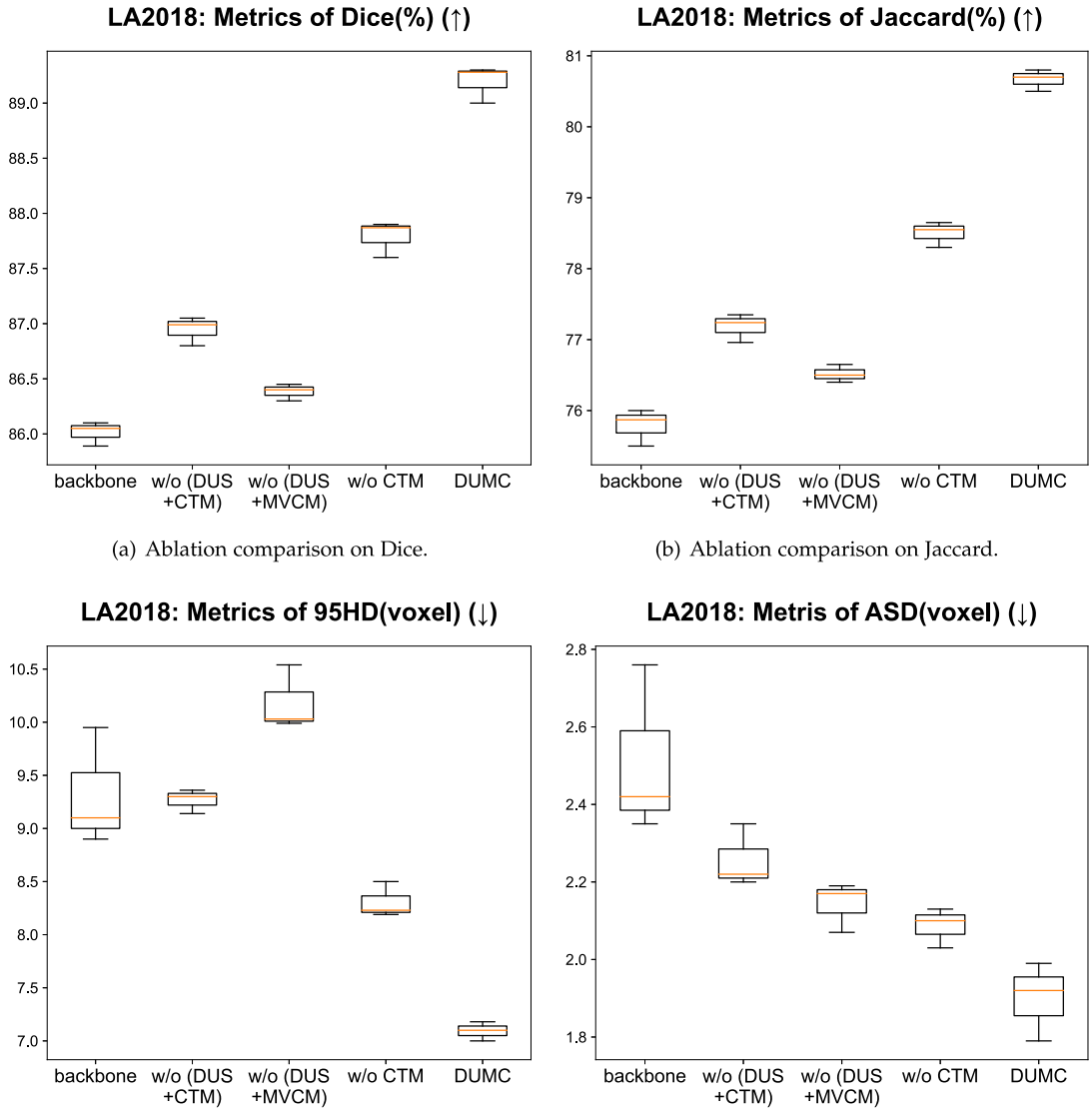
using the method described by [41], with only the flair sequences that were cropped in the center of the tumor area being utilized.

Dice, Jaccard, the Hausdorff Distance (HD)/95% Hausdorff Distance (95HD), and the average surface distance (ASD) were used to quantitatively evaluate segmentation. Dice and Jaccard are measured in percentage, while ASD and 95HD/HD are measured in voxels.

B. Implement Details

The evaluation of the proposed approach on the two datasets were as follows:

BraTS2020: Segmentation was facilitated by a 3D-UNet [42] equipped with two distinct decoders (main network: decoder with up-sampling of trilinear interpolation; auxiliary network: decoder with up-sampling of transpose convolution). As part of our preprocessing, we zeroed the mean and standardized the variance, and we cropped the image to focus on the brain. Two variants of the 3D-UNet were utilized for training. Transposed convolution is used in one decoder for up-sampling. Similar to [41], we trained the network for 30000 iterations, and one of the techniques we used was tri-linear interpolation for up-sampling. The batch size was 4, including two labeled data and two unlabeled data. The input was a grid of $96 \times 96 \times 96$ randomly cropped volumes, and output was an enhancement of the data via 2D rotation and flip. Network parameters were updated using the SGD optimizer (weight decay=0.0001, momentum=0.9), and the initial learning rate was 0.05, which decreased using the ploy learning rate scheduler $(1 - iter/\max\_iter)^{0.9}$. The hyperparameters α, β, τ, λ, σ are respectively set as 0.8, 0.8, 0.5, 0.07,



(a) Ablation comparison on Dice.

(b) Ablation comparison on Jaccard.

(c) Ablation comparison on 95_HD.

(d) Ablation comparison on ASD.

Fig. 7. The ablation comparison result on LA2018 shows that our CTM and DUS and DUMC can improve the performance of segmentation respectively. The combination of the three modules achieves the best performance.

0.01. In terms of equation, is (5) t-value is not a constant. $t = 0.75 + 0.25 * (1 - iter/\max\_iter) * np.\log(2)$, As shown in Table II, the initial value of t is 0.75 and is then incremented using the Gaussian formula, which can acquire the best result. About contrastive learning, the MLP is simply designed with two linear layers and a batch-norm layer. For testing, a sliding window of size 96× 96× 96 with a fixed stride (64, 64, 64) was employed to extract patches. The post-processing strategy (NMS) [35] was used to remove isolated extraneous regions.

LA2018: 3D VNet [43] by Milletari was used as backbone for segmentation. The network was trained for 6000 iterations, having the initial learning rate as 0.03, and afterwards divided 10 at every 2500 iterations. The cropped size was set as 112×112×80 and Adam optimizer was used to update the network parameters. For testing, a sliding window of size 112×112×80 with a fixed stride (18, 18, 4) was set to extract patches. All other experiment settings are the same as that of the BraTS2020 evaluation.

V. Experimental Results

DUMC produces high performance with Dice of 85.94%, 89.28% with only 10% labels in brain tumor segmentation and left atrial segmentation respectively, which demonstrates this method is effective for 3D semi-supervised segmentation.

A. Comparison With Some Approaches

BraTS2020: It can be seen from Table I and Fig. 8 that the proposed network outperforms state-of-the-art classical approaches in Dice, Jaccard, HD, and ASD. The compared methods include the deep adversarial network (DAN) by Zhang et al. [34], the uncertainty-aware self-ensembling model (UA-MT) by Yu et

TABLE I
BRATS2020–QUANTITATIVE COMPARISON WITH 7 STATE-OF-THE-ART ALGORITHMS ON FOUR EVALUATION METRICS DIFFERENT AMOUNTS OF LABELED TRAINING DATA (30 (10%), 60 (20%), 90 (30%) LABELS) RESPECTIVELY

Method	labeled	unlabeled	Dice[%](↑)	Jaccard[%](↑)	HD[voxel](↓)	ASD[voxel](↓)
DAN [34]	30	270	83.90	73.63	36.86	1.95
UA-MT [18]	30	270	84.98	75.00	41.04	1.78
SASSNet [35]	30	270	83.19	73.79	40.93	3.84
DTC [36]	30	270	82.82	73.28	38.40	3.83
ICT [5]	30	270	84.18	74.00	39.78	1.99
URPC [16]	30	270	84.99	75.26	40.93	2.02
SCC [37]	30	270	82.03	71.80	22.10	1.70
Ours	**30**	**270**	**85.94**	**76.32**	**20.56**	**1.50**
DAN [34]	60	240	85.50	75.76	29.45	1.75
UA-MT [18]	60	240	86.17	76.67	33.04	1.75
SASSNet [35]	60	240	84.82	75.61	27.78	2.29
DTC [36]	60	240	84.90	75.53	30.09	2.60
ICT [5]	60	240	84.97	75.19	37.24	1.69
URPC [16]	60	240	85.87	76.40	40.10	2.00
SCC [37]	60	240	85.89	76.82	23.02	2.60
Ours	**60**	**240**	**86.67**	**77.45**	**20.44**	**1.45**
DAN [34]	90	210	86.67	77.42	25.01	1.66
UA-MT [18]	90	210	86.88	77.63	33.05	1.78
SASSNet [35]	90	210	86.39	77.21	29.01	1.46
DTC [36]	90	210	86.19	76.89	23.64	1.61
ICT [5]	90	210	87.01	77.82	26.35	1.68
URPC [16]	90	210	86.33	77.05	34.25	2.17
SCC [37]	90	210	85.99	76.70	22.53	1.46
Ours	**90**	**210**	**87.16**	**78.00**	**19.93**	**1.43**

[1] DAN ⇒ deep adversarial network; UA-MT ⇒ Uncertainty-aware self-ensembling model; SASSNet ⇒ shape-aware semi-supervised segmentation strategy; DTC ⇒ dual-task consistency; ICT ⇒ interpolation consistency training; URPC ⇒ uncertainty recified pyramid consistency; SCC ⇒ contrastive consistency semi-supervised model.

TABLE II
ABLATION RESULTS OF t-VALUE

t	Dice for BraTS2020	Dice for LA2018
0.15-1.0	84.41	89.04
0.25-1.0	84.93	87.97
0.35-1.0	85.04	89.02
0.45-1.0	85.47	87.84
0.55-1.0	85.85	87.98
0.65-1.0	85.30	87.92
0.75-1.0	85.94	89.28
0.85-1.0	83.40	88.95
0.95-1.0	85.21	88.54

al. [18], the shape-aware semi-supervised segmentation strategy (SASSNet) by Li et al. [35], the dual-task consistency (DTC) by Luo et al. [36], the URPC by Luo et al. [16], the ICT by Basak et al. [5], and the contrastive consistency semi-supervised model (SCC) by Liu et al. [37]. It is important to note that for the purpose of a consistent comparison, these approaches all used the same BraTS2020 backbone. There were three different label sizes used to compare these methods to the proposed approach: 30 labels (10%), 60 labels (20%), and 90 labels (30%). In terms of Dice, Jaccard, 95HD, and ASD, DUMC obtains 84.94%, 76.32%, 7.45 voxels, and 1.50 voxels for 30 labels (or 10%) respectively. However, when only 60 labels (20%) were used, DUMC obtains 87.16% of Dice, 78.00% of Jaccard, 6.80 voxels of 95HD, and 1.45 voxels of ASD; with 90 labels, DUMC achieves 87.16% of Dice, 78.00% of Jaccard, 6.80 voxels of 95HD, and 1.43 voxels of ASD.

LA2018: Table III and Fig. 10 validate that the proposed DUMC method performed the best result for left atrial segmentation compared with existing classical methods in Dice, Jaccard, 95HD, and ASD. Unlike the classical methods described above, the DAN method has been replaced with the mutual consistency network (MCNet) by [44].These methods were compared with the proposed method in 8 labels (10%), 16 labels (20%), and 24 labels (30%). For 8 labels, DUMC achieves 89.28%, 80.76%, 7.18 voxels, and 1.95 voxels in terms of Dice, Jaccard, 95HD, and ASD respectively. For 16 labels, DUMC gets 90.82%, 83.26%, 6.21 voxels, and 1.55 voxels in terms of Dice, Jaccard, 95HD, and ASD respectively. DUMC achieves 91.17% of Dice, 83.85% of Jaccard, 5.83 voxels of 95HD, and 1.62 voxels of ASD respectively when 24 labels are used. Among these semi-supervised methods, the proposed (DUMC) method achieves the best performance over them in two datasets. This is because DUMC encourages the similarity of data representations to find more difficult pathological tissues. It also gets the uncertainty from various models that can find more confident areas to effectively avoid error propagation. Likewise,

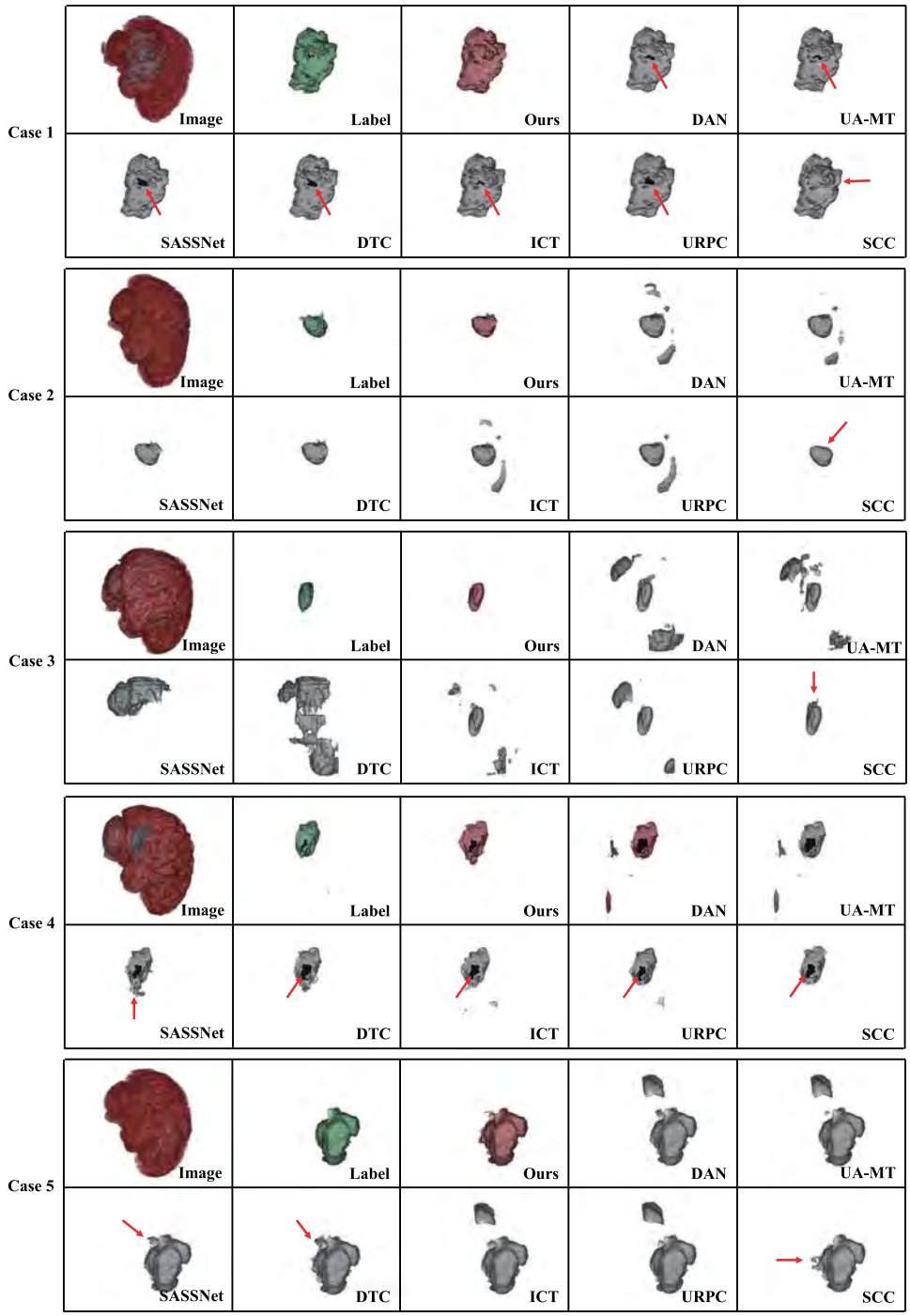


Fig. 8. Comparison with 7 state-of-the-art algorithms with 10% brain and left atrial training data (i.e., 30 cases for BraTS2020 (b) and 8 cases for LA2018 (a)), where red arrow denotes the incorrect segmentation.

DUMC fully uses volume-wise context information to achieve better segmentation.

The proposed (DUMC) method achieved the best performance among the considered semi-supervised methods in two datasets. This is due to the fact that DUMC engages the similarity of data representations in identifying more challenging pathological tissues. It additionally acquires uncertainty from various models that may locate more certain regions to efficiently prevent error propagation, while employing volume-wise context information in its entirety to improve segmentation.

B. Ablation Study

In this subsection, several experiments were conducted to validate the performance of each key component of the proposed DUMC network, including Contrastive Training

TABLE III
LA2018–QUANTITATIVE COMPARISON WITH 7 STATE-OF-THE-ART ALGORITHMS ON FOUR EVALUATION METRICS DIFFERENT AMOUNTS OF LABELED TRAINING DATA (8 (10%), 16 (20%), 24 (30%) LABELS) RESPECTIVELY

Method	labeled	unlabeled	Dice[%](↑)	Jaccard[%](↑)	95HD[voxel](↓)	ASD[voxel](↓)
UA-MT [18]	8	72	84.25	73.48	13.84	3.36
SASSNet [35]	8	72	87.32	77.72	9.62	2.55
DTC [36]	8	72	86.57	76.55	14.47	3.74
MCNet [44]	8	72	87.71	78.31	9.36	2.18
ICT [5]	8	72	86.06	75.76	17.37	4.74
URPC [16]	8	72	78.00	66.04	19.41	3.32
SCC [37]	8	72	86.51	76.54	10.51	2.56
Ours	**8**	**72**	**89.28**	**80.76**	**7.18**	**1.95**
UA-MT [18]	16	64	88.88	80.21	7.32	2.26
SASSNet [35]	16	64	89.54	81.24	8.24	2.20
DTC [36]	16	64	89.42	80.98	7.32	2.10
MCNet [44]	16	64	90.34	82.48	**6.00**	1.77
ICT [5]	16	64	88.34	79.36	11.31	3.20
URPC [16]	16	64	88.15	79.06	10.68	2.45
SCC [37]	16	64	89.81	81.64	7.15	1.82
Ours	**16**	**64**	**90.82**	**83.26**	6.21	**1.55**
UA-MT [18]	24	56	89.10	80.51	9.01	2.13
SASSNet [35]	24	56	89.72	81.51	9.53	2.64
DTC [36]	24	56	89.90	81.75	11.66	3.17
MCNet [44]	24	56	90.01	81.93	7.90	1.72
ICT [5]	24	56	89.02	80.36	8.59	2.35
URPC [16]	24	56	87.63	78.29	9.81	2.24
SCC [37]	24	56	90.52	82.76	6.49	**1.54**
Ours	**24**	**56**	**91.17**	**83.85**	**5.83**	1.62

[1] MCNet ⇒ mutual consistency network.

TABLE IV
ABLATION RESULTS OF α-VALUE AND β-VALUE

α	Dice for BraTS2020	Dice for LA2018	β	Dice for BraTS2020	Dice for LA2018
0	85.00	86.99	0	84.39	87.82
0.1	85.42	86.23	0.1	85.29	86.96
0.2	84.49	88.23	0.2	85.31	88.28
0.3	85.67	87.72	0.3	85.50	89.10
0.4	85.57	86.61	0.4	85.04	87.20
0.5	82.73	87.55	0.5	84.28	87.54
0.6	84.74	86.66	0.6	84.87	86.17
0.7	85.23	86.99	0.7	83.05	86.58
0.8	85.94	89.28	**0.8**	**85.94**	**89.28**
0.9	83.33	85.83	0.9	85.07	86.36
1	84.62	86.78	1	85.14	86.67

Module (CTM), Dual Uncertainty Strategy (DUS) and Mixing Volume Consistency Module (MVCM). This method was compared using a backbone (UNet/VNet with two different up-sampling modes: trilinear interpolation and transpose convolution) and w/o DUS + CTM and w/o DUS + MVCM and w/o CTM.

1) Advantage of Contrastive Training Module (CTM): The quality of the augmented images in both datasets is shown to be enhanced by the proposed CTM in Figs. 6 and 7. To maintain the invariance of data augmentation, which might bring similar images closer together and move diverse images farther apart, CTM computes the contrastive similarity (i.e., contrastive loss) between the original images and their augmentations. Backbone + CTM (i.e., without DUS + MVCM), DUMC without CTM (i.e., w/o CTM). DUMC are the only decoders used throughout training. It appears that the contrastive training module can enhance the quality of augmented images for BraTS2020 and LA2018. Additionally, a comparison was made between the 3D

TABLE V
STATISTIC ANALYSIS OF OUR PROPOSED METHOD

		DAN	UAMT	SASSNet	DTC	ICT	URPC	SCC
BraTS2020	Dice	0.008224	0.000316	0.00416	0.001205	0.000343	0.000401	0.002097
	Jaccard	0.008445	0.000271	0.005284	0.001115	0.000298	0.000305	0.001484
	HD	0.001161	1.09E-05	0.002117	0.001757	0.000405	0.000145	0.020568
	ASD	3.34E-07	1.22E-06	2.68E-07	1.05E-07	6.44E-07	5.21E-08	0.000454
LA2018	Dice	1.91E-06	1.34E-05	5.72E-06	2.67E-05	2.67E-05	1.91E-06	1.91E-05
	Jaccard	1.91E-06	1.34E-05	5.72E-06	2.67E-05	2.67E-05	1.91E-06	1.91E-05
	95HD	3.81E-06	6.29E-05	9.54E-06	0.000168	0.000168	3.81E-06	0.000105
	ASD	1.91E-06	4.77E-05	9.54E-06	0.019234	0.019234	1.91E-06	0.001986

average pooling layer and the 3D adaptive average pooling layer. Better segmentation was achieved with both datasets when using the proposed CTM with a 3D average pooling layer.

2) Advantage of Dual Uncertainty Strategy (DUS): Uncertainty from various models can establish a mutual constraint for improved guidance, as shown in Figs. 6 and 7. This allows the dual uncertainty technique to successfully guide the subsequent segmentation. It has been noticed during training of backbone + MVCM and backbone + MVCM + DUS that dual uncertainty can choose the more confident area to guide the mixing volume consistency as explained in Section V-B3. However, by modifying the uncertainty ratio-α parameter, we can compare the dual uncertainty from two separate models to the single uncertainty from a single model. When the ratio of uncertainties is 0.8 in Table IV, the best result is achieved, demonstrating that dual uncertainty is superior to single uncertainty.

3) Advantage of Mixing Volume Consistency Module (MVCM): The results of mixing volume consistency are shown to be superior to those of original consistency in Figs. 6 and 7. For the most part, MVCM retains more volumetric context information for improved segmentation. In order to gauge MVCM's efficacy, common consistency was substituted for MVCM in the backbone (i.e., backbone Plus MVCM and backbone). The mixing volume consistency is found to preserve volumetric information of the single image but also keeps volumetric features of different images. In addition, the results of using varying mixing ratios as comparisons demonstrate that the mixing ratio-β (the equilibrium between original consistency and mixing consistency) influences the prediction. Based on the ratios presented in Table IV, the evaluation performance of 0.8 preceding and compared with other ratios in all metrics, is credible to DUMC.

Moreover, based on Figs. 6 and 7, these ablation experiments are compared with various degrees, which indicates that our method is not only the most capable but also the most stable among these methods. Specifically, for BraTS 2020, CTM+DUS+MVCM is better than each individual module or incomplete module composition with only 30 (10%) labels. Furthermore, the combination of any two modules is better than each individual module. For example, DUS+MVCM is better than only DUS and only MVCM. For LA 2018, a combination of all modules is better than single module or incomplete module with 8 (10%) labels. To sum up, the above two figures are sufficient to explain the validity of our method.

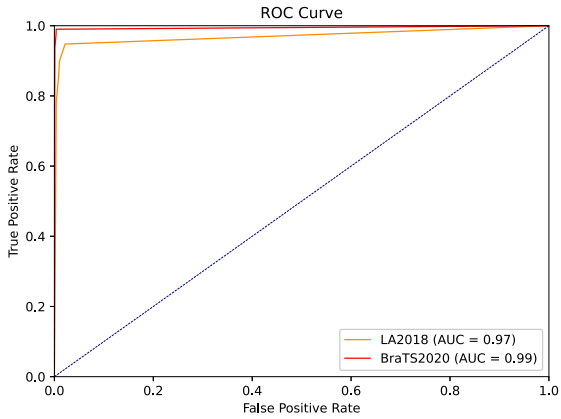


Fig. 9. The receiver operating characteristic (ROC) curve analysis of our proposed method. The orange line denotes the curve of LA2018. And the red line denotes the curve of BraTS2020.

C. Statistic Analysis

The Receiver Operating Characteristic (ROC) Analysis: As shown in Fig. 9, the receiver operating characteristic (ROC) curve, we can get the AUC (the corresponding area under the curve), which demonstrate the performances of the proposed method. The closer the curve is to the top left, the more positive examples take precedence over negative ones, and the better the model will perform overall. For BraTS2020, the AUC is equal to 0.99 with 8 (10%) labels. And the AUC is 0.97 for LA2018 with 8 (10%) labels. We evaluated the classifier performance by computing true positive rate (TPR) and false positive rate (FPR) over the two public datasets: $TPR = \frac{TP}{P} = \frac{TP}{(TP+FN)}$, $FPR = FP/N = FP/(FP + TN)$. where TP denotes true positives, TN true negatives, FP false positives and FN false negatives. The total number of Abnormal and Normal segments are P and N, respectively.

P-Value Test: The p-value is a parameter used to determine the outcome of a hypothesis test. The p-value is simply the probability of an outcome that is more extreme than the sample observations obtained when the null hypothesis is true. We compared our method with state-of-the-art methods by using P-value. As shown in Table V, We get all the p-values that are less than 0.05, which can demonstrate the effectiveness of our proposed DUMC method.

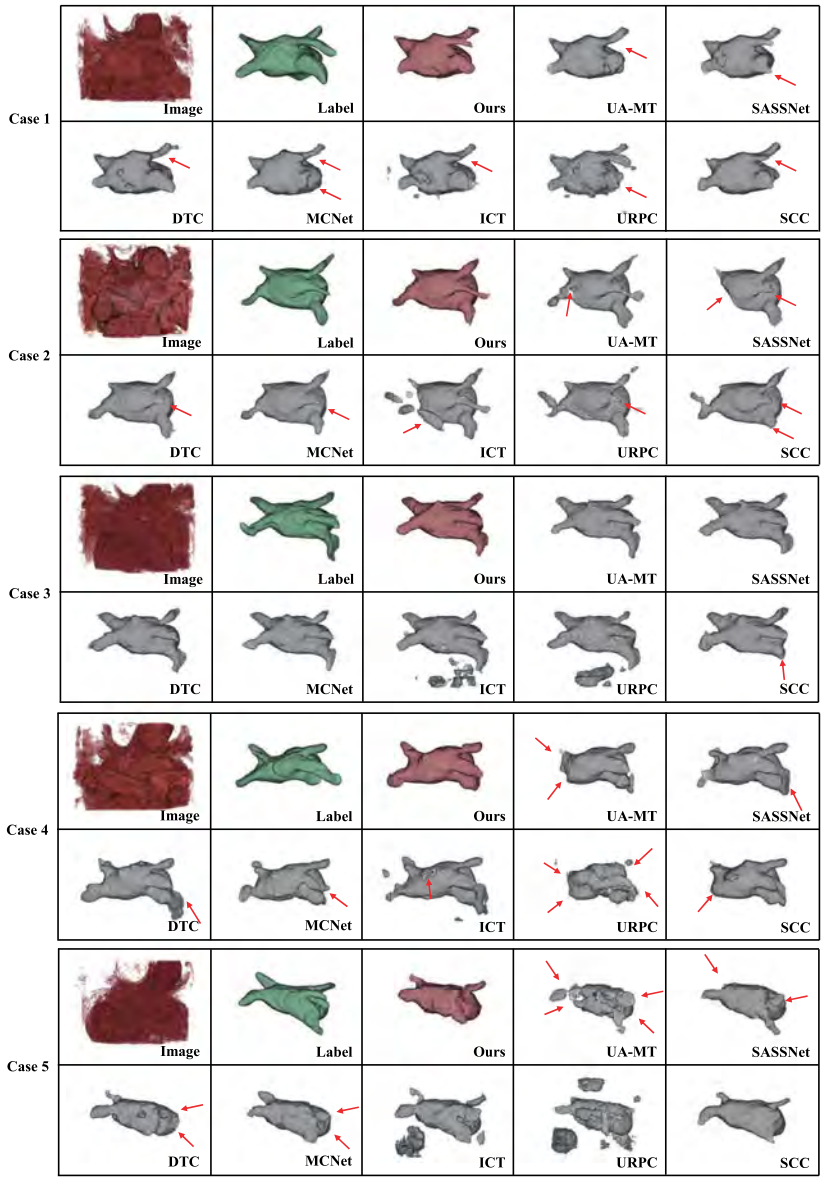


Fig. 10. Comparison with 7 state-of-the-art algorithms with 10% brain and left atrial training data (i.e., 30 cases for BraTS2020 (b) and 8 cases for LA2018 (a)), where red arrow denotes the incorrect segmentation.

D. Limitation

This study has some limitations which have to be pointed out. First, we use two different networks and many times forward to calculate dual uncertainty, which will bring large computation cost. Second, we conduct some experiments in two small public datasets. Thus, our method may not prove the effectiveness of a bigger dataset. Finally, our method is adaptable for only two datasets–BraTS and LA instead of all medical images. We will adopt more appropriate and advance method adapt to medical images, like deep-learning, to solving more practical clinic problems.

VI. Conclusion

For accurate segmentation, a dual uncertainty-guided mixing consistency network that completely exploits unlabeled data and volume-wise context information was proposed. Furthermore, the proposed Contrastive Training Module (CTM) can detect other pathological abnormalities by maintaining the consistency of augmented data. This research applied a Dual Uncertainty Strategy (DUS) to select more certain regions for further segmentation guidance. Finally, the suggested Mixing Volume Consistency Module (MVCM) can improve the varied representation of targets by mixing two unlabeled data and estimating their consistency before and after segmentation. The quantitative and qualitative results demonstrate that the proposed method outperforms state-of-the-art 3D semi-supervised methods, indicating that this method has the potential to become a medical tool for accurate image segmentation. Different contrastive learning approaches will be investigated so that the suggested framework can be modified in the future to improve performance.

REFERENCES

[1] Q. Zhu, B. Du, and P. Yan, "Boundary-weighted domain adaptive neural network for prostate mr image segmentation," *IEEE Trans. Med. Imag.*, vol. 39, no. 3, pp. 753–763, Mar. 2019.

[2] L. Yu et al., "Automatic 3D cardiovascular MR segmentation with densely-connected volumetric convnets," in *Proc. Int. Conf. Med. Image Comput. Comput.- Assist. Interv.*, Springer, 2017, pp. 287–295.

[3] X. Li, L. Yu, H. Chen, C.-W. Fu, L. Xing, and P.-A. Heng, "Transformation-consistent self-ensembling model for semisupervised medical image segmentation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 2, pp. 523–534, Feb. 2021.

[4] W. Luo and M. Yang, "Semi-supervised semantic segmentation via strong-weak dual-branch network," in *Proc. Eur. Conf. Comput. Vis.*, Springer, 2020, pp. 784–800.

[5] H. Basak, R. Bhattacharya, R. Hussain, and A. Chatterjee, "An embarrassingly simple consistency regularization method for semi-supervised medical image segmentation," 2022, *arXiv:2202.00677*.

[6] Y. Zou et al., "PseudoSeg: Designing pseudo labels for semantic segmentation," 2020, *arXiv:2010.09713*.

[7] Z. Zheng and Y. Yang, "Rectifying pseudo label learning via uncertainty estimation for domain adaptive semantic segmentation," *Int. J. Comput. Vis.*, vol. 129, no. 4, pp. 1106–1120, 2021.

[8] Y. Li, J. Chen, X. Xie, K. Ma, and Y. Zheng, "Self-loop uncertainty: A novel pseudo-label for semi-supervised medical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv.*, Springer, 2020, pp. 614–623.

[9] A. Tarvainen and H. Valpola, "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, 2017, pp. 1195–1204.

[10] X. Chen, Y. Yuan, G. Zeng, and J. Wang, "Semi-supervised semantic segmentation with cross pseudo supervision," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 2613–2622.

[11] N. Souly, C. Spampinato, and M. Shah, "Semi supervised semantic segmentation using generative adversarial network," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 5688–5696.

[12] W.-C. Hung, Y.-H. Tsai, Y.-T. Liou, Y.-Y. Lin, and M.-H. Yang, "Adversarial learning for semi-supervised semantic segmentation," 2018, *arXiv:1802.07934*.

[13] D. Nie, Y. Gao, L. Wang, and D. Shen, "ASDNet: Attention based semi-supervised deep networks for medical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.- Assist. Interv.*, Springer, 2018, pp. 370–378.

[14] J. Zhao et al., "United adversarial learning for liver tumor segmentation and detection of multi-modality non-contrast MRI," *Med. Image Anal.*, vol. 73, 2021, Art. no. 102154.

[15] C. Bowles et al., "GAN augmentation: Augmenting training data using generative adversarial networks," 2018, *arXiv:1810.10863*.

[16] X. Luo et al., "Semi-supervised medical image segmentation via uncertainty rectified pyramid consistency," *Med. Image Anal.*, vol. 80, 2022, Art. no. 102517.

[17] Y. Wang et al., "Double-uncertainty weighted method for semi-supervised learning," in *Proc. Int. Conf. Med. Image Comput. Comput.- Assist. Interv.*, Springer, 2020, pp. 542–551.

[18] L. Yu, S. Wang, X. Li, C.-W. Fu, and P.-A. Heng, "Uncertainty-aware self-ensembling model for semi-supervised 3D left atrium segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv.*, Springer, 2019, pp. 605–613.

[19] J. Chen, C. Fu, H. Xie, X. Zheng, R. Geng, and C.-W. Sham, "Uncertainty teacher with dense focal loss for semi-supervised medical image segmentation," *Comput. Biol. Med.*, vol. 149, 2022, Art. no. 106034.

[20] L. Hu et al., "Semi-supervised NPC segmentation with uncertainty and attention guided consistency," *Knowl. Based Syst.*, vol. 239, 2022, Art. no. 108021.

[21] C. Xu et al., "Direct delineation of myocardial infarction without contrast agents using a joint motion feature learning architecture," *Med. Image Anal.*, vol. 50, pp. 82–94, 2018.

[22] C. Xu, J. Howey, P. Ohorodnyk, M. Roth, H. Zhang, and S. Li, "Segmentation and quantification of infarction without contrast agents via spatiotemporal generative adversarial learning," *Med. Image Anal.*, vol. 59, 2020, Art. no. 101568.

[23] Y. Wang, Y. Wang, J. Cai, T. K. Lee, C. Miao, and Z. J. Wang, "SSD-KD: A self-supervised diverse knowledge distillation method for lightweight skin lesion classification using dermoscopic images," *Med. Image Anal.*, vol. 84, 2023, Art. no. 102693.

[24] X. Hu, D. Zeng, X. Xu, and Y. Shi, "Semi-supervised contrastive learning for label-efficient medical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv.*, Springer, 2021, pp. 481–490.

[25] J. Chen et al., "Adaptive hierarchical dual consistency for semi-supervised left atrium segmentation on cross-domain data," *IEEE Trans. Med. Imag.*, vol. 41, no. 2, pp. 420–433, Feb. 2021.

[26] J. Kim, J. Jang, and H. Park, "Structured consistency loss for semi-supervised semantic segmentation," 2020, *arXiv:2001.04647*.

[27] G. French, S. Laine, T. Alia, M. Mackiewicz, and G. Finlayson, "Semi-supervised semantic segmentation needs strong, varied perturbations," 2019, *arXiv:1906.01916*.

[28] J. Peng, P. Wang, C. Desrosiers, and M. Pedersoli, "Self-paced contrastive learning for semi-supervised medical image segmentation with meta-labels," in *Proc. Adv. Neural Inf. Process. Syst.*, 2021, pp. 16686–16699.

[29] X. Chen and K. He, "Exploring simple siamese representation learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 15750–15758.

[30] X. Wang, R. Zhang, C. Shen, T. Kong, and L. Li, "Dense contrastive learning for self-supervised visual pre-training," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 3024–3033.

[31] X. Chen, S. Xie, and K. He, "An empirical study of training self-supervised vision transformers," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 9640–9649.

[32] X. Zhao, C. Fang, D.-J. Fan, X. Lin, F. Gao, and G. Li, "Cross-level contrastive learning and consistency constraint for semi-supervised medical image segmentation," in *Proc. IEEE 19th Int. Symp. Biomed. Imag.*, 2022, pp. 1–5.

[33] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 1597–1607.

[34] Y. Zhang, L. Yang, J. Chen, M. Fredericksen, D. P. Hughes, and D. Z. Chen, "Deep adversarial networks for biomedical image segmentation utilizing unannotated images," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv.*, Springer, 2017, pp. 408–416.

[35] S. Li, C. Zhang, and X. He, "Shape-aware semi-supervised 3D semantic segmentation for medical images," in *Proc. Int. Conf. Med. Image Comput. Comput.- Assist. Interv.*, Springer, 2020, pp. 552–561.

[36] X. Luo, J. Chen, T. Song, and G. Wang, "Semi-supervised medical image segmentation through dual-task consistency," in *Proc. AAAI Conf. Artif. Intell.*, 2021, pp. 8801–8809.

[37] Y. Liu, W. Wang, G. Luo, K. Wang, and S. Li, "A contrastive consistency semi-supervised left atrium segmentation model," *Comput. Med. Imag. Graph.*, vol. 99, 2022, Art. no. 102092.

[38] S. Bakas et al., "Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the brats challenge," 2018, *arXiv:1811.02629*.

[39] S. Bakas et al., "Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features," *Sci. Data*, vol. 4, no. 1, pp. 1–13, 2017.

[40] B. H. Menze et al., "The multimodal brain tumor image segmentation benchmark (BRATS)," *IEEE Trans. Med. Imag.*, vol. 34, no. 10, pp. 1993–2024, Oct. 2015.

[41] S. Chen, G. Bortsova, A. García-Uceda Juárez, G. V. Tulder, and M. D. Bruijne, "Multi-task attention-based semi-supervised learning for medical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Interv.*, Springer, 2019, pp. 457–465.

[42] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning dense volumetric segmentation from sparse annotation," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv.*, Springer, 2016, pp. 424–432.

[43] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. IEEE 4th Int. Conf. 3D Vis.*, 2016, pp. 565–571.

[44] Y. Wu, M. Xu, Z. Ge, J. Cai, and L. Zhang, "Semi-supervised left atrium segmentation with mutual consistency training," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv.*, Springer, 2021, pp. 297–306.

Chenchu Xu received the MSc degree in computer science from the Hefei University of Technology, Hefei, China, in 2013, and the PhD degree in computer science from Anhui University, in 2017. He is currently an associate professor with the School of Computer Science and Technology, Anhui University, Hefei, China. His current research interests include machine learning and medical image processing.

Yuan Yang is with the School of Computer Science and Technology, Key Laboratory of Intelligent Computing and Signal Processing of Ministry of Education, Anhui University, Hefei, China, and the Institute of Artificial Intelligence.

Zhiqiang Xia received the MSc degree from Anhui Medical University, Hefei, China. He is currently working toward the PhD degree with the Hefei University of Technology, Hefei, China. He works with the First Affiliated Hospital of University of Science and Technology of China and Nanshan People's Hospital.

Boyan Wang received the MS and PhD degrees in computer science from the Hefei University of Technology, Hefei, China, in 2017 and 2022, respectively. His current research interests include computer vision, data mining, and machine learning.

Dong Zhang received the bachelor's degree in automation from Northwestern Polytechnical University, in 2018, and the master's degree in biomedical engineering from Western University, in 2020. He is currently working toward the PhD degree with the University of British Columbia in Canada.

Zhang Yanping is a full professor with the School of Computer Science and Technology, Anhui University, China. Her main research interests include Computational intelligence and quotient space theory, machine learning methods and applications, artificial neural network and intelligent information processing.

Shu Zhao is a full professor with the School of Computer Science and Technology, Anhui University, China. Her research interests include network representation learning, knowledge graph, and social network analysis.