

# C3S3: Complementary Competition and Contrastive Selection for Semi-Supervised Medical Image Segmentation

Jiaying He<sup>†,1</sup>, Yitong Lin<sup>†,1</sup>, Jiahe Chen<sup>1</sup>, Honghui Xu<sup>\*,1</sup>, Jianwei Zheng<sup>\*,1</sup>

<sup>1</sup>Zhejiang University of Technology, Hangzhou, China

Email:{xhh, zjw}@zjut.edu.cn

**Abstract**—For the immanent challenge of insufficiently annotated samples in the medical field, semi-supervised medical image segmentation (SSMIS) offers a promising solution. Despite achieving impressive results in delineating primary target areas, most current methodologies struggle to precisely capture the subtle details of boundaries. This deficiency often leads to significant diagnostic inaccuracies. To tackle this issue, we introduce C3S3, a novel semi-supervised segmentation model that synergistically integrates complementary competition and contrastive selection. This design significantly sharpens boundary delineation and enhances overall precision. Specifically, we develop an *Outcome-Driven Contrastive Learning* module dedicated to refining boundary localization. Additionally, we incorporate a *Dynamic Complementary Competition* module that leverages two high-performing sub-networks to generate pseudo-labels, thereby further improving segmentation quality. The proposed C3S3 undergoes rigorous validation on two publicly accessible datasets, encompassing the practices of both MRI and CT scans. The results demonstrate that our method achieves superior performance compared to previous cutting-edge competitors. Especially, on the 95HD and ASD metrics, our approach achieves a notable improvement of at least 6%, highlighting the significant advancements. The code is available at <https://github.com/Y-TARL/C3S3>.

**Index Terms**—Semi-Supervised Learning, Medical Image Segmentation, Consistency Learning, Contrastive Learning

## I. INTRODUCTION

Medical image segmentation (MIS) is essential in healthcare, aiding clinicians in monitoring disease progression and planning treatments. The advancement of deep learning, especially supervised methods, has greatly enhanced MIS, making it the leading approach in the field and significantly improving performance in various practices [1]. Traditional supervised deep learning methods require a large amount of labeled data for training, and pixel-level annotation of medical images is both time-consuming and challenging. To address this issue, semi-supervised medical image segmentation techniques have been developed, aiming to reduce the need for extensive manual annotation while maintaining high performance.

SSMIS primarily stands out by leveraging unlabeled data to improve segmentation performance [2], [3]. Within this framework, two key strategies, *i.e.*, pseudo-label supervision

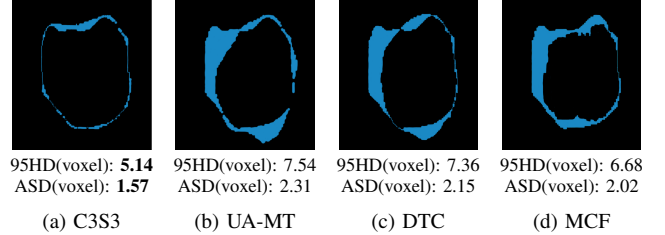


Fig. 1: The visual comparison among C3S3 and others, where the blue areas represent inaccurately localized segments.

and consistency regularization, have been extensively investigated [4]. The former uses pseudo-labels to supervise unlabeled images, while the latter enforces prediction consistency between different models or branches, often based on popular frameworks like Mean Teacher (MT) [5], [6] or co-training [4], [7], [8]. These approaches typically train two networks to ensure consistent predictions, even when the same samples are exposed to random variations. Despite the success of these methods in accurately locating target regions, refining fine boundaries remains a significant challenge.

By integrating two pop techniques, *i.e.*, semi-supervised learning and contrastive learning, we propose a new model to tackle the challenge. To begin with, our approach incorporates an outcome-driven contrastive learning module (ODCL), meticulously designed to enhance the delineation of boundary information. Note that contrastive learning is dedicated to train a model enjoying high discriminability of similar (positive) and dissimilar (negative) data pairs, which allows for a more nuanced understanding of the image content. On that basis, enhanced accuracy and refined details can be earned from the final model, particularly when fused with semi-supervised learning. Second, we elaborate a dynamic complementary competition module (DCC) to innovate beyond the traditions. Traditionally, within Mean Teacher models, the performance of the teacher network depends heavily on the student network, since the former is essentially a weighted combination of the latter's historical states. In contrast, our DCC minimizes the reliance between the student and teacher networks while utilizing pseudo-labeling to promote complementary competition, fostering a collaborative dynamic that enables mutually beneficial improvements. Figure 1 shows an example comparison between our C3S3 and others. As can be seen, our proposal

<sup>†</sup>Equal contributions. <sup>\*</sup>Corresponding author.

This work was supported in part by the National Natural Science Foundation of China under Grant 62276232 and 62406285, the China Postdoctoral Science Foundation under Grant 2024M752863, and the Key Program of Natural Science Foundation of Zhejiang Province under Grant LZ24F030012.

captures the much clearer boundary, with significant advantage over that from the competitors. The numerical values also verify the visual results. Technically, the primary contributions of our work are threefold.

- To strive for an enhanced boundary delineation, a new module named ODCL is proposed, which encompasses two key components, *i.e.*, Spatial Position Binary Masking and Dual-Space Intersection-Union Loss.
- We propose a DCC module that employs a weighted competing mechanism for precise pseudo-labeling, which also enables two subnets of the whole model to mutually learn from each other.
- With VNet and ResVNet behaved as the dual subnets and the proposed two modules seamlessly assembled, a final segmentation network C3S3 is molded. The results on two public benchmark datasets, *i.e.*, left atrium segmentation in MRI and pancreas segmentation in CT scans, verify that C3S3 sets a new state-of-the-art score.

## II. RELATED WORK

### A. Semi-supervised Medical Image Segmentation

Recent advancements in semi-supervised medical image segmentation have introduced a variety of innovative methods. For example, CauSSL [8] employs a novel causal graph to provide a theoretical foundation for mainstream semi-supervised segmentation approaches. Notably, two methodologies closely aligned with our research are self-training with pseudo-labeling [9], [10] and consistency regularization [5], [11], both of which have demonstrated significant potential in enhancing segmentation performance.

Pseudo-labeling involves using a trained model to generate labels for unlabeled data, which are subsequently used to refine the model through iterative updates, thereby improving its learning capability. However, with the abundance of unlabeled data, the risk of error propagation caused by inaccurate pseudo-labels poses a significant challenge. Therefore, enhancing the reliability and accuracy of pseudo-labels becomes a critical priority. C3T [10] introduces a Credibility-Aware Pseudo-Label Filtering mechanism, specifically designed for pixel-wise segmentation tasks, which validates and filters pseudo-labels to enhance inter-network guidance reliability and mitigate error propagation. Additionally, CAML [12] introduces modules that transfer prior knowledge and enforce distribution consistency, enhancing pseudo-labeling reliability.

Consistency regularization is a key technique in semi-supervised learning, enforcing consistent predictions for perturbed samples to reduce error propagation. It has been extended to multi-task learning by incorporating auxiliary tasks to utilize geometric information. In medical image segmentation (MIS), the Mean Teacher (MT) framework is widely adopted, where a student network is updated via gradients and a teacher network via EMA. TraCoCo [13] enhances consistency learning by varying spatial contexts, improving focus on foreground objects and segmentation performance. CORN [14] leverages second-order statistics and a Dynamic

Feature Pool to boost precision and maintain consistency. Besides, ICR-Net [15] integrates labeled data to enhance regularization and improve information extraction from unlabeled data, complementing traditional consistency methods.

### B. Contrastive Learning

Contrastive learning has shown remarkable success in the computer vision field [16], [17], focusing on learning discriminative features by distinguishing similar (positive) and dissimilar (negative) feature pairs. Initially successful in image-level tasks, it has since been adapted for dense prediction tasks like image segmentation, with growing interest in semi-supervised applications. UGCL [17] improves segmentation accuracy in medical image analysis by leveraging uncertainty-aware contrastive learning to address challenges such as boundary fuzziness and high-entropy regions.

More recently, the integration of contrastive learning with consistency regularization in semi-supervised segmentation has marked a significant progress, which leverages pseudo-labels for contrastive learning while simultaneously optimizing contrastive and consistency losses [16]. This dual approach promotes the learning of feature representations and the refinement of pseudo-labels, thereby improving both the efficiency and accuracy of segmentation.

## III. METHODOLOGY

In accordance with MCF [18], the renowned VNet (A) and its variant ResVNet (B) are employed as our dual backbones, both of which own a u-shaped extractor and a tail-ready classifier. Besides, as given in Figure 2, the core ingredient of our proposal is centered on two principal modules, *i.e.*, DCC and ODCL, whose detailed explanations are provided in Subsections III-B and III-C, respectively.

### A. The overview framework

During semi-supervised learning, it is assumed that training dataset contains  $N$  labeled data and  $M$  unlabeled data, where  $M \gg N$ . For convenience, we denote a small amount of labeled data as  $\mathcal{D}_L = \{(x_i^L, y_i^L)\}_{i=1}^N$  and the unlabeled counterpart as  $\mathcal{D}_U = \{x_i^U\}_{i=1}^M$ , where  $x_i \in \mathbb{R}^{H \times W \times D}$  denotes the training image and  $y_i \in \mathbb{B}^{H \times W \times D}$  is the label (if available). Yet with a limited number of labeled samples  $x_i^L$ , the objective of semi-supervised learning is to achieve promising results with the aid of the extra unlabeled data  $x_i^U$ .

With two random data augmentations initially imposed [18], the whole input  $\mathbf{X}$  is then individually fed into two backbones, generating the learned features for labeled data  $\hat{Y}^L$  and unlabeled data  $\hat{Y}^U$ .

$$\hat{Y}_A = A(\mathbf{X}), \hat{Y}_B = B(\mathbf{X}) \quad (1)$$

Referring to previous work [16], for labeled data, cross entropy loss  $L_{ce}$  and dice loss  $L_{dice}$  are utilized for supervised training:

$$L_{seg} = L_{ce}(\hat{Y}^L, y) + L_{dice}(\hat{Y}^L, y) \quad (2)$$

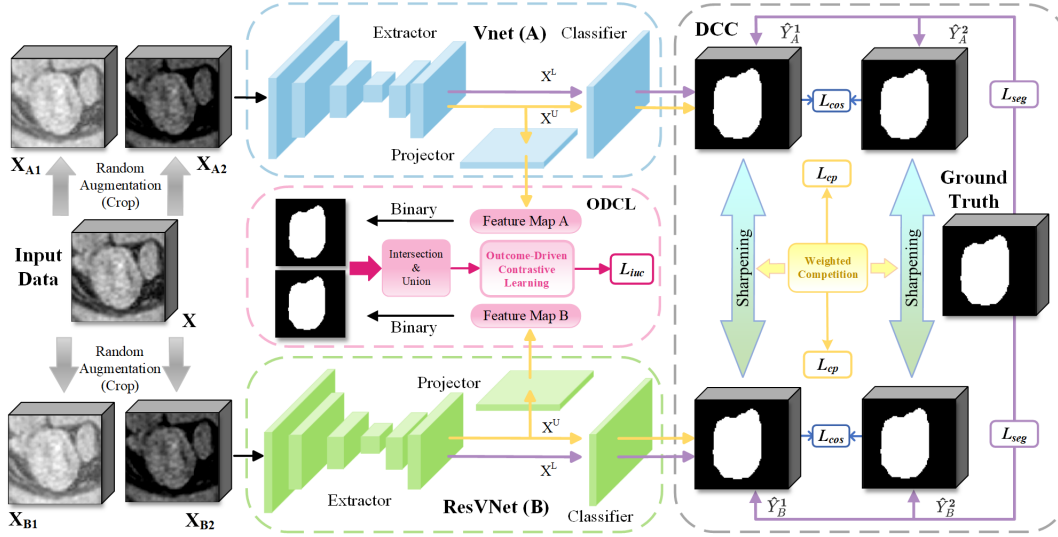


Fig. 2: Overview of our proposal. Upon receiving the input data, C3S3 undergoes two random data augmentations, followed by two separate backbones, i.e., VNet and ResVNet. The yellow line represents unlabeled samples, while the purple line indicates labeled samples. The unlabeled data, after processing through the projector, is directly fed into ODCL, during which a loss  $L_{iuc}$  is incurred. Both VNet and ResVNet yield two prediction results, between which a consistency loss  $L_{cos}$  is computed, followed by the generation of a pseudo-label loss  $L_{cp}$  via the weighted competition module. In addition, labeled data contribute to a supervised loss  $L_{seg}$ .

For unlabeled data, we calculate the pseudo supervision loss  $L_{cp}$  and outcome-driven contrastive loss  $L_{iuc}$ :

$$L_{unsup} = L_{cp}(\hat{Y}_A^U, \hat{Y}_B^U, \hat{Y}^L) + \lambda L_{iuc}(p_A, p_B, p^-) \quad (3)$$

where  $\lambda$  is a weight balancing the two sub-loss, whose detailed setting is dependent on the specific task. The practices of  $p_A$ ,  $p_B$ , and  $p^-$  will be elaborated in Subsection III-C.

Meanwhile, we use consistency regularization to further reduce the difference between two predictions, leading to a consistency loss  $L_{cos}$ . With all the sub-loss assembled, the overall loss is given as  $L_{total} = L_{seg} + L_{cos} + L_{unsup}$ .

### B. DCC Module

We propose a dynamic complementary competition mechanism, enabling two backbones to simultaneously learn from and compete with each other. The dynamic nature of DCC allows for real-time adaptation to changing data patterns, which enjoys a significant improvement over the static version. Simultaneously, as two backbones share equal status, the decoupling of tight dependency should be accomplished, further allowing for more exclusive focus on the respective learning.

In practice, we commence with two random augmentations of the inputs, followed by the respective feeding into backbone A and backbone B. Consequently, the concerned outputs  $\{\hat{Y}_A^1, \hat{Y}_A^2\}$  and  $\{\hat{Y}_B^1, \hat{Y}_B^2\}$  can be generated. After obtaining the supervised loss  $L_{seg}$  using ground truth, a weighted competition function is employed to determine which backbone performs more competitively in the current state.

$$L_{competition} = \alpha L_{ce}(\hat{Y}^L, y) + (1 - \alpha) L_{dice}(\hat{Y}^L, y) \quad (4)$$

where parameter  $\alpha$  is the determining weight.

Then, we utilize the currently better-performing backbone to generate pseudo-labels, supervising the predictions of the other backbone and computing the pseudo supervised loss:

$$L_{cp} = MSE(\hat{Y}^{U1}, Y_p^{U1}) + MSE(\hat{Y}^{U2}, Y_p^{U2}) \quad (5)$$

Meanwhile, to minimize the disagreement between  $\hat{Y}_A^1$  and  $\hat{Y}_B^2$ , (Resp.  $\hat{Y}_B^1$  and  $\hat{Y}_A^2$ ), the cosine distance between two predicted results is utilized to obtain a consistency loss:

$$L_{cos}(\hat{Y}_A^1, \hat{Y}_B^2) = 1 - \frac{\hat{Y}_A^1 \cdot \hat{Y}_B^2}{\|\hat{Y}_A^1\|_2 \cdot \|\hat{Y}_B^2\|_2} \quad (6)$$

The integration of an explicit loss term into the model offers additional regularization during training, which is vital to improving the stability and consistency of the entire training process. Additionally, during the selection of the better-performed backbone, since the two loss terms in the competition are already calculated during the model training, hence introducing additional computations is not required, which further guarantees a reduced computational overhead.

### C. ODCL Module

To ensure consistency between the feature spaces of two backbones, contrastive learning is integrated, which operates in two phases: spatial position binary masking and dual-space intersection-union loss. Contrastive learning enables the model to gain a deeper understanding of complex data structures, offering more detailed data analysis.

**Spatial Position Binary Masking:** Figure 3 provides an overview of the mask sampling scheme. Unlabeled data is firstly input into two backbones, followed by the projection of output from the extractor into low-dimensional feature maps. Subsequently, we binarize the two feature maps, with the

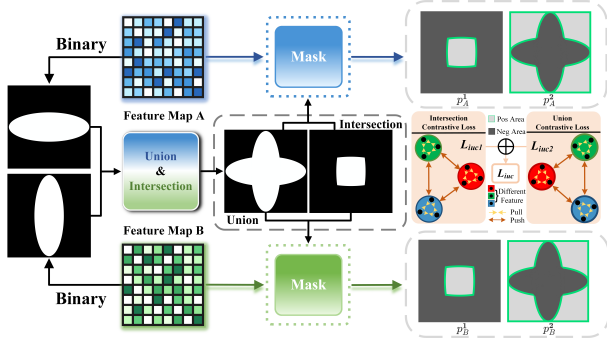


Fig. 3: Illustration of ODCL. ODCL identifies positive and negative sample regions from feature maps processed by intersection and union operations, then calculates the loss  $L_{iuc}$ .

intersection and union of the results calculated to derive the masks. The masks from the intersection and the union are respectively labeled as  $\mathcal{M}_{\cap}$  and  $\mathcal{M}_{\cup}$ . With the feature maps passed through the two masks, the results obtained are denoted as  $p_A^1$  and  $p_A^2$  (Resp.  $p_B^1$  and  $p_B^2$ ), with negative samples defined as  $p^-$ . Note that these notation has appeared in Eq. 3.

We consider the intersection part as the correct foreground prediction by a high probability, and the voxels excluding the union as the correct background prediction. When we set positive samples from the foreground region, the negative sample would be selected from the possible background region, and vice versa, which ensures an evident contrasting effect.

In practice, we design a strategy for selecting positive and negative sample pairs. That is, given an anchor, the corresponding voxel at the same spatial position is chosen as the positive sample. With the utilization of the intersection mask, we further designate the intersecting area as positive sample area and the remaining as negative samples, from which the contrastive loss  $L_{iuc1}$  can be calculated. Similarly, with the union mask, we consider the voxels excluding the union result as positive sample area and the rest as negative samples to compute the contrastive loss  $L_{iuc2}$ .

**Dual-Space Intersection-Union Loss:** Following the respective positive and negative sampling, a contrastive loss is formulated to draw positive pairs closer and negative pairs far away. Specifically, for a foreground voxel feature  $\varphi_1 \in p_A$ , the contrastive loss is computed as follows:

$$L_{iuc1}(\varphi_1, \varphi_2) = -\log \left( \frac{\mathcal{M}_{\cap} e^{\cos(\varphi_1, \varphi_2)/\tau}}{\mathcal{M}_{\cap} e^{\cos(\varphi_1, \varphi_2)/\tau} + (1 - \mathcal{M}_{\cap}) \sum_{n=1}^N e^{\cos(\varphi_1, \varphi_n)/\tau}} \right) \quad (7)$$

where  $\varphi_2$  represents the positive voxel extracted from  $p_B$ , while  $\varphi_n$  refers to the negative voxels.  $N$  indicates the total count of sampled negative voxels, and  $\tau$  stands for the temperature hyperparameter.

Similarly, for the background voxels, the contrastive loss can be formulated as follows:

$$L_{iuc2}(\varphi_1, \varphi_2) = -\log \left( \frac{(1 - \mathcal{M}_{\cup}) e^{\cos(\varphi_1, \varphi_2)/\tau}}{(1 - \mathcal{M}_{\cup}) e^{\cos(\varphi_1, \varphi_2)/\tau} + \mathcal{M}_{\cup} \sum_{n=1}^N e^{\cos(\varphi_1, \varphi_n)/\tau}} \right) \quad (8)$$

Therefore, the whole contrastive loss between two feature maps can be given as:

$$L_{iuc}(p_A, p_B, p^-) = \frac{1}{N_p} \sum_{\varphi_1 \in p_A} L_{iuc1}(\varphi_1, \varphi_2) + \frac{1}{N_p} \sum_{\varphi_1 \in p_A} L_{iuc2}(\varphi_1, \varphi_2) \quad (9)$$

where  $N_p$  denotes the number of voxels in  $p_A$ .

With four feature maps from two backbones, we maximize their information utility by computing contrastive loss between feature map pairs from different branches, culminating in the definition of the dual-space intersection-union loss:

$$L_{all}(p_A, p_B, p^-) = L_{iuc}(p_A^1, p_B^1, p^-) + L_{iuc}(p_A^1, p_B^2, p^-) + L_{iuc}(p_A^2, p_B^1, p^-) + L_{iuc}(p_A^2, p_B^2, p^-) \quad (10)$$

#### IV. EXPERIMENTS

We evaluate our approach on two datasets: the Left Atrial (LA) dataset [19] with 100 MR volumes ( $112 \times 112 \times 80$  crops for training and sliding windows for inference) and the Pancreas-CT dataset [20] with 82 CT scans ( $96 \times 96 \times 96$  crops and  $K$ -fold cross-validation for evaluation).

In line with recent state-of-the-art methods [3], [18], we employ four widely used metrics: Dice Similarity Coefficient (Dice), Jaccard Similarity Coefficient (Jaccard), 95% Hausdorff Distance (95HD), and Average Surface Distance (ASD). Dice and Jaccard evaluate regional overlap between predicted segmentations and ground truth, while 95HD and ASD assess boundary alignment. Specifically, 95HD measures the 95th percentile of surface-to-surface distances, and ASD computes the average surface distance. For comparison, we benchmark against several advanced methods, including UA-MT [21], SASSNet [22], DTC [23], SS-Net [24], MCF [18], CAML [12], CauSSL [8], and TraCoCo [13].

##### A. Results on LA

Table I presents the evaluation results on the LA dataset, including comparisons with standalone VNet and ResVNet under fully supervised (100% labeled) and semi-supervised (20% labeled) settings. C3S3 achieves a 95HD of 5.14, outperforming the fully supervised ResVNet (5.61) and closely approaching VNet (4.97). For ASD, C3S3 achieves 1.57, the closest to fully supervised performance. Among semi-supervised methods, C3S3 demonstrates substantial improvements over competitors. Notably, C3S3 achieves the highest Dice (91.24%), surpassing semi-supervised ResVNet (91.05%) and nearly matching fully supervised VNet (91.47%). Similarly, its Jaccard index (84.01%) is comparable to VNet (84.36%). These results highlight the precise boundary localization and segmentation accuracy of C3S3. Figure 4 visually confirms these findings, demonstrating that C3S3 captures finer details with superior boundary delineation.

##### B. Results on Pancreas-CT

Table II presents the evaluation results on the Pancreas-CT dataset, a more challenging task than LA MRIs due to the complex background of pancreatic CT volumes. The performance

TABLE I: Performance comparison on LA Dataset. The best and second-best performers are highlighted by bold and underline, respectively.

Method	Volumes used		Metrics			
	Labeled	Unlabeled	95HD(voxel)↓	ASD(voxel)↓	Dice(%)↑	Jaccard(%)↑
VNet	80(100%)	0	5.61	1.51	91.47	84.36
ResVNet	80(100%)	0	4.97	1.81	91.05	83.48
VNet	16(20%)	0	14.82	3.91	86.33	73.56
ResVNet	16(20%)	0	17.38	5.02	85.61	73.09
UA-MT (MICCAI 2019)	16(20%)	64	7.54	2.31	88.82	80.18
SASSNet (MICCAI 2020)	16(20%)	64	8.75	3.10	89.29	80.85
DTC (AAAI 2021)	16(20%)	64	7.36	2.15	89.45	81.01
SS-Net (MICCAI 2022)	16(20%)	64	6.96	1.84	90.21	82.36
MCF (CVPR 2023)	16(20%)	64	6.68	2.02	88.02	79.51
CAML (MICCAI 2023)	16(20%)	64	6.11	<u>1.68</u>	90.78	83.19
TraCoCo(TMI 2024)	16(20%)	64	<u>5.63</u>	1.79	<u>90.94</u>	<u>83.47</u>
<b>C3S3 (ours)</b>	16(20%)	64	<b>5.14</b>	<b>1.57</b>	<b>91.24</b>	<b>84.01</b>

TABLE II: Comparisons on Pancreas-CT Dataset. The best and second-best performers are highlighted by bold and underline, respectively.

Method	Volumes used		Metrics			
	Labeled	Unlabeled	95HD(voxel)↓	ASD(voxel)↓	Dice(%)↑	Jaccard(%)↑
VNet	62(100%)	0	7.21	1.69	80.72	68.90
ResVNet	62(100%)	0	7.38	1.63	79.84	67.35
VNet	12(20%)	0	17.69	4.64	64.04	48.17
ResVNet	12(20%)	0	18.95	5.63	66.45	51.12
UA-MT (MICCAI 2019)	12(20%)	50	17.06	5.23	74.02	60.10
SASSNet (MICCAI 2020)	12(20%)	50	13.89	3.56	73.58	59.73
DTC (AAAI 2021)	12(20%)	50	13.18	3.75	73.31	59.26
SS-Net (MICCAI 2022)	12(20%)	50	12.64	3.59	74.27	60.56
MCF (CVPR 2023)	12(20%)	50	11.59	3.27	75.00	61.27
CauSSL (ICCV 2023)	12(20%)	50	8.11	2.34	79.97	67.32
TraCoCo(TMI 2024)	12(20%)	50	<u>7.34</u>	<u>1.84</u>	<u>80.90</u>	<b>69.26</b>
<b>C3S3 (ours)</b>	12(20%)	50	<b>6.96</b>	<b>1.74</b>	<b>80.93</b>	<u>68.06</u>

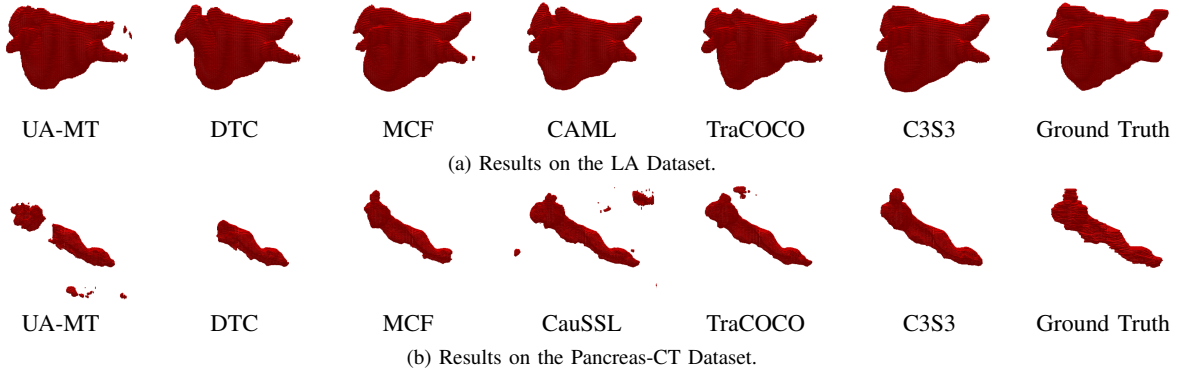


Fig. 4: Qualitative segmentation results. (a) Results on the LA dataset. (b) Results on the Pancreas-CT dataset.

of C3S3 is benchmarked against VNet and ResVNet under both fully supervised (100% labeled data) and semi-supervised (20% labeled data) settings. A four-fold cross-validation strategy is applied, using 12 labeled and 50 unlabeled samples to ensure a robust evaluation. C3S3 achieves outstanding results, particularly in boundary-sensitive metrics, with 95HD reduced to 6.96 and ASD to 1.74, significantly outperforming VNet and ResVNet trained with 20% labeled data. Compared to other semi-supervised models, C3S3 also achieves the best results in 95HD and ASD, highlighting its superior boundary delineation capabilities. In terms of Dice and Jaccard metrics, C3S3 achieves the highest Dice score (80.93%) and the second-highest Jaccard index (68.06%) among all semi-supervised methods. Figure 4 visually illustrates these results, showcasing C3S3's precise boundary localization and improved segmentation accuracy compared to other approaches.

## V. ABLATION STUDY

**Influence of hyperparameter  $\alpha$ :** In the DCC module, a weighted competition evaluates the practical performance of two backbone networks, selecting the most competitive one for pseudo-labeling. Analysis in the left atrial dataset shows that different values of the hyperparameter  $\alpha$  have a relatively mild impact on model performance, with the best results achieved when  $\alpha$  is set to 0.8, as shown in Table III.

**The Role of ODCL:** The ODCL module aims to enhance precision in target boundary delineation. Experimental comparisons between models with and without ODCL demonstrate significant difference, as shown in Table IV. Specifically, the 95HD decreases from 6.97 to 5.64, ASD drops from 2.03 to 1.81, the Dice coefficient improves from 87.12% to 89.92%, and the Jaccard index increases from 80.14% to 81.88%. These results confirm that ODCL effectively improves the accuracy



TABLE III: Influence of different  $\alpha$  values in the weighted competition mechanism on LA Dataset.

Parameter	Metrics			
	95HD(voxel)↓	ASD(voxel)↓	Dice(%)↑	Jaccard(%)↑
$\alpha$				
1.0	5.94	1.97	88.89	80.58
0.9	5.62	1.90	89.70	81.57
0.8	<b>5.14</b>	<b>1.57</b>	<b>91.24</b>	<b>84.01</b>
0.7	5.40	1.66	90.34	82.56
0.6	5.87	1.84	89.17	80.91
0.5	5.83	1.88	89.59	81.44
0.4	5.76	2.00	89.26	80.97
0.3	7.14	2.16	89.15	80.78
0.2	5.83	2.02	88.51	80.20
0.1	5.75	1.95	89.91	81.85
0.0	6.47	2.01	88.70	80.27

TABLE IV: Ablation results of ODCL and DCC Module.

Method	$L_{iuc}$	$L_{cp}$	Metrics			
			95HD(voxel)↓	ASD(voxel)↓	Dice(%)↑	Jaccard(%)↑
(1)	—	—	6.97	2.03	87.12	80.14
(2)	✓	—	5.64	1.81	89.92	81.88
(3)	—	✓	6.27	1.86	90.18	82.26
(4)	✓	✓	<b>5.14</b>	<b>1.57</b>	<b>91.24</b>	<b>84.01</b>

of target localization across various metrics.

**The Role of DCC:** A comprehensive ablation analysis highlights the impact of the DCC module in the C3S3 framework. DCC refines predictions by dynamically generating pseudo-labels through a competitive mechanism between dual backbones, using the better-performing backbone to supervise the other. This approach improves boundary delineation accuracy and enhances both boundary-sensitive and overlap metrics. As shown in Table IV, for example, 95HD decreases from 6.97 to 6.27, ASD reduces from 2.03 to 1.86, while Dice and Jaccard scores increase to 90.18% and 82.26%, respectively. These findings demonstrate DCC’s critical role in reducing pseudo-labeling errors and enhancing framework robustness. Additionally, its integration with other loss components further enhances the stability and robustness of the overall framework.

## VI. CONCLUSION

In this study, we propose a new semi-supervised method for 3D medical image segmentation, which is specifically designed to provide refined boundaries and then support enriched details. For that purpose, an outcome-driven contrastive learning module is meticulously developed, which further encompasses a spatial position binary masking scheme and a dual-space intersection-union loss, both of which significantly improve the boundary molding. Additionally, we have implemented a dynamic complementary competition module, which generates pseudo-labels by selecting the more competitive backbone via a weighted competition mechanism. Consistency regularization is also employed to reduce discrepancies between different predictions. The effectiveness of our C3S3 in SSMIS is validated on two public benchmark datasets. Concretely, new state-of-the-art scores have been achieved in terms of the vast majority of metrics. Moreover, we find that the proposed ODCL and DCC modules can behave as a plug-and-play role and be applicable across various fields.

## REFERENCES

[1] Ziyun Yang and Sina Farsiu, “Directional connectivity-based segmentation of medical images,” in *CVPR*, 2023, pp. 11525–11535.

[2] Duowen Chen, Yunhao Bai, Wei Shen, et al., “Magicnet: Semi-supervised multi-organ segmentation via magic-cube partition and recovery,” in *CVPR*, 2022, pp. 23869–23878.

[3] Yunhao Bai, Duowen Chen, Qingli Li, et al., “Bidirectional copy-paste for semi-supervised medical image segmentation,” in *CVPR*, 2023, pp. 11514–11524.

[4] Zhiqiang Shen, Peng Cao, Hua Yang, et al., “Co-training with high-confidence pseudo labels for semi-supervised medical image segmentation,” in *IJCAI*, 2023, pp. 4199–4207.

[5] Xiaomeng Li, Lequan Yu, Hao Chen, et al., “Transformation-consistent self-ensembling model for semi-supervised medical image segmentation,” *IEEE Transactions on Neural Networks and Learning Systems*, pp. 523–534, 2020.

[6] Yuyan Shi, Yichi Zhang, and Shasha Wang, “Competitive ensembling teacher-student framework for semi-supervised left atrium mri segmentation,” *BIBM*, pp. 2231–2234, 2023.

[7] Bo Qian, Yang Wen, and Bin Sheng, “Self-paced co-training and foundation model for semi-supervised medical image segmentation,” in *ICME*, 2024, pp. 1–6.

[8] Juzheng Miao, Cheng Chen, Furui Liu, et al., “Causl: Causality-inspired semi-supervised learning for medical image segmentation,” in *ICCV*, Oct. 2023, pp. 21369–21380.

[9] Hritam Basak and Zhaozheng Yin, “Pseudo-label guided contrastive learning for semi-supervised medical image segmentation,” in *CVPR*, 2023, pp. 19786–19797.

[10] Yucheng Shu, Jiaxin Xie, Lihong Qiao, et al., “C3t: Contrastive consistency cross-network learning for semi-supervised semantic segmentation,” in *ICME*, 2024, pp. 1–6.

[11] Lequan Yu, Shujun Wang, Xiaomeng Li, et al., “Uncertainty-aware self-ensembling model for semi-supervised 3d left atrium segmentation,” in *MICCAI*, 2019, pp. 605–613.

[12] Shengbo Gao, Ziji Zhang, Jiechao Ma, et al., “Correlation-aware mutual learning for semi-supervised medical image segmentation,” in *MICCAI*, 2023, pp. 98–108.

[13] Yuyuan Liu, Yu Tian, Chong Wang, et al., “Translation Consistent Semi-supervised Segmentation for 3D Medical Images,” *IEEE Transactions on Medical Imaging*, pp. 1–1, 2024.

[14] Xinze Li, Runlin Huang, Zhenghao Wu, et al., “Leveraging coral-correlation consistency network for semi-supervised left atrium mri segmentation,” *ArXiv*, 2024.

[15] Xianpeng Cao, Weixing Xie, Xianxing Cao, et al., “Icr-net: Semi-supervised medical image segmentation guided by intra-sample cross reconstruction,” in *ICME*, 2024, pp. 1–6.

[16] Xiangyu Zhao, Zengxin Qi, Sheng Wang, et al., “Rcps: Rectified contrastive pseudo supervision for semi-supervised medical image segmentation,” *IEEE Journal of Biomedical and Health Informatics*, pp. 251–261, 2023.

[17] Hengyang Liu, Pengcheng Ren, Yang Yuan, et al., “Uncertainty global contrastive learning framework for semi-supervised medical image segmentation,” *IEEE Journal of Biomedical and Health Informatics*, pp. 1–10, 2024.

[18] Yongchao Wang, Bin Xiao, Xiuli Bi, et al., “Mcf: Mutual correction framework for semi-supervised medical image segmentation,” in *CVPR*, 2023, pp. 15651–15660.

[19] Zhaohan Xiong, Qing Xia, Zhiqiang Hu, et al., “A global benchmark of algorithms for segmenting late gadolinium-enhanced cardiac magnetic resonance imaging,” *Medical Image Analysis*, p. 101832, 2020.

[20] Holger R. Roth, Le Lu, Amal A. Farag, et al., “Deeporgan: Multi-level deep convolutional networks for automated pancreas segmentation,” in *MICCAI*, 2015, p. 556–564.

[21] Lequan Yu, Shujun Wang, Xiaomeng Li, et al., “Uncertainty-aware self-ensembling model for semi-supervised 3d left atrium segmentation,” in *MICCAI*, 2019, pp. 605–613.

[22] Shuailin Li, Chuyu Zhang, and Xuming He, “Shape-aware semi-supervised 3d semantic segmentation for medical images,” in *MICCAI*, 2020, pp. 552–561.

[23] Xiangde Luo, Jieneng Chen, Tao Song, et al., “Semi-supervised medical image segmentation through dual-task consistency,” in *AAAI*, 2021, pp. 8801–8809.

[24] Yicheng Wu, Zhonghua Wu, Qianyi Wu, et al., “Exploring smoothness and class-separation for semi-supervised medical image segmentation,” in *MICCAI*, 2022, pp. 34–43.