Computers, Materials & Continua

Tech Science Press

DOI: 10.32604/cmc.2024.056478

ARTICLE

Stochastic Augmented-Based Dual-Teaching for Semi-Supervised Medical Image Segmentation

Hengyang Liu[1], Yang Yuan[1,*], Pengcheng Ren[1], Chengyun Song[1] and Fen Luo[2]

[1]School of Computer Science and Engineering, Chongqing University of Technology, Chongqing, 400054, China

[2]College of Artificial Intelligence, Chongqing Technology and Business University, Chongqing, 400067, China

*Corresponding Author: Yang Yuan. Email: yuan0224@stu.cqut.edu.cn

Received: 23 July 2024 Accepted: 06 October 2024 Published: 03 January 2025

ABSTRACT

Existing semi-supervised medical image segmentation algorithms use copy-paste data augmentation to correct the labeled-unlabeled data distribution mismatch. However, current copy-paste methods have three limitations: (1) training the model solely with copy-paste mixed pictures from labeled and unlabeled input loses a lot of labeled information; (2) low-quality pseudo-labels can cause confirmation bias in pseudo-supervised learning on unlabeled data; (3) the segmentation performance in low-contrast and local regions is less than optimal. We design a Stochastic Augmentation-Based Dual-Teaching Auxiliary Training Strategy (SADT), which enhances feature diversity and learns high-quality features to overcome these problems. To be more precise, SADT trains the Student Network by using pseudo-label-based training from Teacher Network 1 and supervised learning with labeled data, which prevents the loss of rare labeled data. We introduce a bi-directional copy-paste mask with progressive high-entropy filtering to reduce data distribution disparities and mitigate confirmation bias in pseudo-supervision. For the mixed images, Deep-Shallow Spatial Contrastive Learning (DSSCL) is proposed in the feature spaces of Teacher Network 2 and the Student Network to improve the segmentation capabilities in low-contrast and local areas. In this procedure, the features retrieved by the Student Network are subjected to a random feature perturbation technique. On two openly available datasets, extensive trials show that our proposed SADT performs much better than the state-of-the-art semi-supervised medical segmentation techniques. Using only 10% of the labeled data for training, SADT was able to acquire a Dice score of 90.10% on the ACDC (Automatic Cardiac Diagnosis Challenge) dataset.

KEYWORDS

Semi-supervised; medical image segmentation; contrastive learning; stochastic augmented

1 Introduction

1.1 Background

Segmenting medical images is a core task in medical image processing that plays an indispensable role in assisting diagnosis. By separating lesions or other regions of interest from the backdrop of the picture, it lays the groundwork for accurate localization and in-depth examination of the lesions. Large volumes of labeled data have enabled supervised medical image segmentation networks [1,2] to achieve impressive results. However, medical imaging variety and complexity provide difficulties. These include

high annotation costs and data scarcity, which restrict the training of models for segmenting medical images. Learning with semi-supervision (SSL) strategies train models by combining scarce labeled data with abundant unlabeled data. They effectively address the issue of sparse annotations by leveraging precious labeled information to provide a wealth of a priori knowledge for unlabeled images, thereby helping to establish connections between labeled and unlabeled images. As a result, medical image segmentation techniques based on SSL have emerged as a promising area of study.

SSL has advanced significantly in the realm of medical image segmentation [3,4] in recent years. Its core strategies can be summarized as: pseudo-supervision techniques [5], prior knowledge-based techniques [4], and consistency regularization techniques [6]. Learning strategies based on the Mean Teacher structure [7–9] or cross-task consistency learning strategies [10] have demonstrated their usefulness. Recently, consistency regularization paradigms have gained more attention through the application of strong-weak augmentation [6] or the introduction of contrastive learning [11]. Strong-weak augmentation strategies involve processing the same image in different ways and use consistency loss to ensure that predictions from different views remain consistent. By calculating the distances (measured by cosine similarity) between positive and negative sample pairs taken from the data, contrastive learning seeks to reduce the gap between similar characteristics and increase the distance between dissimilar features. The significance of contrastive learning approaches is heightened in clinical settings due to the ease with which unlabeled data may be obtained.

1.2 Motivation

A crucial presumption for SSL's effectiveness is that the dataset's labeled and unlabeled data have the same distribution. However, with a relatively little quantity of labeled data in clinical data collection, it is not viable to extrapolate the distribution of the complete dataset. An empirical distribution mismatch between labeled and unlabeled data results from this [12]. To address this issue, References [6,13] proposed a copy-paste augmentation technique, generating partially labeled images by copying small patches from labeled images and pasting them onto unlabeled images. Strictly depending on these partially labeled images means that the Teacher Network's pseudo-labels are crucial. Confirmation bias and the widespread disregard of sparsely labeled data may result from this. Additionally, these methods perform poorly in low-contrast regions and local area segmentation.

In order to solve the aforementioned problems, we created SADT specifically for medical image segmentation tasks. In order to reduce confirmation bias, we suggest using a bidirectional copy-paste mask with progressive high-entropy filtering to weed out untrustworthy pixels in the early training phases. We introduce the use of pseudo-supervised and supervised learning to train the Student Network, addressing the issue of losing valuable labeled information. For mixed images, we design DSSCL in the feature spaces of Student Network with random feature perturbation and Teacher Network 2 to learn high-quality features, thereby improving segmentation capabilities in low-contrast and local regions.

1.3 Contributions

Two publicly accessible datasets, the ACDC [14] and ISIC (International Skin Imaging Collaboration) [15] datasets, were used to validate the designed SADT. Through a series of comparison trials, our offered SADT technique shows greater efficiency than existing leading methods in the semi-supervised medical image segmentation area, namely Bidirectional Copy-Paste (BCP [13]). Using the ACDC data collection, our model using just 5% labeled data achieved an 88.68% Dice score, and at 10% labeled data, it outperformed the leading BCP method by 1.26%. The distribution gap between labeled and

unlabeled data is reduced by the provided SADT. It preserves scarce labeled information. It improves segmentation in low-contrast and local regions. Ablation studies were conducted to further confirm the efficiency of each component. We used Unet [1] as segmentation network baseline.

To sum up, this paper's primary contributions are as follows:

(1) Proposing a bidirectional copy-paste mask with progressive high-entropy filtering to reduce confirmation bias caused by unreliable pixels and narrow the discrepancy in distribution between data with and without labels.
(2) Proposing a dual-teacher-assisted training strategy that combines supervision, pseudo-supervision, and contrastive learning to assist in training the Student Network.
(3) Designing a Stochastic Feature Perturbation Pool (SFPP), which applies perturbations of random intensity to the features to increase feature diversity and thereby improve the capability of segmenting low-contrast areas.
(4) Designing DSSCL to encourage the learning of high-quality features in both deep and representational layers, thereby improving the model's low-contrast area and local segmentation performance.
(5) Proving the efficacy of the offered SADT and validating it on two public datasets.

2 Related Work

2.1 Medical Image Segmentation

In medical image segmentation tasks, research based on the Unet model targeting various refinement directions has achieved significant improvements. Self-attention components were included by Azad et al. [16] in order to lower computing costs and improve the segmentation performance of the model. Skip connections are optimized using MultiResUnet [17] in order to include multi-scale characteristics. Rahman et al. [18,19] stated cascaded multi-scale features and attention mechanisms to address inconsistencies between local features. A universal medical image segmentation method that handled data heterogeneity and annotation variability was published by Gao et al. [20]. However, our method differs from these by leveraging an enormous volume of unlabeled data alongside scarce labeled data to train models for segmenting medical images, tackling challenges such as the difficulty of pixel-level annotations.

2.2 Semi-Supervised Learning (SSL)

SSL combines both marked and unmarked data to train a model. Unsupervised methods have higher uncertainty, while supervised training faces difficulties in data annotation. Therefore, research on SSL for semantic segmentation has rapidly developed [21–24]. Recent SSL strategies using strong and weak augmentations to enforce consistency between the results of teacher and student networks have shown good performance [25–27], where a unique method of dual-stream perturbation was presented by Yang et al. [28]. A pseudo-label correction technique was presented by Zhao et al. [29] in order to lessen the effect of noisy pseudo-labels. However, these approaches do not effectively leverage the connection between labeled and unlabeled data. Given the scarcity and diversity of medical image data, pixel-level annotation is particularly challenging. As a result, semi-supervised learning techniques for medical image segmentation have to be able to narrow the distribution gap between labeled and unlabeled images while fully using the important labeled data.

2.3 Contrastive Learning (CL)

CL is fundamentally a form of metric learning that utilizes the information in a dataset to learn a powerful feature representation space. Regarding the semi-supervised segmentation of medical images, CL techniques are employed to effectively learn features [9]. Wang et al. [22] discussed use CL to move sparse area anchor features toward highly regularized key centers. ELN [23] demonstrates how the quality of features the model extracts is enhanced by patch-based per-pixel contrastive loss. The proposed DSSCL method designs a contrastive loss that considers both detailed deep features and geometric edge features across the entire batch. This approach aims to cluster similar features in the feature space, improving the quality of learned features and enhancing discrimination in low-contrast and local pixel regions.

2.4 Copy-Paste Data Enhancement Technology

Copy-paste is an effective data augmentation technique. In SSL, it helps to increase data diversity and establish connections between labeled and unlabeled data. Augmentation matters [26] adaptively copy-pastes low-confidence unlabeled samples onto labeled ones. Cut-paste consistency [6] matches unlabeled with labeled images based on color similarity and pastes more matching labeled lesion information onto unlabeled images using Gaussian blurring. Chi et al. [30] proposed an adaptive bidirectional displacement to generate new unlabeled samples, addressing the impact of perturbation methods. BCP [13] copy-pastes labeled onto unlabeled images using zero-centered masks, processing both types of images in the same way to reduce their distribution gap.

Inspired by [13,23,26,28], the proposed SADT model addresses significant loss of label information and low-quality pseudo-labels leading to confirmation bias and poor segmentation performance in low-contrast and local regions in challenges involving semi-supervised medical image segmentation using a dual-teacher assisted training strategy.

3 Methodology

This section describes the construction of a semi-supervised learning technique for medical picture segmentation, including important terminology and symbols. Firstly, a summary of the SADT framework is given in Section 3.1. Secondly, Section 3.2 describes how pseudo-labels are copied and pasted onto true labels using a progressive high-entropy filtering mask to achieve partial supervision is described. Thirdly, in Section 3.3, the proposed SFPP strategy to enhance feature diversity in a simplified manner is detailed. Finally, in Section 3.4, the proposed DSSCL method to learn high-quality features and improve segmentation capabilities in local and low-contrast regions is described.

3.1 Overview

During the training process, the medical image dataset includes N labeled and M unlabeled data (where $N \ll M$). Therefore, the training dataset $D = D^l \cup D^u$ includes labeled data $D^l = \left\{\left(\mathrm{X}_i^l, \mathrm{Y}_i^l\right)\right\}_{i=1}^{N}$ and unlabeled data $D^{\mathrm{u}} = \left\{\mathrm{X}_i^u\right\}_{i=N+1}^{M+N}$. As shown in Fig. 1, a student network and two teacher networks make up the recommended structure, forming a dual-teacher-assisted training framework, parameterized by the parameters θ_s, θ_{t1}, and θ_{t2}, respectively. The improved BCP method utilizes the copy-paste technique to process images. Specifically, there are two processing methods, as illustrated in Fig. 2: copy-pasting between labeled data generates X^{mix}; bidirectional copy-paste transferring between labeled and unlabeled data generates X^{in} and X^{out}.

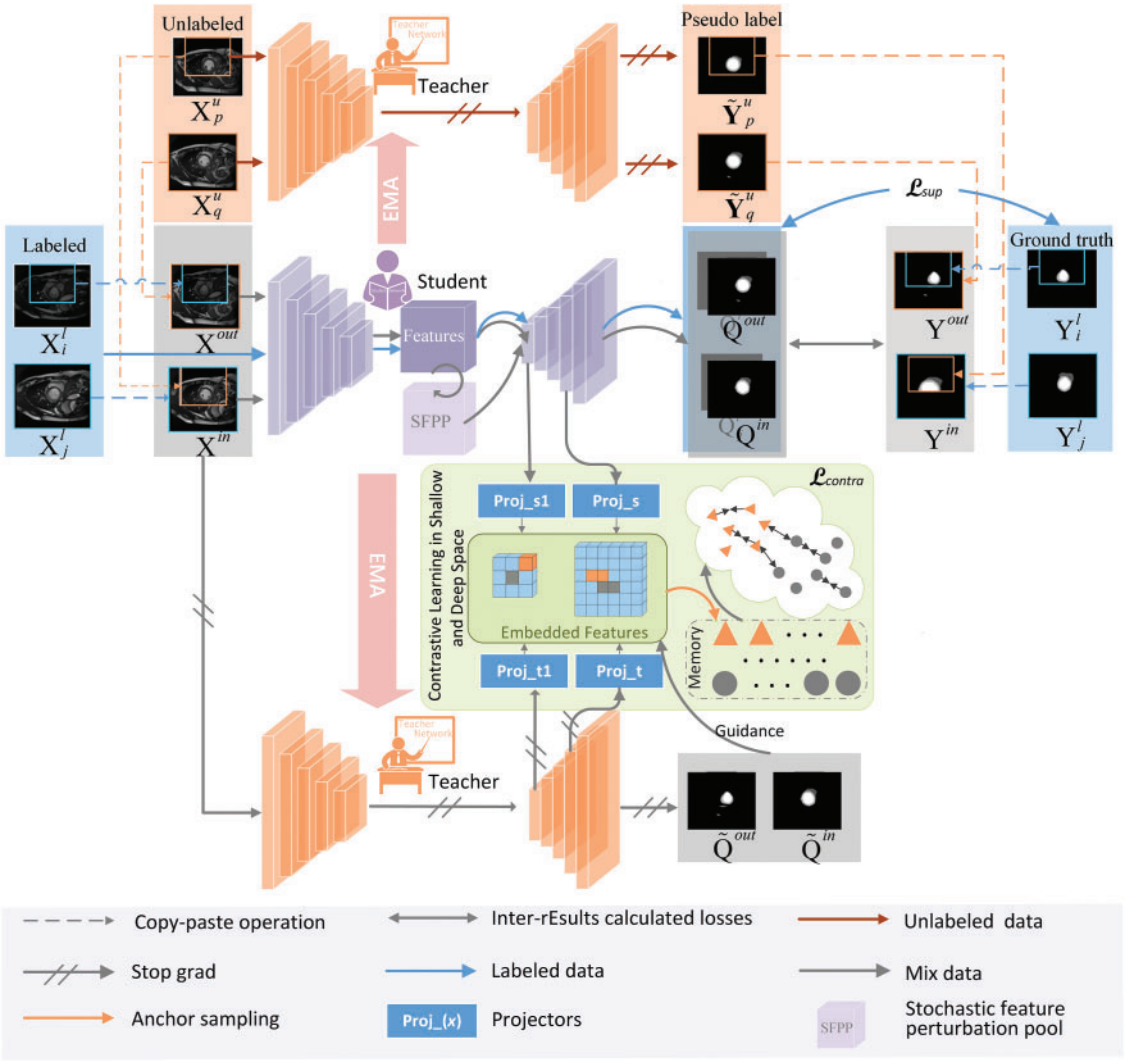

Figure 1: Overall framework

The training phase of the proposed SADT consists of two steps. The first step involves pre-training the Unet network using labeled data X^l and mixed data X^{mix}. Joint training employing contrastive learning, supervised learning, and pseudo-supervised learning is the second phase. Ultimately, Stochastic gradient descent is used to optimize the Student Network, and an exponential moving average of the Student Network weights is used to update the Teacher Networks: $\theta_{t1} = \alpha\theta_{t1} + (1-\alpha)\,\theta_s$; $\theta_{t2} = \alpha\theta_{t2} + (1-\alpha)\,\theta_s$, where $\alpha = 0.999$ is a commonly used momentum parameter.

A batch of tagged and unlabeled samples is obtained in each round. By reducing the supervised loss, the partial pseudo-supervised loss, and the contrastive loss, the Student Network is to be optimized. Thus, the total loss of training the Student Network is: $L = L_{sup} + L_{part} + L_{constrat}$.

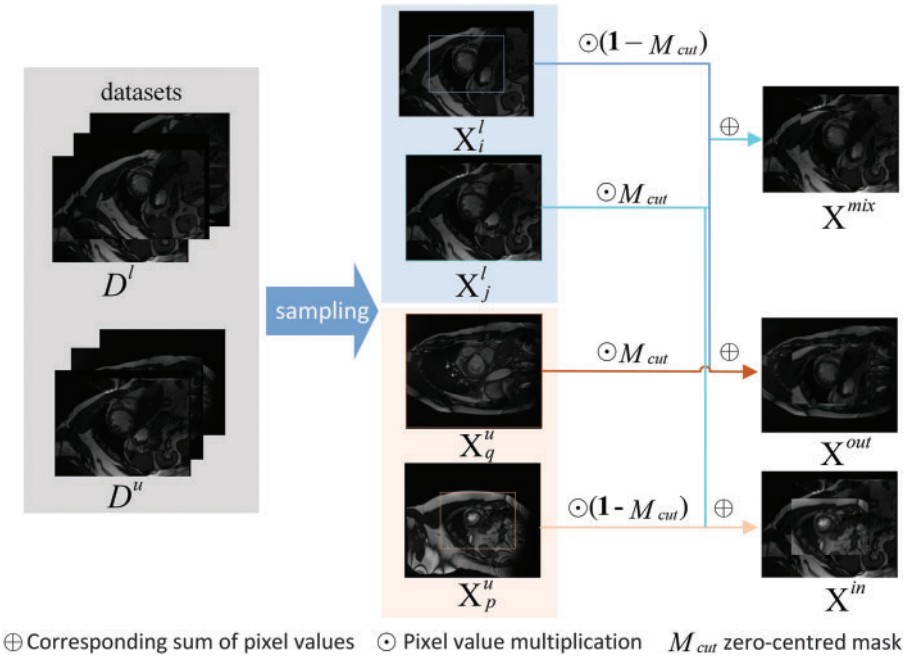

Figure 2: The detailed process for generating three types of mixed images through copy-paste

3.2 Effective Copy-Paste

Pre-training In order to properly acquire previous knowledge of picture augmentation noise and labeling information, the pre-training procedure for the BCP method is enhanced. $\left(X^l, X^{mix}\right)$ is pretrained on the main segmentation network with a Unet structure, obtaining prediction probabilities $\left(P^l, P^{mix}\right)$. Thus, the pre-training loss function is $L_{pre} = L_{seg}\left(P^l, Y^l\right) + L_{seg}\left(P^{mix}, Y^{mix}\right)$, where L_{seg} is a linear combination of the cross-entropy loss and the dice loss.

Bidirectional copy-paste mask with progressive high-entropy filtering During the early stages of training with pseudo-supervision, some pixels are difficult to confidently classify, leading to severe confirmation bias. Therefore, we design a progressive high-entropy filtering mask for their pseudo-labeling to reduce the unreliability of pseudo-supervision. That is, in the initial stage, only pixels with low entropy in the Teacher Network's predictions are selected. As learning progresses, the range of predicted pixels involved in loss calculation gradually expands. Specifically, firstly, obtain the prediction probabilities according to $p_p^u = f\left(X_p^u; \theta_{t1}\right)$; $p_q^u = f\left(X_q^u; \theta_{t1}\right)$, where $X_p^u, X_q^u \in D^u$, $p \neq q$. Secondly, the entropy of the prediction probabilities is calculated using $\mathrm{H}\left(\hat{p}_{ij}\right) = -\sum_{c=1}^{C} \hat{p}_{ij}^c \log \hat{p}_{ij}^c$, where, $\hat{p}_{ij}$ represents the prediction probability of the j-th pixel in the i-th image by the Teacher Network. Finally, filter out pixel predictions with entropy values higher than the β_t percentile, where β_t decreases from β_0 to 0 over t iterations, with a decrement method of $\beta_t = \beta_0\left(1 - t/T\right)$, and T is the total number of iterations. The threshold $\eta\beta_t$ is extracted from the current batch's entropies $\left\{H\left(\hat{p}_{ij}\right)\right\}$ as the $\eta\beta_t$-th percentile value. This defines a progressive high-entropy filtering mask $M_{filter} \in \{0\ ,\ 1\}^{W \times H}$, which indicates whether a pixel meets the $H\left(p_{ij}^u\right) < \eta\beta_t$ condition (1) or not (0).

A zero-centered mask $M_{cut} \in \{0, 1\}^{W \times H}$ is designed. The zero-value region's dimensions is $\beta W \times \beta H$, where $\beta = 2/3$. After that, processing images with and without labels produces X^{in} and X^{out}

according to

$$X^{in} = X_j^l \odot M_{cut} + X_p^u \odot (\mathbf{1} - M_{cut}); X^{out} = X_q^u \odot M_{cut} + X_i^l \odot (\mathbf{1} - M_{cut}), \tag{1}$$

where $X_i^l, X_j^l \in D^l$, $i \neq j$, $\mathbf{1} \in \{1\}^{W \times H}$, and $\odot$ denote element-by-element multiplication.

According to $\hat{Y}^u = argmaxP^u$, calculate the pseudo-labels $\hat{Y}^u$ for the unlabeled predictions. Train the student network using (X^{in}, X^{out}) and its corresponding partial pseudo-labels (Y^{in}, Y^{out}), together with $\left(X^l, X^{mix}\right)$ and its corresponding labels $\left(Y^l, Y^{mix}\right)$. The calculation methods for Y^{in} and Y^{out} are as follows:

$$Y^{in} = Y_j^l \odot M_{cut} + \hat{Y}_p^u \odot \left[(1 - M_{cut}) \odot M_{filter}\right]; Y^{out} = \hat{Y}_q^u \odot \left(M_{cut} \odot M_{filter}\right) + Y_i^l \odot (1 - M_{cut}). \tag{2}$$

The loss computation supervised by Y^{in} and Y^{out} is as follows:

$$L^{in} = L_{seg}\left(Q^{in}, Y^{in}\right) \odot M_{cut} + \alpha L_{seg}\left(Q^{in}, Y^{in}\right) \odot \left[(1 - M_{cut}) \odot M_{filter}\right], \tag{3}$$

$$L^{out} = L_{seg}\left(Q^{out}, Y^{out}\right) \odot (1 - M_{cut}) + \alpha L_{seg}\left(Q^{out}, Y^{out}\right) \odot \left(M_{cut} \odot M_{filter}\right), \tag{4}$$

where $\alpha = 0.5$ is used to manage the loss function's unlabeled pixel contribution. Q^{in} and Q^{out} represent the predicted output of the Student Network, which are computed according to

$$Q^{in} = f\left(X^{in}; \theta_s\right); Q^{out} = f\left(X^{out}; \theta_s\right). \tag{5}$$

Therefore, the partial supervised loss is computed as $L_{part} = L^{in} + L^{out}$. Additionally, the Student Network is trained using labeled images, with the supervised loss represented as $L_{\text{sup}} = L_{seg}\left(Q^l + Q^{mix}; Y^l + Y^{mix}\right)$, where $\left(Q^l + Q^{mix}\right) = f\left(X^l + X^{mix}; \theta_s\right)$.

3.3 Deep-Shallow Spatial Contrastive Learning (DSSCL)

Pseudo-supervision is weak in learning feature patterns of the entire dataset and in segmenting local low-contrast regions. Therefore, we introduce a new contrastive learning paradigm for mixed images—DSSCL—forming a dual-teacher-assisted training structure. As shown in Fig. 1, the Student Network and Teacher Network 2 each contain two feature projectors $(H_d * (\cdot), H_s * (\cdot))$. Positive and negative sample pairs are selected in the deep and representation layer feature spaces, and a patch-based local contrastive learning loss is adopted (i.e., dividing the feature maps of the current batch into small patches of size $h \times h$), enabling effective attention to the feature representation of local regions.

In each feature embedding space, define Ω_p^i as the collection of pixels in the same class as the original pixel i, and Ω_n^i as those not belonging to the same class. Also define d as the distance function (i.e., $d(f_1, f_2) = \exp(\cos(f_1, f_2)/\tau)$), where cos reflects the similarity of the cosine, and τ is the hyper-parameter for temperature. Therefore, extracting deep and shallow features from the decoder allows for the calculation of the contrastive loss in the deep feature space L_{contra_d} and the contrastive loss in the shallow feature space L_{contra_s} by

$$L_{contra\_x} = -\frac{1}{|V|} \sum_{i \in V} \sum_{j \in \Omega_p^i} \log \frac{d\left(f_i, \tilde{f}_j\right)}{d\left(f_i, \tilde{f}_j\right) + \sum_{k \in \Omega_n^i} d\left(f_i, \tilde{f}_k\right)} \tag{6}$$

where V represents the set of pixels of mixed images (X^{in}, X^{out}) in the current batch, and f_i as well as $\tilde{f}_i$ reflect the characteristic embedding of pixel i from the Student Network and Teacher Network 2,

correspondingly. The whole loss of contrast is $L_{contra} = \lambda_d L_{contra\_d} + \lambda_s L_{contra\_s}$, where λ_d and λ_s are the hyperparameters controlling the contributions of deep and shallow contrastive learning, respectively.

3.4 Stochastic Feature Perturbation Pool (SFPP)

In medical image segmentation tasks, the insufficient segmentation capability in low-contrast regions is particularly prominent. Medical images have a complicated backdrop with very little contrast between the target and background. To enhance feature diversity and improve the model's segmentation performance in low-contrast regions, a random feature perturbation pool is designed. This method applies perturbations of random intensity to the features extracted by the Student Network. Existing feature processing uses a fixed dropout method, which limits the effectiveness of semi-supervised learning. Random feature perturbation is a simpler way to generate feature diversity, whereas fixing multiple perturbation types may overly distort the features, harming their distribution.

Specifically, it randomly selects no more than K perturbation methods from the feature perturbation pool to generate a feature perturbation strategy, rather than using a fixed number. The candidate methods are shown in Table 1, including various feature perturbation methods at different scales, with their methodologies detailed in the corresponding clarification columns. Therefore, the designed randomness-based feature perturbation method can more simply generate diverse features. At the same time, our proposed fully random selection strategy does not severely impact the feature distribution, which makes it more appropriate for semi-supervised learning.

Table 1: The feature perturbation methods and their perturbation scales within the SFPP

Methods	Scale	Clarification
Feature_None	None	Not apply any perturbation.
Dropout	(0.4/0.2)	Use dropout technology to set (40/20)% of the features to 0.
Random	(0.7/0.4)	Generate random numbers within the range [-scale, scale] that follow a uniform distribution and add them as noise to the feature map.
Gaussian	(0.6/0.3)	Create a normal distribution with a mean of 0 and a scale standard deviation of random numbers, then add them as noise to the feature map.

4 Experimentation

4.1 Datasets

ACDC Dataset Three kinds of annotated short-axis cardiac MR-cine (Magnetic Resonance Cine) images are included in the ACDC [14] dataset. The detailed characteristics of the dataset are shown in the first row of Table 2. ACDC is aimed at segmenting LV, RV, and Myo in cardiac dynamic magnetic resonance imaging diastolic (ED) and systolic (ES) frames. There are five cases, with 30 cases per category. The data split is conducted following the BCP method, having 100 examples established as patient scans for testing, validation, and training, respectively, of 70, 10, and 20.

ISIC Dataset The ISIC dataset [15] is dermoscopic image dataset aimed at identifying melanoma regions through image recognition. The detailed characteristics of the dataset are shown in Table 2's second row. The experimental design from earlier work, with 1815 images serving as the training set and 779 images serving as the validation set, to assure experiment fairness. Of the images in the training set, 5% (91) and 10% (181) include labels and are applied to various semi-supervised trial scenarios.

Table 2: The characteristics of the two datasets

Datasets	Dimension	Modal	Structure	Classifications	Number of cases
ACDC	3D (sliced 2D)	MR	Hearts	RV: right ventricle; Myo: myocardium; LV: left ventricle.	(1) Normalcy; (2) Myocardial infarction with systolic heart failure; (3) Dilated cardiomyopathy; (4) Hypertrophic cardiomyopathy; (5) Right ventricular anomalies;
ISIC	2D	Dermatoscope	Skin	Melanomas	Melanoma

4.2 Evaluation Indicators

Four widely-used assessment measures were chosen in order to assess the effectiveness of SADT: Average Surface Distance (ASD), 95% Hausdorff Distance (95HD), Dice Score (%), and Jaccard Score (%). ASD determines the average distance between the borders, 95HD estimates the distance of the nearest point between the two regions, and Dice Score and Jaccard Score compute the proportion of overlapping areas between two target areas.

4.3 Implementation Details

ACDC Dataset To ensure fairness, the experimental parameters were set following the BCP method, with Unet selected as the backbone network. The input picture size was set to 256 × 256 (2D slices) during training, and 170 × 170 was chosen as the size of the mask M_{cut} was zero-valued area. Furthermore, the batch size, pre-training iteration count, and self-training iteration count were set to 24, 10 and 30k, respectively.

ISIC Dataset The experimental parameters were set following the Uncertainty-Guided Pixel Contrastive Learning (UGCL [31]) method, with Unet serving as the backbone structure. 8 labeled images were included in the batch size of 16. Pre-training and self-training iteration counts were set to 2 and 6 k, respectively, during the training phase of all trials.

We decided on the stochastic gradient descent (SGD) optimizer, which has a momentum of 0.9 and a weight decay of 0.0005. A polynomial scheduling method was used to lower the initial learning rate from 0.01 to 0.001. We trained using an NVIDIA RTX 3090 GPU and used the PyTorch package to implement the proposed approach.

4.4 Comparison with State-of-the-Art Results

ACDC Dataset The offered SADT was trained using only 5% and 10% of the labeled data, and the test was conducted on the average segmentation results performed for the four categories (background, left ventricle, right ventricle, and myocardium) using the ACDC dataset. The comparative experimental results are shown in Table 3. From the results, our method outperforms all the most recent state-of-the-art (SOTA) methods in terms of performance. The Dice score is enhanced by 1.26% as

compared to the BCP approach at a labeling rate of 10%. Visualization results comparing the method to advanced methods are shown in Fig. 3. Consequently, our approach fully utilizes the labeled data while also reducing the distribution gap between labeled and unlabeled data. Furthermore, it enhances the segmentation capabilities in low-contrast and local regions, achieving more accurate segmentation of edge region pixels.

Table 3: Comparing the ACDC dataset with cutting-edge semi-supervised segmentation techniques

Method	Scans used		Metrics			
	Labeled	Unlabeled	Dice ↑	Jaccard ↑	95HD ↓	ASD ↓
U-Net [1]	3 (5%)	0	47.83	37.01	31.16	12.62
U-Net	7 (10%)	0	79.41	68.11	9.35	2.70
U-Net	70 (All)	0	91.44	84.59	4.30	0.99
UA-MT [32]	3 (5%)	67 (95%)	46.04	35.97	20.08	7.75
SASSNet [33]			57.77	46.14	20.05	6.06
DTC [10]			56.90	45.67	23.36	7.39
URPC [34]			55.87	44.64	13.60	3.74
MC-Net [35]			62.85	52.29	7.62	2.33
SS-Net [5]			65.83	55.38	6.67	2.28
BCP [13]			87.59	78.67	1.90	0.67
Ours			**88.68**	**80.26**	**1.69**	**0.51**
UA-MT [32]	7 (10%)	63 (10%)	81.65	70.64	6.88	2.02
SASSNet [33]			84.50	74.34	5.42	1.86
DTC [10]			84.29	73.92	12.81	4.01
URPC [34]			83.10	72.41	4.84	1.53
MC-Net [35]			86.44	77.04	5.50	1.84
SS-Net [5]			86.78	77.67	6.07	1.40
BCP [13]			88.84	80.62	3.98	1.17
Ours			**90.10**	**82.47**	**2.12**	**0.71**

ISIC Dataset The proposed model was also trained on the ISIC dataset in order to confirm the model's capacity for generalization. The average segmentation results performance for the two categories (melanoma and background) under conditions where the labeled data make up just 5% and 10% of the training data is shown in Table 4. The outcomes demonstrate that, for this dataset, our approach performs better than cutting-edge techniques. Dice and Jaccard scores were employed as assessment measures in accordance with the UGCL methodology. As demonstrated by the performance metrics when the model was trained with solely labeled data, as indicated in the first row of data in Table 4. Our proposed technique performs much better than previous methods under different semi-supervised scenarios. With a labeling ratio of 5%, the Dice score improves by 8.49% compared to the advanced UGCL method.

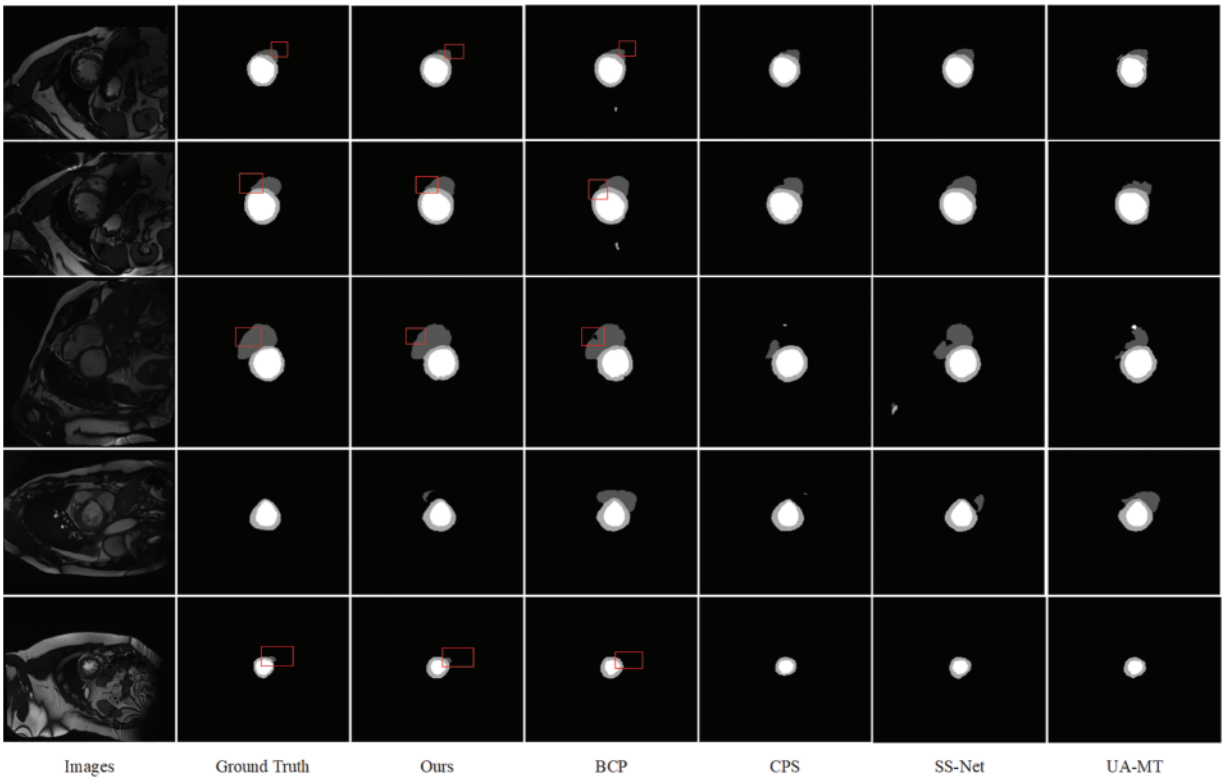

Figure 3: Comparison of several semi-supervised segmentation methods, prediction results of models trained with 10% labeled data on the ACDC dataset, and visualization of images and their labels

Table 4: Comparison of different methods on the ISIC dataset under different semi-supervised settings

Method	Metrics (5%)		Metrics (10%)	
	Dice ↑	Jaccard ↑	Dice ↑	Jaccard ↑
U-Net	67.02	53.49	68.91	56.69
MT [7]	69.87	57.19	70.64	58.63
UA-MT [32]	69.33	56.20	75.67	63.42
EM [36]	66.11	51.89	71.10	58.54
DCT [37]	70.09	57.13	75.98	63.68
CCT [38]	69.53	56.94	73.42	62.13
CPS [21]	71.87	57.35	78.09	65.74
UGCL [31]	72.67	57.85	79.48	67.25
Ours	**81.16**	**68.99**	**82.03**	**70.57**

4.5 Ablation Experiments

Through ablation experiments, the importance of each component was validated, including the improved pre-training strategy, progressively filtered bidirectional copy-paste mask, the dual-teacher

structure combining supervised, pseudo-supervised, and DSSCL, and the SFPP strategy. The outcomes of these ablation tests are shown in Table 5, which also details the improvement process and considerable impact of our technique on the ACDC dataset at a 10% labeling rate.

Table 5: Ablation experiments were conducted on each component of the method on the ACDC dataset, with 10% labeled data

Pre_training	L_{sup}	$M_{cut} \odot M_{filter}$	L_{contra}	SFPP	Dice
					88.84
√					89.53
√	√				89.67
√	√	√			89.74
√	√	√	√		89.98
√	√	√	√	√	**90.10**

4.5.1 *The Effectiveness of the Pre-Training Strategy*

In the lead-up to training, the Unet was trained using the original labeled data and mixed images obtained through copy-paste operations between labeled data. Three pre-training input data schemes are designed to investigate the impact of different strategies on model training. As shown in Table 6, in each mini-batch, 'a' represented the first half of the labeled data, and 'b' represented the second half. $a \rightarrow b$ indicated copy-pasting the part 'a' of the images onto the edge region of 'b', while $b \rightarrow a$ indicated pasting onto the center region of 'b', and $a + b$ represented all the original images. The purpose of the pre-training was to help the model learn about the noise introduced by data augmentation early on, while avoiding the loss of valuable labeled information as seen in the first strategy. Our pre-training technique outperforms employing simply mixed images, according to experimental results; the second strategy introduces an excessive amount of redundant data.

Table 6: The ablation experiment for the pre-training input data strategy

Pre-training strategies	Dice	Jaccard	95HD	ASD
$a \rightarrow b$	89.76	81.97	5.43	1.23
$a \rightarrow b, b \rightarrow a, a + b$	88.77	80.47	8.40	2.17
$a \rightarrow b, a + b$ (**ours**)	**90.10**	**82.47**	**2.12**	**0.71**

4.5.2 *Validity of Filtered Pseudo-Label*

The bidirectional copy-paste mask between pseudo-labels and true labels is generated through a dot product of the progressive high-entropy filter mask and the zero-centered mask. This process simultaneously reduces the gap in data distribution and the confirmation bias in pseudo-supervision. As shown in Table 7, M_{cut} represents the direct bidirectional copy-paste of pseudo-labels with true labels, whereas $M_{filter} \odot M_{cut}$ indicates the bidirectional copy-paste of pseudo-labels after undergoing

progressive filtering. Experimental results show that using a copy-paste mask with progressive high-entropy filtering reduces confirmation bias caused by low-quality pseudo-labels in the early stages of training, effectively improving the training efficiency of pseudo-supervision.

Table 7: The ablation experiment for pseudo-label copy-paste

Masking strategy	Dice	Jaccard	95HD	ASD
M_{cut}	89.88	82.07	3.14	0.79
$M_{filter} \otimes M_{cut}$ (ours)	**90.10**	**82.47**	**2.12**	**0.71**

4.5.3 The Effectiveness of the Dual-Teacher-Assisted Training

To effectively guide the training of the Student Network, Teacher Network 1 provides pseudo-supervised learning, while Teacher Network 2 offers contrastive learning for the feature space of (X^{in}, X^{out}). Meanwhile, labeled data provide supervised learning. By leveraging partial pseudo-supervision to reduce the distribution gap between labeled and unlabeled data, we design a joint training paradigm of pseudo-supervision and contrastive learning for the Student Network, effectively addressing the challenge of poor segmentation in low-contrast regions. Experimental results, as shown in Table 8, L_{part} denotes using only partial pseudo-supervision; $L_{part} + L_{\text{sup}}$ denotes partial pseudo-supervision combined with supervised learning; $L_{part} + L_{\text{sup}} + L_{cl}$ denotes the proposed dual-teacher-assisted training structure. The last row validates the introduction of a consistency loss function, which enforces consistent output results, affecting the quality of contrastive learning clustering features in the feature space. According to experimental findings, the recommended dual-teacher networks can help the Student Network be trained using various learning paradigms to provide superior results.

Table 8: The ablation experiment for the dual-teacher assisted training strategy

Model	Dice	Jaccard	95HD	ASD
L_{part}	89.70	81.88	2.16	0.73
$L_{part} + L_{\text{sup}}$	89.74	81.92	3.04	1.96
$L_{part} + L_{\text{sup}} + L_{contra}$ (ours)	**90.10**	**82.47**	**2.12**	**0.71**
$L_{part} + L_{con} + L_{\text{sup}} + L_{contra}$	89.79	82.00	3.64	1.10

4.5.4 The Effectiveness of the SFPP

By perturbing features randomly, we aim to simplify the increase of feature diversity in contrastive learning. The effectiveness of randomly selecting up to K perturbation methods from a SFPP (with experimental validation of the value of K) *vs*. choosing two fixed perturbation methods was validated. When K is too large, it damages the feature distribution; when too small, the number of random selections may be zero, leading to no perturbation effect. Table 9 illustrates the experimental findings that randomly selecting up to two feature perturbation methods more effectively improves the performance of semi-supervised learning for medical image segmentation tasks.

Table 9: The ablation experiment for the random feature perturbation pool

Strategy	Dice	Jaccard	95HD	ASD
$K = 1$	89.53	81.64	3.68	1.06
$K = 2$ (ours)	**90.10**	**82.47**	**2.12**	**0.71**
$K = 3$	90.02	82.37	4.28	0.99
Two fixed perturbations	88.78	80.45	7.34	2.24

4.5.5 The Effectiveness of DSSCL

By designing contrastive learning losses at both the deep and representation layers of the decoder, the aim is to push features away from other classes and shorten the distance between comparable features. As shown in Table 10, the designed DSSCL was compared with contrastive learning at the Representation Layer alone (RLCL). The values of λ_d and λ_s in the contrastive learning loss were also validated. The outcomes of our experiments demonstrate the great effectiveness of our planned contrastive learning technique, where $\lambda_d = 0.5, \lambda_s = 0.5$. Values of λ that are too high can cause the overall training to focus excessively on the contrastive learning method, potentially neglecting other training strategies. An imbalance in λ values at the deep and shallow layers can lead to a greater focus on the feature space with the higher λ value during contrastive learning, easily neglecting the other feature space. Therefore, balancing the contribution weights is crucial for enhancing the model's overall performance.

Table 10: The ablation experiment for DSSCL

Strategy	Dice	Jaccard	95HD	ASD
RLCL	88.89	80.56	8.74	2.35
DSSCL (ours)	**90.10**	**82.47**	**2.12**	**0.71**
$\lambda_s = 1, \lambda_d = 1$	89.51	81.47	6.35	1.99
$\lambda_s = 1, \lambda_d = 0.5$	89.79	81.99	3.19	0.84
$\lambda_s = 0.5, \lambda_d = 1$	89.64	81.72	5.52	1.30
$\lambda_s = 0.5, \lambda_d = 0.5$ (ours)	**90.10**	**82.47**	**2.12**	**0.71**

5 Discussion

We introduce SADT, a novel semi-supervised learning process, to improve medical picture segmentation accuracy. SADT combines random information enhancement techniques with DSSCL. Specifically, we adopt two teachers playing distinct roles help direct the student model's learning process. This dual-teacher mechanism helps improve the empirical distribution matching between labeled and unlabeled data, reduces the confirmation bias of pseudo-labels, and facilitates the learning of high-quality features. Additionally, introduce the stochastic intensity-based feature perturbation pool strategy to increase feature diversity in a randomized manner. For DSSCL, we compute patch-based contrast losses at both the deep and shallow (representation) layers of the decoder, which aids in capturing local texture features and edge geometric features, thereby improving the quality of segmentation in low-contrast and local regions.

The experimental results show that with a labeling ratio of 10%, the model trained on the ACDC dataset, the SADT method improves the Dice score by 1.26% compared to the BCP method; and on the ISIC dataset, it improves by 2.55% compared to the UGCL method. These results demonstrate the superior performance of SADT. Additionally, as shown in Fig. 4, we conducted a statistical significance analysis. From the figure, it can be observed that the test result distribution of SADT is better compared to the BCP method. The difference between the highest and lowest values of the test results is smaller, the median values of the Dice and Jaccard metrics are higher than those of the BCP method, and the median values of the 95 HD and ASD metrics are lower than those of the BCP method.

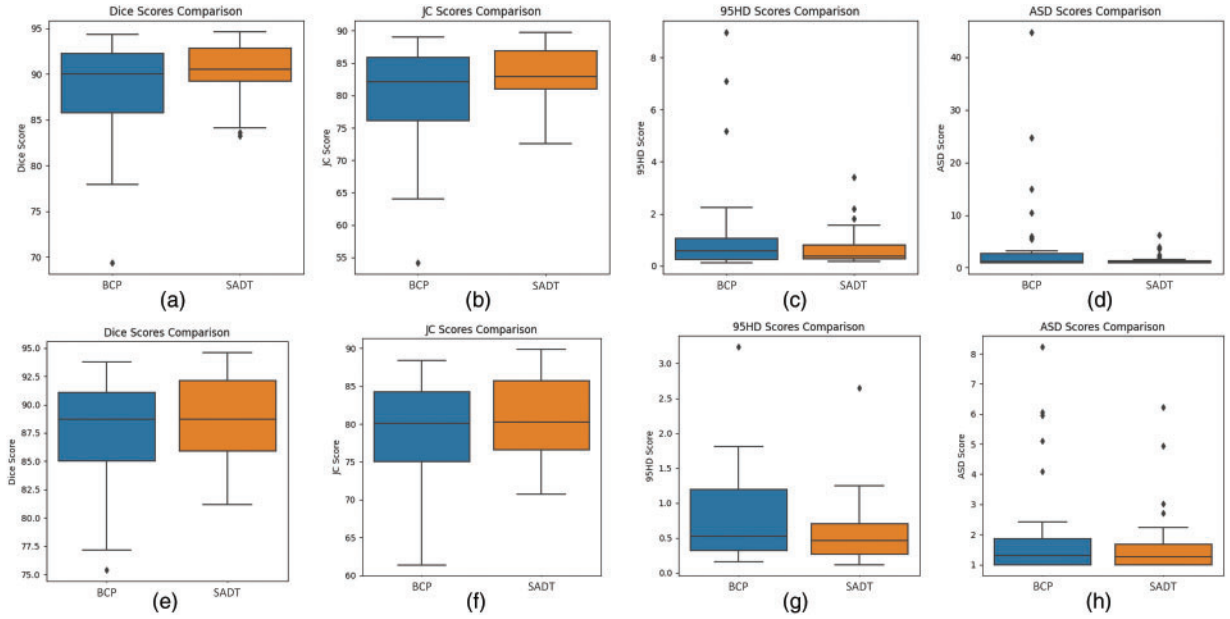

Figure 4: Boxplot comparing the distribution of metric scores for the SADT and BCP methods on the ACDC dataset test results. The labeled (a–d) represent the distribution of test scores after training the model with 10% labeled data. The box plots labeled (e–h) represent the distribution of test scores after training the model with 5% labeled data. The two approaches, BCP and SADT, are represented on the x-axis of each plot, and the evaluation metric score, with units in percentage (%), is represented on the y-axis

Although the proposed SADT method outperforms state-of-the-art semi-supervised learning strategies, there are still some limitations. Due to the use of supervised learning with real labeled data and comparative learning with mixed data in the feature space to guide the update of the student network, additional training costs are required. As shown in Table 11, the statistical t-test significance analysis results indicate that the Dice metric of the model trained with 10% labeled data shows a very significant difference compared to BCP (p-value < 0.05). However, the significance of the differences in Jaccard, 95HD, and ASD metrics weakens, with less significant advantages in outlier edge segmentation compared to BCP. The significance of differences between models trained with 5% labeled data is weaker due to the scarcity of labeled data, resulting in insignificant differences in prediction accuracy in complex and difficult-to-learn feature regions compared to the BCP method. Therefore, we will further enhance the distinctiveness of SADT compared to other advanced methods in semi-supervised medical image segmentation. In future work, we will continue to investigate

the SADT framework, optimize its contrast learning strategy, and effectively apply it to different segmentation network baselines, while controlling the additional training costs.

Table 11: The *p*-values for the statistical significance analysis of the SADT and BCP methods' results on the ACDC dataset tests

Labeling ratios	Dice	Jaccard	95HD	ASD
5%	0.128	0.100	0.174	0.120
10%	**0.049**	0.053	0.056	0.094

6 Conclusion

The proposed SADT strategy enriches data augmentation with partial pseudo-supervision and contrastive learning. The distribution gap between labeled and unlabeled data is reduced by this method. It enhances segmentation capabilities in low-contrast and local regions. It also makes full use of expensive labeled information. Experimental comparisons show that the provided semi-supervised learning approach works better than the most advanced techniques, including BCP. With just 10% of the ACDC dataset's data labeled, the Dice score reached 90.10%. On the ISIC dataset, using only 5% labeled data, the Dice score reached 81.16%, an improvement of 8.49% over UGCL.

Acknowledgement: We would like to thank the editor and the anonymous reviewers for taking the time to review and provide valuable comments.

Funding Statement: This study was supported by the Natural Science Foundation of China (No. 41804112; author: Chengyun Song).

Author Contributions: The authors confirm contribution to the paper as follows: study conception and design: Yang Yuan; data collection: Yang Yuan, Pengcheng Ren; analysis and interpretation of results: Yang Yuan, Hengyang Liu; draft manuscript preparation: Yang Yuan; checking and revising the manuscript preparation: Yang Yuan, Hengyang Liu, Chengyun Song, Fen Luo. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: We evaluate the semi-supervised medical image segmentation model proposed in this paper using two publicly available medical image datasets, i.e., ACDC dataset (https://humanheart-project.creatis.insa-lyon.fr/database/#collection/637218c173e9f0047faa00fb, accessed on 16 June 2023), ISIC dataset (https://challenge.isic-archive.com/data/#2018, accessed on 03 January 2024). All the data generated or analyzed in the course of this study are included in the tables of this paper.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

[1] O. Ronneberger, P. Fischer, and T. Brox, "Unet: Convolutional networks for biomedical image segmentation," in *Med. Image Comput. Comput.-Assist. Interv.-MICCAI 2015: 18th Int. Conf.*, Munich, Germany, Oct. 5–9, 2015, pp. 234–241.

[2] A. Tragakis, C. Kaul, R. Murray-Smith, and D. Husmeier, "The fully convolutional transformer for medical image segmentation," in *Proc. IEEE/CVF Conf. WACV*, Waikoloa, HI, USA, 2023, pp. 3660–3669.

[3] R. Jiao *et al.*, "Learning with limited annotations: A survey on deep semi-supervised learning for medical image segmentation," *Comput. Biol. Med.*, vol. 169, 2023, Art. no. 107840. doi: 10.1016/j.compbiomed.2023.107840.

[4] F. Wu and X. Zhuang, "Minimizing estimated risks on unlabeled data: A new formulation for semi-supervised medical image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 5, pp. 6021–6036, 2023.

[5] Y. Wu, Z. Wu, Q. Wu, Z. Ge, and J. Cai, "Exploring smoothness and class-separation for semi-supervised medical image segmentation," in *Int. Conf. Med. Imag. Comput. Comput. Interv.*, 2022, pp. 34–43. doi: 10.1007/978-3-031-16443-9_4.

[6] B. P. Yap and B. K. Ng, "Cut-paste consistency learning for semi-supervised lesion segmentation," in *WACV*, Waikoloa, HI, USA, 2023, pp. 6149–6158.

[7] A. Tarvainen and H. Valpola, "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results," *Adv. Neural Inf. Process Syst.*, 2017. doi: 10.48550/arXiv.1703.01780.

[8] Y. Chen *et al.*, "Dual-decoder consistency via pseudo-labels guided data augmentation for semi-supervised medical image segmentation," 2023, *arXiv:2308.16573*.

[9] C. You, Y. Zhou, R. Zhao, L. Staib, and J. S. Duncan, "SimCVD: Simple contrastive voxel-wise representation distillation for semi-supervised medical image segmentation," *IEEE Trans. Med. Imaging*, vol. 41, no. 9, pp. 2228–2237, 2022. doi: 10.1109/TMI.2022.3161829.

[10] X. Luo, J. Chen, T. Song, and G. Wang, "Semi-supervised medical image segmentation through dual-task consistency," in *Proc. AAAI Conf. Artif. Intell.*, vol. 35, no. 10, pp. 8801–8809, 2021. doi: 10.1609/aaai.v35i10.17066.

[11] H. Basak and Z. Yin, "Pseudo-label guided contrastive learning for semi-supervised medical image segmentation," in *Proc. IEEE/CVF Conf. CVPR*, Vancouver, BC, Canada, 2023, pp. 19786–19797.

[12] Q. Wang, W. Li, and L. V. Gool, "Semi-supervised learning by augmented distribution alignment," in *Proc. IEEE/CVF Conf. ICCV*, Seoul, Republic of Korea, 2019, pp. 1466–1475.

[13] Y. Bai, D. Chen, Q. Li, W. Shen, and Y. Wang, "Bidirectional copy-paste for semi-supervised medical image segmentation," in *Proc. IEEE/CVF Conf. CVPR*, Vancouver, BC, Canada, 2023, pp. 11514–11524.

[14] O. Bernard *et al.*, "Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: Is the problem solved?" *IEEE Trans. Med. Imaging*, vol. 37, no. 11, pp. 2514–2525, 2018. doi: 10.1109/TMI.2018.2837502.

[15] N. C. F. Codella *et al.*, "Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC)," in *2018 IEEE 15th ISBI 2018*, Brighton, UK, IEEE, 2018, pp. 168–172.

[16] R. Azad *et al.*, "Beyond self-attention: Deformable large kernel attention for medical image segmentation," in *Proc. IEEE/CVF Conf. WACV*, Waikoloa, HI, USA, 2024, pp. 1287–1297.

[17] N. Ibtehaz and M. S. Rahman, "MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation," *Neural Netw.*, vol. 121, no. 11, pp. 74–87, 2019. doi: 10.1016/j.neunet.2019.08.025.

[18] M. M. Rahman and R. Marculescu, "Medical image segmentation via cascaded attention decoding," in *Proc. IEEE/CVF Conf. WACV*, Waikoloa, HI, USA, 2023, pp. 6222–6231.

[19] M. M. Rahman, M. Munir, and R. Marculescu, "EMCAD: Efficient multi-scale convolutional attention decoding for medical image segmentation," in *Proc. IEEE/CVF Conf. CVPR*, Seattle, WA, USA, 2024, pp. 11769–11779.

[20] Y. Gao, "Training like a medical resident: Context-prior learning toward universal medical image segmentation," in *Proc. IEEE/CVF Conf. CVPR*, Seattle, WA, USA, 2024, pp. 11194–11204.
[21] X. Chen, Y. Yuan, G. Zeng, and J. Wang, "Semi-supervised semantic segmentation with cross pseudo supervision," in *Proc. IEEE/CVF Conf. CVPR*, Nashville, TN, USA, 2021, pp. 2613–2622.
[22] X. Wang, B. Zhang, L. Yu, and J. Xiao, "Hunting sparsity: Density-guided contrastive learning for semi-supervised semantic segmentation," in *Proc. IEEE/CVF Conf. CVPR*, Vancouver, BC, Canada, 2023, pp. 3114–3123.
[23] D. Kwon and S. Kwak, "Semi-supervised semantic segmentation with error localization network," in *Proc. IEEE/CVF Conf. CVPR*, New Orleans, LA, USA, 2022, pp. 9957–9967.
[24] A. Peláez-Vegas, P. Mesejo, and J. Luengo, "A survey on semi-supervised semantic segmentation," 2023, *arXiv:2302.09899*.
[25] Y. Liu, Y. Tian, Y. Chen, F. Liu, V. Belagiannis and G. Carneiro, "Perturbed and strict mean teachers for semi-supervised semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recogn. (CVPR)*, New Orleans, LA, USA, 2022, pp. 4248–4257.
[26] Z. Zhao, L. Yang, S. Long, J. Pi, L. Zhou and J. Wang, "Augmentation matters: A simple-yet-effective approach to semi-supervised semantic segmentation," in *Proc. IEEE/CVF Conf. CVPR*, Vancouver, BC, Canada, 2023, pp. 11350–11359.
[27] K. B. Nguyen, "SequenceMatch: Revisiting the design of weak-strong augmentations for Semi-supervised learning," in *Proc. IEEE/CVF WACV*, Waikoloa, HI, USA, 2024, pp. 96–106.
[28] L. Yang, L. Qi, L. Feng, W. Zhang, and Y. Shi, "Revisiting weak-to-strong consistency in semi-supervised semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recogn. (CVPR)*, Vancouver, BC, Canada, 2023, pp. 7236–7246.
[29] X. Zhao *et al.*, "RCPS: Rectified contrastive pseudo supervision for semi-supervised medical image segmentation," *IEEE J. Biomed. Health Inform.*, vol. 28, no. 1, pp. 251–261, 2024. doi: 10.1109/JBHI.2023.3322590.
[30] H. Chi, J. Pang, B. Zhang, and W. Liu, "Adaptive bidirectional displacement for semi-supervised medical image segmentation," in *Proc. IEEE/CVF Conf. CVPR*, Seattle, WA, USA, 2024, pp. 4070–4080.
[31] T. Wang *et al.*, "Uncertainty-guided pixel contrastive learning for semi-supervised medical image segmentation," in *Europ. Conf. Artif. Intell. (ECAI)*, Vienna, Austria, 2022, pp. 1444–1450.
[32] 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 *Med. Image Comput. Comput. Assist. Interv.–MICCAI 2019: 22nd Int. Conf.*, Shenzhen, China, Oct. 13–17, 2019, pp. 605–613. doi: 10.1007/978-3-030-32245-8_67.
[33] S. Li, C. Zhang, and X. He, "Shape-aware semi-supervised 3D semantic segmentation for medical images," in *Med. Image Comput. Comput. Assist. Interv.–MICCAI 2020: 23rd Int. Conf.*, Lima, Peru, Oct. 4–8, 2020, pp. 552–561.
[34] X. Luo *et al.*, "Efficient semi-supervised gross target volume of nasopharyngeal carcinoma segmentation via uncertainty rectified pyramid consistency," in *Med. Image Comput. Comput. Assist. Interv.–MICCAI 2021: 24th Int. Conf.*, Strasbourg, France, Sep. 27–Oct. 1, 2021, pp. 318–329.
[35] Y. Wu, M. Xu, Z. Ge, J. Cai, and L. Zhang, "Semi-supervised left atrium segmentation with mutual consistency training," in *Med. Image Comput. Comput. Assist. Interv.–MICCAI 2021: 24th Int. Conf.*, Strasbourg, France, Sep. 27–Oct. 1, 2021, pp. 297–306.
[36] T. H. Vu, H. Jain, M. Bucher, M. Cord, and P. Pérez, "Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation," in *Proc. IEEE/CVF Conf. CVPR*, Long Beach, CA, USA, 2019, pp. 2517–2526.
[37] S. Qiao, W. Shen, Z. Zhang, B. Wang, and A. Yuille, "Deep co-training for semi-supervised image recognition," in *Proc. ECCV*, Munich, Germany, 2018, pp. 135–152.
[38] Y. Ouali, C. Hudelot, and M. Tami, "Semi-supervised semantic segmentation with cross-consistency training," in *Proc. IEEE/CVF Conf. CVPR*, Seattle, WA, USA, 2020, pp. 12674–12684.