Computerized Medical Imaging and Graphics 99 (2022) 102092

Contents lists available at ScienceDirect

Computerized Medical Imaging and Graphics

journal homepage: www.elsevier.com/locate/compmedimag

A contrastive consistency semi-supervised left atrium segmentation model

Yashu Liu [a], Wei Wang [a,*], Gongning Luo [a], Kuanquan Wang [a], Shuo Li [b]

[a] School of Computer Science and Technology, Harbin Institute of Technology (HIT), China
[b] Department of Medical Imaging, Western University, London, Canada

ARTICLE INFO

Keywords:
Left atrium segmentation
Semi-supervised learning
Contrastive learning

ABSTRACT

Accurate segmentation for the left atrium (LA) is a key process of clinical diagnosis and therapy for atrial fibrillation. In clinical, the semantic-level segmentation of LA consumes much time and labor. Although supervised deep learning methods can somewhat solve this problem, a high-efficient deep learning model requires abundant labeled data that is hard to acquire. Therefore, the research on automatic LA segmentation of leveraging unlabeled data is highly required. In this paper, we propose a semi-supervised LA segmentation framework including a segmentation model and a classification model. The segmentation model takes volumes from both labeled and unlabeled data as input and generates predictions of LAs. And then, a classification model maps these predictions to class-vectors for each input. Afterward, to leverage the class information, we construct a contrastive consistency loss function based on these class-vectors, so that the model can enlarge the discrepancy of the inter-class and compact the similarity of the intra-class for learning more distinguishable representation. Moreover, we set the class-vectors from the labeled data as references to the class-vectors from the unlabeled data to relieve the influence of the unreliable prediction for the unlabeled data. At last, we evaluate our semi-supervised LA segmentation framework on a public LA dataset using four universal metrics and compare it with recent state-of-the-art models. The proposed model achieves the best performance on all metrics with a Dice Score of 89.81 %, Jaccard of 81.64 %, 95 % Hausdorff distance of 7.15 mm, and Average Surface Distance of 1.82 mm. The outstanding performance of the proposed framework shows that it may have a significant contribution to assisting the therapy of patients with atrial fibrillation. Code is available at: https://github.com/PerceptionComputingLab/SCC.

1. Introduction

Atrial fibrillation (AF) is a common heart disease and the risk of it increases with age (Feinberg et al., 1995). Patients with AF may have heart palpitations, breathlessness, low energy, and an increased risk of stroke (Center, 2009). Catheter ablation is a current routine therapy for patients with AF (Kalla et al., 2017). However, the success ratio of catheter ablation is unsatisfactory, after which AF recurrence and the second ablation often happen (Chelu et al., 2018). According to the clinical experience, ablation strategies and AF recurrence are dominated by the degree of atrial fibrosis and the ablation-related scar (Akoum et al., 2011; Wu et al., 2021). And learning the topology of the left atrium (LA) is crucial for evaluating the degree of atrial fibrosis and ablation-related scar in patients with AF. Therefore, to improve the success ratio of the catheter ablation, accurate segmentation of the LA in medical images is a critical process that can assist the clinic in understanding the topology of LA, assessing the risk of AF, and making patient-specific treatment plan. Recently, late gadolinium-enhanced MRI (LGE MRI) provides a promising visualizing ability for myocardial scar tissues through brightening scar signal intensities to differentiate them from the healthy tissues, which results in the poor boundary of the LA (Yang et al., 2020). The LA segmentation involves the LA cavity, pulmonary veins, LA appendage, etc. These complex structures and the fuzzy boundary problem make the acquirement of the semantic-level label of the LA consuming much more time and labor. Therefore, accurate and automatic segmenting of the LA in LGE MRI is a challenging and necessary task.

For the past few years, deep learning models have taken impressive improvements on several medical image segmentation tasks (Shen et al., 2017). However, a high-efficient supervised deep learning model requires abundant labeled data. And the requirement of plenty of data with dense annotations somewhat slows down the process of deep learning application in medical image analysis. On the other hand, a large amount of unlabeled data may be available with the development

* Corresponding author.
E-mail address: wangwei2019@hit.edu.cn (W. Wang).

https://doi.org/10.1016/j.compmedimag.2022.102092
Received 15 January 2022; Received in revised form 30 May 2022; Accepted 9 June 2022
Available online 16 June 2022

of the wise information technology of med (Cheplygina et al., 2019). Hence, research on leveraging unlabeled data for medical image analysis is highly required.

In this work, we focus on semi-supervised learning (SSL) to learn representations from both labeled and unlabeled data for LA segmentation. SSL is an intermediate way between supervised learning and unsupervised learning (Chapelle et al., 2006), and its efficiency has been verified in many computer vision tasks (Van Engelen and Hoos, 2020). Typically, SSL attempts to train a model with a limited amount of labeled data and a large amount of unlabeled data. The unlabeled data supervises the model in a self-training manner with the consistent regularization which is based on the assumption that predictions of the model should be consistent under minor perturbations for the same input (Van Engelen and Hoos, 2020). Notably, the scope of this work is the standard SSL whose involved data has the same categories and modality (e.g. MRI).

Recently, several LA segmentation works have been done with SSL to relieve the requirement of expensive dense annotations for deep learning models. Primary works of these SSL models for LA segmentation are based on consistent regularization. To be specific, they can either make model predictions consistent with the original unlabeled data and its random perturbed data (e.g. noise, scaling) or make the model learn distribution consistency between labeled and unlabeled data by adversarial learning. Due to the consistency is calculated among predictions of unlabeled data (also called pseudo labels), the false prediction has the potential to make the training unstable. To mitigate the effect of unreliable predictions on the stability of training, UA-MT leveraged an uncertainty map of predictions for perturbed data to filter out the high uncertainty regions (Yu et al., 2019). This model adopted the mean-teacher (Tarvainen and Valpola, 2017) framework that required two networks and multiple forward propagations to formulate the uncertainty information. To reduce the time and memory cost, Wu et al. designed a network with two decoders and formulated the discrepancy of these two predictions as model uncertainty information to construct an unsupervised loss (Wu et al., 2021b). However, this model just considered the consistency in the output-level. To embed the geometric information into training, Li et al. (2020) took a distance map regression as an auxiliary task and adopted a discriminator to distinguish the source of the predicted distance map to learn the representation from unlabeled data while learning the shape information (). Following this work, Luo et al., 2021a; Luo et al., 2021b extended the concept of consistency to the task-level and proposed a dual-task model that jointly optimized the segmentation task and a distance map regression task to utilize geometric information and unlabeled data at the same time. Most of these models leveraged unlabeled data by forcing the model to be consistent in either image-/ouput-level or feature-level (Wang et al., 2020). But they ignored the class-level information and became class-agnostic approaches. However, the class-level information is crucial to improve the distinguishability of the segmentation model.

Contrastive learning has achieved major advances in self-supervised representation learning. The main idea of it is to pull the positive samples together and push the negative samples apart. And the sample construction strategy is commonly based on data augmentations at the image-level. Augmentations of the same input are positive samples, and the other data are negative samples (Khosla et al., 2020; Chaitanya et al., 2020). The performance of contrastive learning has shown great potential and achieved state-of-the-art results in downstream visual tasks (He et al., 2020; Chen et al., 2020). However, the representation learning of contrastive learning is usually on the image-level. It is too rough to fit the semantic segmentation task. To learn more specific representations, Chaitanya et al. (2020) proposed a local version contrastive learning to encourage the model to learn local representations (). Following this local contrastive learning idea, Xiang et al. embedded a contrastive loss at the feature level for SSL based on a teacher-student model (Xiang et al., 2021). Although these models constructed sampling based on local or feature levels, the class information is still ignored.

Inspired by the idea of contrastive learning (Chen et al., 2020; Chaitanya et al., 2020; Khosla et al., 2020; Chen et al., 2021), we embedded a contrastive consistency loss at the class-level in an unsupervised manner to enable the class-aware SSL. For learning the class-level representation, we constructed a classification model following a segmentation model that takes the segmentation predictions as input and maps them into a class-vector space. Then, we set class-vectors of the same class as intra-class samples and class-vectors of different classes as inter-class samples. At last, the contrastive consistency loss based on these samples is embedded in the supervised segmentation loss to jointly optimize the segmentation framework.

In summary, the main contributions of our model are three folds:

Firstly, we proposed a class-aware semi-supervised LA segmentation framework. Compared with the class-agnostic SSL models, the framework can leverage the class-level information to learn representations from both labeled and unlabeled data to improve the distinguishability of the segmentation model.

Secondly, we proposed a contrastive consistency loss on the class-vector space. Compared with the sample construction strategy at the image-level, our class-level sample construction strategy can enable the model to learn more distinguishable representations that will be beneficial to the pixel-level segmentation task. Moreover, we set the samples of labeled data as the reference to samples of unlabeled data to alleviate the effect of the unreliable predictions for unlabeled data.

Thirdly, we verified our framework on the popular left atrial segmentation dataset and performed plenty of ablation and comparative experiments. Both quantitative and qualitative results demonstrated the superiority of the proposed framework.

2. Materials and methods

In this section, we will introduce the detail of the proposed LA semi-supervised segmentation framework. We first briefly present the involved data in this work. Afterward, we describe details of our framework and loss functions. At last, details of the implementation and metrics are described.

2.1. Materials

The dataset used in this work was disclosed by 2018 Atrial Segmentation Challenge organizers (Xiong et al., 2021). There are 100 training data scanned by LGE MRI of patients with AF. The isotropic resolution of these data is $0.625mm \times 0.625mm \times 0.625mm$ with spatial dimensions of either $512 \times 512 \times 88$ or $640 \times 640 \times 88$. Most of the data are provided by the University of Utah, while the rest are provided by multiple centers. Hence, the quality and distribution of these data are not consistent, which increases the challenge for the automatic LA segmentation task. The binary label of these data is provided by three trained radiologists using the Corview image processing software (Merrk Inc, Salt Lake City, UT). A label contains the LA cavity, pulmonary veins, and LA appendage.

For a fair comparison, the data pre-process and division strategy in this work are according to related works (Luo et al., 2021a; Li et al., 2020; Yu et al., 2019). In detail, the data is firstly center-cropped around the heart region and normalized into zero mean and unit variation. Then, we split the data into 80 training samples and 20 validation samples. For semi-supervised learning, we divide 20 % of training volumes as the labeled data and the rest 80 % of training volumes as the unlabeled data. All experiments in this work are trained on the training samples and tested on the validation set.

2.2. Framework

The overall framework of the proposed method is illustrated in Fig. 1, which consists of two sub-models. The main sub-model is a

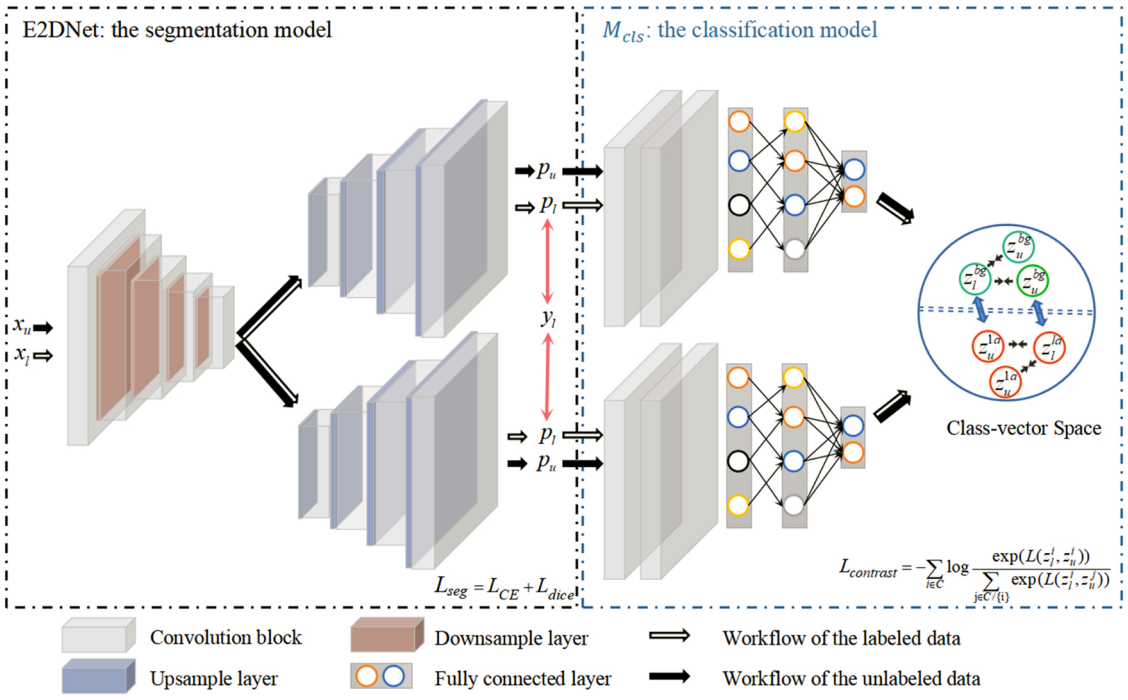

Fig. 1. The framework of the proposed left atrium semi-supervised segmentation model. There are two sub-models: the segmentation model E2DNet shown in the black dash box and the classification model M_{cls} shown in the blue dash box. The ↔ indicates the segmentation loss function for the labeled data. The class-vector space is shown in the blue circle, as indicate class-vectors of the labeled and unlabeled data respectively. Class-vectors with different colors indicate different classes.

segmentation model M_{seg} with a dual decoder structure that is shown in the black dash box in Fig. 1, termed as E2DNet. It takes 3D volumes as input and predicts pixel-level segmentation probabilities at the same time. Then a classification model M_{cls} following E2DNet maps the segmentation probabilities to class-vector space.

2.2.1. Segmentation model

Our segmentation model E2DNet is a variant VNet (Fausto et al., 2016) with an additional decoder. Specifically, E2DNet contains one encoder W_e and two decoders W_{dA} and W_{dB}. It takes both labeled and unlabeled volumes as input and generates two probabilities for each input. A brief mathematical representation of the segmentation model is shown in Eq. (1).

$$p_{dA}(x) = \sigma(f_{W_{dA}}(f_{W_e}(x))), \quad p_{dB}(x) = \sigma(f_{W_{dB}}(f_{W_e}(x))) \tag{1}$$

where σ is the softmax activation function; $x = \{x_l, x_u\}$ is a batch of training data, x_l indicates the labeled data, and x_u indicates the unlabeled data; $p_{d.}(x) = \{p_{d.}^{la}(x), p_{d.}^{bg}(x)\}$ indicates probability maps of each class for the input x. The final prediction in the inference stage is the average of these two predictions. Based on predictions of dual decoders, our model can benefit from the ensemble strategy to enhance the performance of the segmentation model in challenging regions.

In the training stage, the segmentation loss function only involves the labeled data. Following previous LA segmentation works (Yu et al., 2019), we leveraged a powerful combination of the dice loss function (Fausto et al., 2016) and the cross-entropy loss function that is shown in Eq. (2) as our supervised segmentation loss function.

$$L_{seg} = L_{CE}(p_{dA}(x_l), y_l) + L_{dice}(p_{dA}(x_l), y_l) + L_{CE}(p_{dB}(x_l), y_l) + L_{dice}(p_{dB}(x_l), y_l) \tag{2}$$

where $L_{CE}(p_l, y_l) = y_l \log p_l + (1 - y_l)\log(1 - p_l)$, $L_{dice}(p_l, y_l) = -2 \times |p_l \cap y_l| / (|p_l| + |y_l|)$, and y_l, p_l indicate the ground truth and prediction of the corresponding labeled data x_l respectively.

2.2.2. Classification model and contrastive consistency loss function

To leverage the class information on the labeled and unlabeled data for representation learning, we designed a classification model after the segmentation model to extract class-level information. As shown in the blue dash region of Fig. 1, the classification model takes the segmentation probabilities as input and maps the input into the class-vector space (shown in the blue circle region of Fig. 1). Specifically, the classification model predicts a class-vector z^i for each input$p_{d.}^i(x)$, where $\{i \in C | C = \{la, bg\}\}$ indicates class, la denotes LA, and bg denotes background. A brief mathematical representation of the classification model is shown in Eq. (3).

$$z^i = \sigma(f_{W_{cls}}(p_{d.}^i(x))) \tag{3}$$

Based on the class-vector of each class, we constructed the contrastive consistency loss function based on the class-level sample construction strategy. As shown in the blue circle of Fig. 1, we set class-vectors from the same class as intra-class samples, i.e. $\{z_l^{la}, z_u^{la}\}$. And inter-class samples are class-vectors from different classes, i.e. $\{z_l^{bg}, z_u^{la}\}$. To alleviate the effect of the unreliable predictions of the unlabeled data, we set class-vectors from the labeled data z_l as the reference to class-vectors from the unlabeled data z_u. At last, to minimize the margin of intra-class and maximize the margin of inter-class, the proposed contrastive consistency loss function is shown in Eq. (4).

$$L_{contrast} = -\sum_{i \in C} \log \frac{\exp(L(z_l^i, z_u^i))}{\sum_{j \in C/\{i\}} \exp(L(z_l^i, z_u^j))} \tag{4}$$

where L denotes the similarity function. By optimizing this $L_{contrast}$, the framework can make the similarity of intra-class samples more compact and the discrepancy of inter-class samples larger.

The final loss function of our proposed framework is the combination of the segmentation loss and the contrastive consistency loss as shown in Eq. (5).

$$L = L_{seg} + \lambda L_{contrast} \tag{5}$$

where λ is a weight to balance the importance of the segmentation loss and the contrastive consistency loss, we empirically set it to 1 in the experiment. With an end-to-end optimization manner, the segmentation model can be benefited from the contrastive consistency loss.

2.2.3. Implementation details and metrics

The model was implemented by PyTorch framework on a server with an NVIDIA RTX 2080Ti GPU. In this work, we adopted VNet (Fausto et al., 2016) as the backbone for all experiments. Then we implemented our E2DNet by adding an auxiliary decoder to the classical VNet. Most of the hyperparameters of our segmentation model followed recent state-of-the-art LA segmentation models (Yu et al., 2019; Li et al., 2020; Wu et al., 2021b). In detail, the framework was trained based on stochastic gradient descent (SGD) with a momentum of 0.9. The initial learning rate lr_{init} was 0.01 with a step decay strategy as shown in Eq. (6), where t denotes the current iteration and the step was set as 2,500. And it was trained 6,000 iterations with a batch size of 4. Following UA-MT (Yu et al., 2019), we adopted a two-stream sampling strategy to load the labeled and unlabeled data in the same batch size. Specifically, there were 2 labeled and 2 unlabeled data in each batch. Therefore, it did not require extra memory or time cost for saving previous parameters to back up former predictions.

$$lr = lr_{init} \times 0.1^{t/step} \tag{6}$$

Due to the memory limitation, the volumes were randomly cropped into patches with size $112 \times 112 \times 80$ during training. To alleviate the over-fitting problem, we adopted online augmentations including random crop, random flip, and random rotation 90° at the axial view. In the inference stage, only segmentation model E2DNet was required. Hence, there was no extra time or memory cost compared to other models in the inference stage. Predicted patches were extracted by a sliding window of size $112 \times 112 \times 80$ with a fixed $18 \times 18 \times 4$ stride. The final segmentation result was the average of predictions from two decoders of E2DNet of all patches.

Following recent works, we adopted four universal metrics (Yeghiazaryan and Voiculescu, 2015) to validate the performance of the proposed framework. Two overlap metrics, Dice score (Dice, 1945) and Jaccard (Liu et al., 2012) as shown in Eq. (7), verify the performance on the volumetric level. The interval of them is $[0\%, 100\%]$, and the best result approaches 100%. Two surface metrics, 95 % Hausdorff distance (95HD) (Gerig et al., 2001) and average surface distance (ASD) (Heimann et al., 2009) as shown in Eq. (8), evaluate the performance on the surface level. The interval of them is $[0mm, \infty)$, and the best result approaches 0 mm.

$$Dice = \frac{2|P \cap G|}{|P| + |G|} \times 100\%, \quad Jaccard = \frac{|P \cap G|}{|P \cup G|} \times 100\% \tag{7}$$

where P is the 3D prediction, and G is the corresponding 3D ground truth.

$$HD = \max\left\{ \min_{s_g \in S(G)} d(s_g, S(P)), \min_{s_p \in S(P)} d(s_p, S(G)) \right\}$$
$$ASD = \frac{\sum_{s_p \in S(P)} \min d(s_p, s_g)^{s_g \in s_G}}{|S(P)|} \tag{8}$$

where $S(\cdot)$ is the set of the surface of the corresponding volume, $d(\cdot)$ is the shortest Euclidean distance function to calculate the distance of a surface voxel s to another surface $S(\cdot)$.

3. Experiments and Results

3.1. Comparative Experiments and Results

Firstly, we compared our framework with four state-of-the-art LA semi-supervised segmentation works, including the uncertainty-aware mean teacher approach (UA-MT) (Yu et al., 2019), shape-aware adversarial network (SASSNet) (Li et al., 2020), local and global structure-aware entropy regularized mean teacher model (LG-ER-MT) (Hang et al., 2020), and dual-task consistency framework (DTC) (Luo et al., 2021a). Table 1 demonstrates the quantitative comparative result of these methods. The first two rows are the results of a fully supervised VNet. We set the performance of VNet that trained with 100 % labeled training data as the upper bound, and the performance of VNet that trained with 20 % labeled data as the lower bound. As shown in Table 1, all semi-supervised models benefit from the unlabeled data involved in the training stage. In detail, they achieve 2.85–3.78 % improvement in Dice score, 4.15–5.58 % improvement in Jaccard, 5.43–7.11 mm reduction in 95HD, and 0.38–1.69 mm reduction in ASD compared with the lower bound. Notably, our framework achieves the best performance on all metrics and is just lower than the upper bound Dice score of 1.33 %.

To further illustrate the performance of these comparative models, we randomly displayed three segmentation results in the inference stage in Fig. 2 for UA-MT, SASSNet, LG-ER-MT, DTC, and our framework. As shown in Fig. 2, other comparative models failed in the connection region between the LA cavity and the pulmonary vein, especially the LG-ER-MT. The proposed framework can fit the ground truth well and handle the topology of LA better (such as regions pointed out by red arrows). These comparative results verify that our framework has superior performance to other models.

Secondly, we also compared with other three latest semi-supervised models proposed for different datasets, including the semi-supervised contrastive learning approach (SSCL, Hu et al., 2021), dual-consistency semi-supervised learning network (UDC-Net, Li et al., 2021), and uncertainty rectified pyramid consistency model (URPC, Luo et al., 2021b). We re-implemented these models based on their public codes and trained them on the 20 % labeled training data set. Table 2. demonstrates the comparative experiment results. UDC-Net achieves comparative performance with us and gets the best surface metrics. Compared with our framework, it requires seven different auxiliary decoders during training to learn extensive features. And the additional dual uncertainty-guided mechanism improves its robustness. But our framework is based on a simpler structure and achieves the best performance on the overlap metrics under a minor weakness on the surface metrics than UDC-Net. This further indicates the superiority of our framework.

Table 1
Comparative experiment results with the latest and previous studies in the same database. 'Data Used' indicates the amount of labeled/unlabeled data involved in the training stage.

Methods	Data Used (train)		Metrics (inference)			
	Labeled	Unlabeled	Dice (%)	Jaccard (%)	95HD (mm)	ASD (mm)
VNet	80(100 %)	0	91.14	83.82	5.75	1.52
VNet	16(20 %)	0	86.03	76.06	14.26	3.51
UA-MT (Yu et al., 2019)	16(20 %)	64	88.88	80.21	7.32	2.26
SASSNet (Li et al., 2020)	16(20 %)	64	89.27	80.82	8.83	3.13
LG-ER-MT (Hang et al., 2020)	16(20 %)	64	89.62	81.31	7.16	2.06
DTC (Luo et al.,)	16(20 %)	64	89.42	80.98	7.32	2.10
Ours	16(20 %)	64	**89.81**	**81.64**	**7.15**	**1.82**

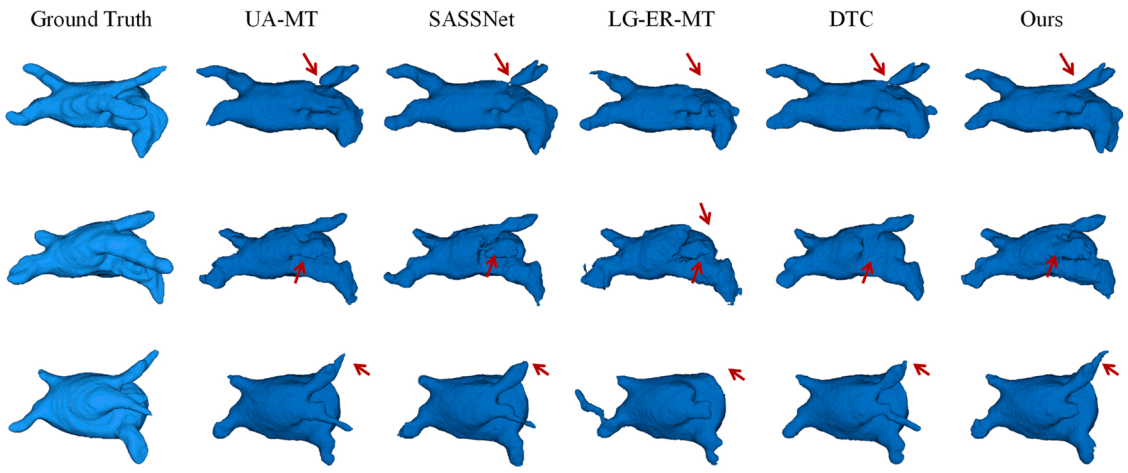

Fig. 2. Three segmentation results for comparative models. The first column denotes the ground truth. Red arrows point out the error-prone regions.

Table 2
Comparative experiment results with other latest semi-supervised learning methods on 20 % labeled training data set.

Methods	Metrics (inference)			
	Dice (%)	Jaccard (%)	95HD (mm)	ASD (mm)
SSCL (Hu et al., 2021)	84.93	74.45	6.95	2.18
UDC-Net (Li et al., 2021)	89.14	80.60	**6.43**	**1.76**
URPC (Luo et al., 2021b)	88.15	79.06	10.68	2.45
Ours	**89.81**	**81.64**	7.15	1.82

3.2. Ablation experiments and results

In this section, we demonstrated the effectiveness of involved modules in our proposed framework, including the dual decoder structure (E2DNet), and the additional classification model with the proposed contrastive consistency loss function. We also verified the impact of the ratio of labeled data involved in the training stage on the performance of our framework. Table 3 displays the quantitative results for all ablation models. These results are grouped by the ratio of labeled data involved in the training stage. In both groups of experiments, all proposed modules have a positive effect on the performance of the LA segmentation.

3.2.1. Effectiveness of dual decoder for fully supervised segmentation

The dual decoder structure was inspired by the multi-head ensemble strategy (Lee et al., 2015; Li et al., 2019). Ensemble modeling is a powerful strategy to improve performance by learning more variant representations. The multi-head ensemble has several variants. In this work, we focused on the shared encoder and multiple decoders structure that can perform inference at one time with fewer parameters. To investigate the effectiveness of our dual decoder segmentation structure compared with the basic VNet, we firstly conducted an ablation experiment on the VNet and E2DNet in a fully supervised manner. As shown in Table 3, E2DNet achieves improvement in all groups of ablation experiments compared with the classical VNet. For example, the E2DNet obtains 0.74 % improvement in Dice score, 1.24 % improvement in Jaccard, 2.35 mm reduction in 95HD, and 0.71 mm reduction in ASD on the 20 % labeled data set. Moreover, the performance improvement on the 10 % labeled data set is higher. Fig. 3 demonstrates the performance on all metrics in all groups of experiments on the ratio of labeled data involved in the training stage. E2DNet achieves better or comparable performance on all experiments. These improvements demonstrate the effectiveness of the E2DNet.

Table 3
Ablation experiments results of the proposed modules on the 10 % and 20 % labeled data set.

Methods	Data Used (train)		Metrics (inference)			
	Labeled	Unlabeled	Dice (%)	Jaccard (%)	95HD (mm)	ASD (mm)
VNet	8(10 %)	0	79.99	68.12	21.11	5.48
E2DNet	8(10 %)	0	81.63	70.75	14.25	3.64
Ours	8(10 %)	72	**86.51**	**76.54**	**10.51**	**2.56**
VNet	16(20 %)	0	86.03	76.06	14.26	3.51
E2DNet	16(20 %)	0	86.77	77.30	11.91	2.80
Ours	16(20 %)	64	**89.81**	**81.64**	**7.15**	**1.82**

3.2.2. Effectiveness of classification model with the contrastive consistency loss function for semi-supervised segmentation

In this section, we verified the impact of the additional classification model and the contrastive consistency loss function on the performance of E2DNet in a semi-supervised manner. As shown in Table 3, the proposed framework obtains improvement on all metrics in both groups of experiments with the proposed contrastive consistency loss function. In the 20 % labeled data group, the proposed framework achieves 89.51 % Dice score, 81.20 % Jaccard, 7.05 mm 95HD, and 1.82 mm ASD. Compared with the E2DNet, it improves 3.04 % in Dice score, 4.34 % in Jaccard, and reduced 4.76 mm in 95HD, and 0.98 mm in ASD on the 20 % labeled data set. This indicates the proposed framework can learn representations from unlabeled data and make a positive impact on the segmentation performance. We thought this was because the classification model learned more global representations to enhance the performance of our segmentation model. Moreover, the classification model only appears in the training stage, thereby there is no extra time cost in the inference stage.

3.2.3. Effectiveness of the ratio of used labeled data for semi-supervised segmentation

In former sections, we verified the efficiency of our framework on 10 % and 20 % labeled data sets, and the proposed framework achieves superior performance in both experiments. In this section, we conducted an ablation experiment to further evaluate the impact of the proportion of labeled data involved in the training stage on the performance of VNet, E2DNet, and the proposed framework. Fig. 3 presents the performance of these three models with the proportion of labeled data varying from 10 % to 100 % involved in the training stage. As shown in Fig. 3, the proposed framework achieves better performance than both fully supervised models in all data settings. This indicates the effectiveness of our framework to utilize the unlabeled data. With the

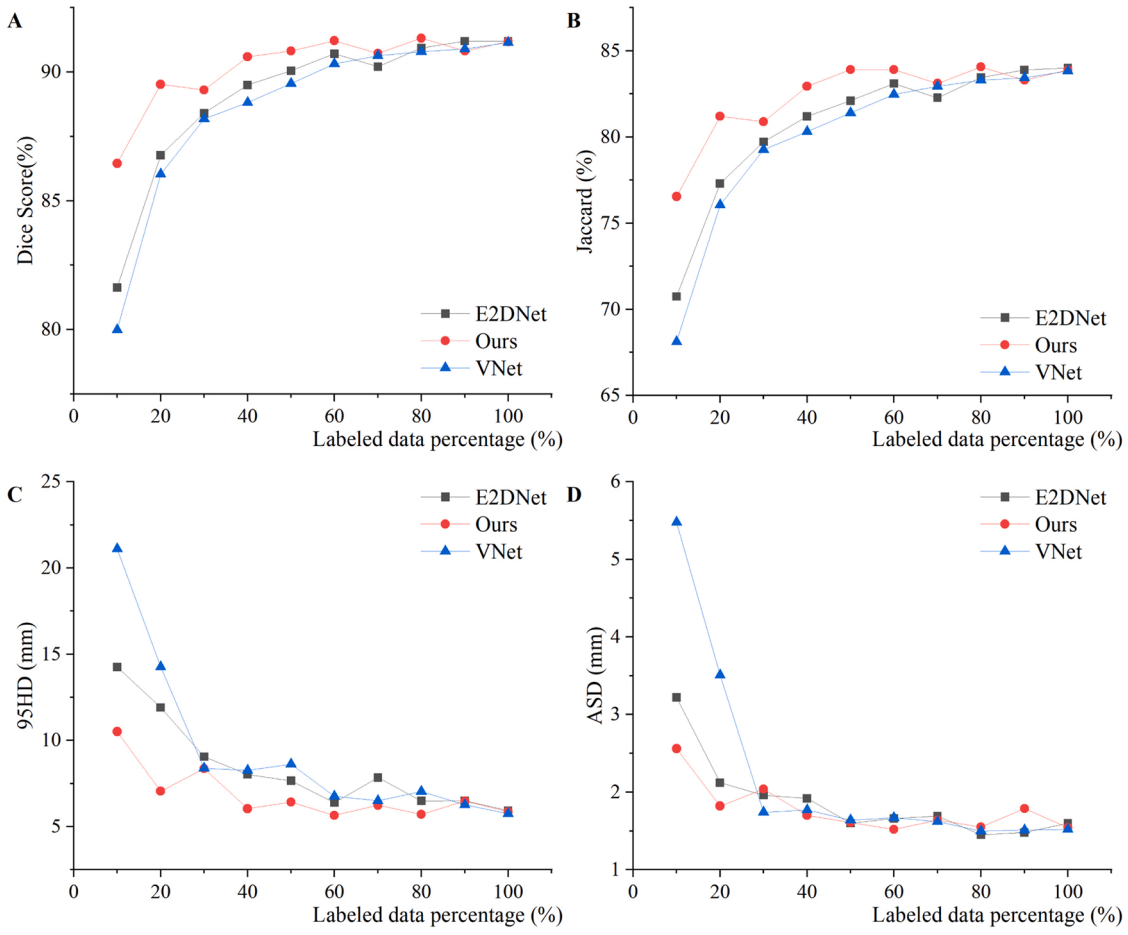

Fig. 3. Line diagrams of (A) Dice Score, (B) Jaccard, (C) 95HD, and (D) ASD in experiments with different percentages of labeled data for the VNet, E2DNet, and our proposed semi-supervised segmentation framework. E2DNet and VNet were trained in a fully supervised manner.

proportion of labeled data increasing, the margin between semi-supervised and fully supervised models gradually narrows. The superiority of our semi-supervised segmentation framework is more pronounced when the number of labeled data is small.

4. Discussion

In this work, we aimed to develop a class-aware semi-supervised LA segmentation framework on LGE MRI for patients with AF. Extensive experiments demonstrate that learning representation from the unlabeled data in the training stage can improve segmentation performance. Current mainstream semi-supervised LA segmentation works focused on a consistent regularization strategy to leverage the unlabeled data. While this kind of model usually requires a complex structure, such as a mean teacher with two networks. As shown in Table 1, these models do improve the segmentation performance on the semi-supervised LA segmentation task. However, they are class-agnostic models. And as shown in Fig. 2, their predictions in the connection region between the pulmonary vein and the LA cavity are error-prone. Inspired by the idea of contrastive learning, we proposed a contrastive consistency loss function at class-level to enable a class-aware semi-supervised LA segmentation model. With the additional class-level loss function, the model can learn more distinguishable representations by maximizing the discrepancy of inter-class and minimizing the margin of intra-class. Apart from this, the structure of the dual-decoder in our segmentation model also promotes the performance of our model. The dual-decoder structure is a simple structure of the multi-head ensemble strategy, which is a well-known technology to learn diverse features from an input to increase accuracy. It leverages smaller parameters and less inference time to achieve a comparable performance against the multiple sampling and multiple models strategies.

Even though the proposed method could relieve the label shortage problem, it still confronts the current data shortage problem in medical image analysis, because deep learning is a data-driven approach. To address the data shortage problem, the best way is to acquire data from the hospital. But this way still requires time to bring it to reality. A more efficient way to enlarge the training data may adopt the generative adversarial network to synthesize data based on currently available data. Combining the synthesized data and real data to perform the semi-supervised learning may address the data shortage and label shortage problems simultaneously. Recently, Chen et al. have tried to leverage the adversarial learning strategy to address the LA and scar segmentation simultaneously (Chen et al., 2021), and cross-domain LA segmentation (Chen et al., 2022). Their results demonstrated the technological feasibility to deal with the LA segmentation task under the generative adversarial network. In the future, we will pay attention to this technology to further address the data shortage problem of medical image analysis.

5. Conclusions

In this study, we constructed a semi-supervised LA segmentation framework with a segmentation model followed by a classification model. The E2DNet takes patches as input to predict probability maps for each class. And the classification model maps these probability maps into the class-vector space. At last, the framework is supervised by the

segmentation loss of labeled data and self-supervised by the contrastive consistency loss between labeled data and the unlabeled data. Thanks to the dual-decoder structure of our segmentation model, E2DNet achieves better performance than the baseline in a fully supervised manner. Then with the additional classification model and the proposed contrastive consistency loss function, the proposed semi-supervised segmentation framework achieves state-of-the-art performance with limited labeled data. Due to the classification model is only involved in the training stage, thereby, there is no extra time or memory cost in the inference stage compared with other models. The extensive quantitative and qualitative results have illustrated the promising potential of our framework in the future computer-aided therapy of AF.

CRediT authorship contribution statement

Yashu Liu: Conceptualization; Formal analysis; Investigation; Methodology; Validation; Visualization; Writing-original draft. **Wei Wang**, Funding acquisition; Supervision; Writing-review & editing. **Gongning Luo:** Funding acquisition; Project administration; Writing-review & editing. **Kuanquan Wang:** Funding acquisition; Project administration; Supervision; Writing-review & editing. **Shuo Li:** Investigation; Methodology; Validation; Writing-review & editing.

Funding

This work was supported by the National Natural Science Foundation of China [grant numbers 62001141, 62001144]; and the Science and Technology Innovation Committee of Shenzhen Municipality [grant number JCYJ20210324131800002].

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Thanks to the authors of Yu et al. (2019) and Ma et al. (2020). Their code repositories are the fundament of our work. Thanks to the organizers of the 2018 Atrial Segmentation Challenge to publish the LA segmentation dataset.

References

Akoum, N., Daccarett, M., Mcgann, C., Segerson, N., Vergara, G., Kuppahally, S., Badger, T., Burgon, N., Haslam, T., Kholmovski, E., et al., 2011. Atrial fibrosis helps select the appropriate patient and strategy in catheter ablation of atrial fibrillation: a DE-MRI guided approach. J. Cardiovasc. Electr. 22 (1), 16–22. https://doi.org/10.1111/j.1540-8167.2010.01876.x.

Center, E., 2009. Radiofrequency Ablation for Atrial Fibrillation: A Guide for Adults, Comparative Effectiveness Review Summary Guides for Consumers. Agency for Healthcare Research and Quality (US), Rockville.

Chaitanya, K., Erdil, E., Karani, N., Konukoglu, E., 2020. Contrastive Learning of Global and Local Features for Medical Image Segmentation with Limited Annotations, Advances in Neural Information Processing Systems.

Chapelle, O., Schölkopf, B., Zien, A., 2006. Semi-Supervised Learning The MIT Press, Cambridge, Massachusetts, London, England.

Chelu, M.G., King, J.B., Kholmovski, E.G., Ma, J., Gal, P., Marashly, Q., Aljuaid, M.A., Kaur, G., Silver, M.A., Johnson, K.A., et al., 2018. Atrial fibrosis by late gadolinium enhancement magnetic resonance imaging and catheter ablation of atrial fibrillation: 5-year follow-up data. J. Am. Heart Assoc. 7 (23), e6313 https://doi.org/10.1161/JAHA.117.006313.

Chen, J., Yang, G., Khan, H., Zhang, H., Zhang, Y., Zhao, S., Mohiaddin, R., Wong, T., Firmin, D., Keegan, J., 2021. JAS-GAN: generative adversarial network based joint atrium and scar segmentation on unbalanced atrial targets. IEEE J. Biomed. Health 1. https://doi.org/10.1109/JBHI.2021.3077469.

Chen, J., Zhang, H., Mohiaddin, R., Wong, T., Firmin, D., Keegan, J., Yang, G., 2022. Adaptive hierarchical dual consistency for semi-supervised left atrium segmentation on cross-domain data. IEEE T. Med. Imaging 41 (2), 420–433. https://doi.org/10.1109/TMI.2021.3113678.

Chen, T., Kornblith, S., Norouzi, M., Hinton, G. , 2020. A Simple Framework for Contrastive Learning of Visual Representations. In: Proceedings of the 37th International Conference on Machine Learning, pp. 1597–1607.

Chen, T., Luo, C., Li, L., 2021. Intriguing properties of contrastive losses. In: Advances in Neural Information Processing Systems - NeurIPS 2021. pp. 11834-11845.

Cheplygina, V., de Bruijne, M., Pluim, J.P.W., 2019. Not-so-supervised: a survey of semi-supervised, multi-instance, and transfer learning in medical image analysis. Med. Image Anal. 54, 280–296. https://doi.org/10.1016/j.media.2019.03.009.

Dice, L.R., 1945. Measures of the amount of ecologic association between species. Ecology 26 (3), 297–302. https://doi.org/10.2307/1932409.

Feinberg, W.M., Blackshear, J.L., Laupacis, A., Kronmal, R., Hart, R.G., 1995. Prevalence, age distribution, and gender of patients with atrial fibrillation: analysis and implications. Arch. Intern. Med. 155 (5), 469–473. https://doi.org/10.1001/archinte.1995.00430050045005.

Fausto, M., Nassir, N., Seyed-Ahmad, A., 2016. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. In: 2016 Fourth International Conference on 3D Vision (3DV), pp. 565–571.

Gerig, G., Jomier, M., Chakos, M., 2001. Valmet: A New Validation Tool for Assessing and Improving 3D Object Segmentation. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2001, pp. 516–523.

Hang, W., Feng, W., Liang, S., Yu, L., Wang, Q., Choi, K., Qin, J., 2020. Local and Global Structure-Aware Entropy Regularized Mean Teacher Model for 3D Left Atrium Segmentation. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2020,pp. 562–571.

He, K., Fan, H., Wu, Y., Xie, S., Grishick, R., 2020. Momentum Contrast for Unsupervised Visual Representation Learning. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 9726–9735.

Heimann, T., van Ginneken, B., Styner, M.A., Arzhaeva, Y., Aurich, V., Bauer, C., Beck, A., Becker, C., Beichel, R., Bekes, G., et al., 2009. Comparison and evaluation of methods for liver segmentation from CT datasets. IEEE Trans. Med Imaging 28 (8), 1251–1265. https://doi.org/10.1109/TMI.2009.2013851.

Hu, X., Zeng, D., Xu, X., Shi, Y., 2021. Semi-supervised Contrastive Learning for Label-Efficient Medical Image Segmentation. In: Medical Image Computing and Computer Assisted Intervention –. MICCAI 2021, 481–490.

Kalla, M., Sanders, P., Kalman, J.M., Lee, G., 2017. Radiofrequency catheter ablation for atrial fibrillation: approaches and outcomes. Heart, Lung Circ. 26 (9), 941–949. https://doi.org/10.1016/j.hlc.2017.05.125.

Khosla, P., Teterwak, P., Wang, C., Sarna, A., Tian, Y., Isola, P., Maschinot, A., Liu, C., Krishnan, D., 2020. Supervised Contrastive Learning. In: Advances in Neural Information Processing Systems,pp.18661–18673.

Lee, S., Purushwalkam, S., Cogswell, M., Crandall, D., Batra, D., 2015. Why M Heads are Better than One: Training a Diverse Ensemble of Deep Networks. arXiv. doi: 1511.06314.

Li, Y., Luo, L., Lin, H., Chen, H., Heng, P. , 2021. Dual-Consistency Semi-supervised Learning with Uncertainty Quantification for COVID-19 Lesion Segmentation from CT Images. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2021. pp. 199–209.

Li, H., Ng, J.Y., Natsev, P., 2019. EnsembleNet: End-to-End Optimization of Multi-headed Models. arXiv. doi:1905.09979.

Li, S., Zhang, C., He, X., 2020. Shape-Aware Semi-supervised 3D Semantic Segmentation for Medical Images. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2020. pp. 552–561.

Liu, Y., Cheng, H.D., Huang, J., Zhang, Y., Tang, X., 2012. An effective approach of lesion segmentation within the breast ultrasound image based on the cellular automata principle. J. Digit. Imaging 25 (5), 580–590. https://doi.org/10.1007/s10278-011-9450-6.

Luo, X., Chen, J., Song, T., Wang, G., 2021a. Semi-Supervised Medical Image Segmentation Through Dual-Task Consistency. In: AAAI Conf. Artificial Intelligence. 8801–8809.

Luo, X., Liao, W., Chen, J., Song, T., Chen, Y., Zhang, S., Chen, N., Wang, G., Zhang, S., 2021b. Efficient Semi-supervised Gross Target Volume of Nasopharyngeal Carcinoma Segmentation via Uncertainty Rectified Pyramid Consistency. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2021, pp. 318–329.

Ma, J., Wei, Z., Zhang, Y., Wang, Y., Lv, R., Zhu, C., et al., 2020. How Distance Transform Maps Boost Segmentation CNNs: An Empirical Study. In: Proceedings of the Third Conference on Medical Imaging with Deep Learning, pp. 479-492.

Shen, D., Wu, G., Suk, H., 2017. Deep learning in medical image analysis. Annu. Rev. Biomed. Eng. 19, 221–248. https://doi.org/10.1146/annurev-bioeng-071516-044442.

Tarvainen, A. , Valpola, H. , 2017. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In: Proceedings of the 31st International Conference on Neural Information Processing Systems, pp. 1195–1204.

Van Engelen, J.E., Hoos, H.H., 2020. A survey on semi-supervised learning. Mach. Learn. 109 (2), 373–440. https://doi.org/10.1007/s10994-019-05855-6.

Wang, Y. , Zhang, Y. , Tian, J. , Zhong, C. , Shi, Z. , Zhang, Y. , He, Z., 2020. Double-Uncertainty Weighted Method for Semi-supervised Learning. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2020. pp. 542–551.

Wu, Y., Tang, Z., Li, B., Firmin, D., Yang, G., 2021. Recent advances in fibrosis and scar segmentation from cardiac mri: a state-of-the-art review and future perspectives. Front. Physiol. 12. https://doi.org/10.3389/fphys.2021.709230.

Wu, Y. , Xu, M. , Ge, Z. , Cai, J. , Zhang, L. , 2021b. Semi-supervised Left Atrium Segmentation with Mutual Consistency Training. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2021, pp. 297–306.

Xiang, J., Li, Z., Wang, W., Xia, Q., Zhang, S., 2021. Self-ensembling contrastive learning for semi-supervised medical image segmentation. arXiv preprint arXiv:2105.12924.

Xiong, Z., Xia, Q., Hu, Z., Huang, N., Bian, C., Zheng, Y., Vesal, S., Ravikumar, N., Maier, A., Yang, X., et al., 2021. A global benchmark of algorithms for segmenting the left atrium from late gadolinium-enhanced cardiac magnetic resonance imaging. Med. Image Anal. 67, 101832 https://doi.org/10.1016/j.media.2020.101832.

Yang, Guang, Chen, Jun, Gao, Zhifan, Li, Shuo, Ni, Hao, Angelini, Elsa, et al., 2020. Simultaneous left atrium anatomy and scar segmentations via deep learning in multiview information with attention. Future Generation Computer Systems 107, 215–228. https://doi.org/10.1016/j.future.2020.02.005.

Yeghiazaryan, V., Voiculescu, I., 2015. An Overview of Current Evaluation Methods Used in Medical Image Segmentation. Department of Computer Science. University of Oxford, p. N12.

Yu, L., Wang, S., Li, X., Fu, C., Heng, P., 2019. Uncertainty-Aware Self-ensembling Model for Semi-supervised 3D Left Atrium Segmentation. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2019, pp. 605–613.