



*Research article*

## **An adversarially consensus model of augmented unlabeled data for cardiac image segmentation (CAU<sup>+</sup>)**

**Wenli Cheng and Jiajia Jiao\***

College of Information Engineering, Shanghai Maritime University, Shanghai, China

\* **Correspondence:** Email: [jiaojiajia@shmtu.edu.cn](mailto:jiaojiajia@shmtu.edu.cn).

**Abstract:** High quality medical images play an important role in intelligent medical analyses. However, the difficulty of acquiring medical images with professional annotation makes the required medical image datasets, very expensive and time-consuming. In this paper, we propose a semi-supervised method, CAU<sup>+</sup>, which is a consensus model of augmented unlabeled data for cardiac image segmentation. First, the whole is divided into two parts: the segmentation network and the discriminator network. The segmentation network is based on the teacher student model. A labeled image is sent to the student model, while an unlabeled image is processed by CTAugment. The strongly augmented samples are sent to the student model and the weakly augmented samples are sent to the teacher model. Second, CAU<sup>+</sup> adopts a hybrid loss function, which mixes the supervised loss for labeled data with the unsupervised loss for unlabeled data. Third, an adversarial learning is introduced to facilitate the semi-supervised learning of unlabeled images by using the confidence map generated by the discriminator as a supervised signal. After evaluating on an automated cardiac diagnosis challenge (ACDC), our proposed method CAU<sup>+</sup> has good effectiveness and generality and CAU<sup>+</sup> is confirmed to have a improves dice coefficient (DSC) by up to 18.01, Jaccard coefficient (JC) by up to 16.72, relative absolute volume difference (RAVD) by up to 0.8, average surface distance (ASD) and 95% Hausdorff distance ( $HD_{95}$ ) reduced by over 50% than the latest semi-supervised learning methods.

**Keywords:** semi-supervised learning; adversarial learning; medical image segmentation; data augmentation; convolutional neural networks

---

## 1. Introduction

According to the World Health Organization (WHO), cardiovascular diseases (CVDs) are the leading cause of death worldwide. 17.9 million people died from CVDs in 2016, and heart disease and strokes are classified as the leading CVDs. This number is increasing every year. Significant advances in cardiovascular research and practices have been made in recent decades, aimed at improving the diagnosis and treatment of heart diseases, as well as reducing CVD mortality. Modern medical imaging techniques, such as magnetic resonance imaging (MRI), computed tomography (CT), and ultrasound are widely used, they allow the non-invasive qualitative and quantitative assessment of cardiac anatomy and function for diagnosis, disease monitoring, treatment planning, and prognosis [1].

It is worth noting that cardiac image segmentation is an important first step in many applications. It segments the image into a number of semantically (i.e., anatomically) meaningful regions, on the basis of which quantitative metrics such as myocardial mass, wall thickness, left ventricle (LV) and right ventricle (RV) volumes, and ejection fraction (EF) can be extracted. Typically, the anatomical structures of interest for cardiac image segmentation includes the left ventricle, right ventricle, left atrium, right atrium, and coronary arteries.

Cardiac MRI (constructed from a series of parallel short-axis slices) is considered the gold standard for the functional analysis of the heart because of its well-known ability to differentiate between different types of tissue [2]. However, there are some difficulties in cardiac MRI segmentation. For example, this includes inherent noise caused by motion artefacts and cardiac dynamics, as well as variations in the shape and intensity of cardiac structures from patient to patient and from condition to condition [3].

Current fully supervised segmentation methods in the field of cardiac image segmentation are mainly based on convolutional neural networks using fully convolutional networks (FCN) [4–7] or U-Net [8–10] architectures. However, one of the major challenges for deep learning methods is the scarcity of annotated data, especially in the field of medical imaging, where data is scarce. Most studies have used a fully supervised approach to train their networks, but this requires many annotated images. In fact, annotating cardiac images is time consuming and requires a lot of expertise.

Based on this, many semi-supervised algorithms have been applied to the field of medical image segmentation. In a semi-supervised learning task, only a small fraction of the training images is assumed to have full pixel-level annotations, while a large number of unlabeled images are available to improve accuracy and generalization. Since unlabeled data does not require labor-intensive annotation, any performance gains from using unlabeled data are low-cost. The challenge in this learning scenario is to use the large amount of unlabeled data effectively and thoroughly.

In this paper, we propose a semi-supervised based method, CAU<sup>+</sup>, for cardiac image segmentation. This method is based on our original proposed method CAU [11]. Two mechanisms are used for upgrading from CAU to CAU<sup>+</sup>. First, inspired by ReMixMatch, we replaced the data augmentation of CAU with CTAugment so that the model can dynamically learn the augmentation strategy during the training process. Second, inspired by AdvSemiSeg [12], adversarial learning is introduced by replacing the generator with a segmentation network (i.e., the teacher-student model in CAU), and a confidence map generated by the discriminator is made to guide the loss function as a supervised signal.

The contributions of this paper include the following:

- (1) We propose a semi-supervised algorithm for cardiac image segmentation, namely “an

adversarially consensus model of augmented unlabeled data (CAU<sup>+</sup>)”, which enables low-cost and the high-precision segmentation of cardiac images.

(2) Our method combines a teacher-student model, and the overall framework is based on a weighted combination of supervised and unsupervised losses from this model. In this way, false identification is avoided and the regularization effect is improved.

(3) We extend the strong and weak augmentation of the data into CTAugment, which using the idea of control theory to dynamically learn the magnitude of each transformation during the training process.

(4) We propose a combination of unsupervised loss to make full use of unlabeled data, i.e., minimizing the difference between network predictions under different data augmentation treatments and using minimized entropy for the output of both networks. Based on this, adversarial learning is introduced to add adversarial loss to unsupervised loss and train discriminators to facilitate semi-supervised learning of unlabeled images by using the confidence map generated by the discriminators as a supervised signal.

(5) We validate CAU<sup>+</sup> on the ACDC dataset, experimentally demonstrating the effectiveness of our method. Experiments shows that CAU<sup>+</sup> has improved over the original CAU in almost all experiments with the same amount of data (up to 1.17 higher DSC, up to 5.64 lower ASD, up to 1.94 lower  $HD_{95}$ , up to 3.24 improvement in JC and up to 0.06 improvement in RAVD). CAU<sup>+</sup> improves DSC by up to 18.01, JC by up to 16.72, RAVD by up to 0.8 and reduces ASD and  $HD_{95}$  reduced by more than 50% than the latest semi-supervised learning methods. It also outperforms a fully supervised algorithm using all labeled data in the ACDC dataset with 35% and 50% labeled data.

## 2. Related works

### 2.1. Semi-supervised learning

Many semi-supervised learning methods provide better generalization of the model by adding a loss term to the unlabeled data. The loss term usually consists of the following:

1) Entropy minimization, which encourages the model to output high confidence predictions on unlabeled data.

2) Consistency regularization, which encourages the model to output the same probability distribution after perturbing the data.

3) Generic regularization, which encourages better generalization and reduces overfitting.

MixMatch [15] achieves good results by combining these methods into one loss. ReMixMatch [16] improves on MixMatch with two components: Distribution Alignment, which distribute the predictions of unlabeled data aligned to labeled data, and Augmentation Anchor, which uses the predictions of the weakly augmented samples as the training target for the strongly augmented version. To generate a strong augmentation, ReMixMatch proposes a variant of AutoAugment, also known as CTAugment, which can learn the augmentation strategy simultaneously during training. FixMatch [17] simplifies MixMatch and ReMixMatch by using the weak enhancement method to obtain a pseudo-label for unlabeled data, and then uses the pseudo-label to monitor the output values of the strong enhancement.

## 2.2. Automated data augmentation based on control theory

Data augmentation is an effective technique to improve the accuracy of modern image classifiers. AutoAugment [18] is a method for learning data augmentation strategies to improve the accuracy of validation sets. The augmentation strategy consists of a set of transformation parameter magnitude tuples to be applied to each image. Crucially, however, AutoAugment is learned under supervision (i.e., the magnitude and order of the transformations are determined by training many models on an agent task). This makes the application of the AutoAugment method problematic for semi-supervised learning on low-labeled semi-supervised learning, especially for medical images with sparsely labeled images. To compensate for the need to train the strategy on labeled data, RandAugment [19] uses uniform random sampling transformations, though this requires tuning the hyperparameters of random sampling on the validation set; however, this is also methodologically difficult when very little labeled data is available.

ReMixMatch introduces a control-theory-based variant of AutoAugment, called CTAugment, which uses ideas from control theory to eliminate the need for augmentation learning in AutoAugment. Unlike AutoAugment, CTAugment learns the augmentation strategy while the model is being trained, making it particularly convenient to set up in semi-supervised learning.

In CTAugment, there is a set of 18 possible transformations and the magnitude values of the transformations are divided into bins, with each bin assigned a different weight. Initially, all bins have a weight of 1. Now two transformations are randomly selected from this set with equal probability to form a sequence of transformations, similar to RandAugment. For each transformation, a magnitude bin is randomly selected based on the normalized bin weights; labeled samples are augmented by these two transformations and fed to the model to predict how close the model predictions are to the actual labels. Then, the bin weights of these transformations are updated. In this way, CTAugment learns to select models that have a higher chance of predicting the correct label and thus augment within the network tolerance.

## 2.3. Adversarial learning for image segmentation

In a game-theoretic sense, generative adversarial networks are based on a game between two machine learning models. The game between two machine learning models in the game theoretic sense is usually implemented using neural networks.

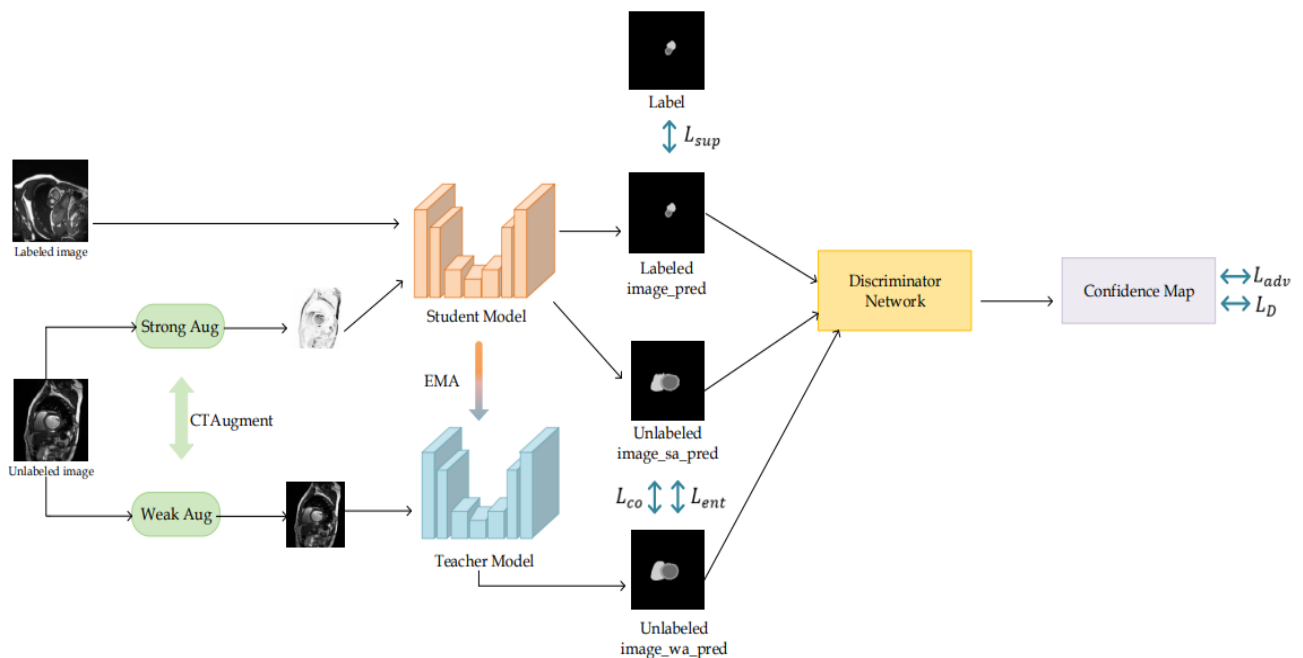
We can think of generative adversarial learning as being a bit like counterfeiters and policeman: counterfeiters create counterfeit currency, while policeman try to arrest counterfeiters and keep legitimate currency in circulation. The competition between the counterfeiters and the police leads to increasingly realistic counterfeits, until the counterfeiters create perfect counterfeits and the police are unable to tell the difference. A complication of this analogy is that the generator learns from the gradient of the discriminator, as if the counterfeiter had planted a mole among the police to report the specific methods the police use to detect counterfeit currency.

Since the framework of generative adversarial network (GAN) and its theoretical foundations were proposed, it has provided ideas for research in many directions in the field of images. In the area of the semi-supervised semantic segmentation of images, several studies have used adversarial methods to make the segmentation of unlabeled images to be more like the segmentation of labeled images [20–23]. Considering the spatial resolution, Hung [12]. proposed a method for semi-supervised

semantic segmentation using adversarial networks to design a discriminator in a fully convolutional manner to distinguish the predicted probability map from the true value segmentation distribution. In the field of medical image segmentation, Xu [25] proposed an adversarial model that allowed for a boundary mining model to learn from additional unlabeled data by evaluating segmentation performance and by providing pseudo-supervision. Zhang [21]. introduced adversarial learning to encourage the segmentation output of unlabeled data to be similar to the annotation of labeled data. Chen [20]. added a discriminator after the segmentation network to distinguish whether the input signed distance map is from either a labeled or an unlabeled image. These methods always include a discriminator to distinguish whether the input image is an annotation from either a labeled image or a prediction from an unlabeled image.

### 3. Proposed method CAU<sup>+</sup>

#### 3.1. Overview of our method CAU<sup>+</sup>



**Figure 1.** General framework of our proposed CAU<sup>+</sup> model. The whole model is divided into two parts: the segmentation network and the discriminator network. The segmentation network, like CAU, is based on the teacher-student model. The unlabeled samples are augmented by CTAugment. Strongly augmented samples are fed into the student model and weakly augmented samples are fed into the teacher model. The labeled samples are trained using  $L_{sup}$ , and the unlabeled samples are trained by  $L_{co}$ ,  $L_{ent}$  and  $L_{adv}$ . The discriminator network contains the discriminator,  $L_D$  trains the discriminator, and  $L_{adv}$  is for adversarial training.

Figure 1 shows the improved CAU<sup>+</sup> model. The whole is divided into two parts: the segmentation network, and the discriminator network. The segmentation network, like CAU, is based on the teacher-

student model: therefore the teacher model and the student model share the same architecture, and in this paper, we use U-Net. The labeled image is sent to the student model, and the unlabeled image is processed by CTAugment. The strongly augmented samples are sent to the student model and the weakly augmented samples are sent to the teacher model. Meanwhile, the similarity measures of the two models' outputs are calculated and the entropy of the two models' outputs is minimized. In the discriminator network, we add the adversarial loss  $L_{adv}$ , which is used to compute the confidence map through the discriminator network, and in turn, the confidence map is used as a supervised signal to guide the segmentation network (i.e., the teacher-student model). We use all prediction data to train the discriminator network and the loss function  $L_D$  is used to train the discriminator network.

### 3.2. Adversarial learning in CAU<sup>+</sup>

Similar to AdvSemiSeg, the model consists of two parts: a segmentation network and a discriminator network. The former can be any network designed for semantic segmentation; in this paper, we use U-Net [26], (i.e., given an input image of size  $H \times W \times 3$ ); the segmentation network outputs a class probability map of size  $H \times W \times C$ , where  $C$  is the number of semantic classes. The framework of the segmentation network is based on the same teacher-student model as the CAU. The discriminator network takes the class probability map as the input, and the class probability map is a spatial probability map of size  $H \times W \times 1$  obtained from either the segmentation network or from the ground truth label after one-hot encoding, and the discriminator outputs each pixel  $P$  of the map, with a pixel value  $P = 1$  indicating from the ground truth label, and a pixel value  $P = 0$  indicating from the segmentation network. When using labeled data, the segmentation network is supervised by  $L_{sup}$ , and for unlabeled data, the loss function adds the adversarial loss  $L_{adv}$  to the CAU. After obtaining the initial segmentation prediction of unlabeled data from the segmentation network, we compute the confidence map by the discriminator network, and in turn use this confidence map as a supervisory signal to guide the segmentation network. AdvSemiSeg trains the discriminator using only labeled data, but due to the sparsity of medical image data, we use the full data to train the discriminator.

### 3.3. Augmentation in CAU<sup>+</sup>

CAU<sup>+</sup> extends the strong augmentation and weak augmentation processing of data in CAU to CTAugment. CTAugment not only compensates for the disadvantage that AutoAugment [18] must be trained on an agent task before it can be used, but also compensates for the disadvantage that RandAugment cannot be trained on the rare cases where there are labeled images. It uses uniform random sampling transforms and is able to perform dynamic inference of the magnitude of each transform during training. Intuitively, CTAugment learns the likelihood that it will produce an image that is classified as correctly labeled. Using these possibilities, CTAugment samples only those enhancements that fall within the tolerance of the network. First, as in AutoAugment, CTAugment divides each parameter of each transform into different deformation magnitudes. Let  $m$  be a vector of bin weights for a certain deformation parameter amplitude. At the beginning of training, all magnitude bins are set with a set of weights initialized to 1.

These weights are used to determine which magnitude bin is applicable to a given image. In each training step, two transforms are sampled uniformly and randomly for each image. To enhance the images for training, CTAugment generates a set of modified bin weights  $\hat{m}$  for each parameter of

these transforms. If  $m_i > 0.8$ ,  $\hat{m}_i = m_i$ , otherwise  $\hat{m}_i = 0$ , and the magnitude bins are drawn from categorical (normalize ( $\hat{m}$ )) [15]. To update the weights of the sampled transforms, CTAugment first samples one magnitude bin  $m_i$  uniformly and randomly for each transform parameter. The resulting transforms will then be applied to the image  $x$  with label  $p$ , resulting in an enhanced version of the image  $\hat{x}$ . Then, according to  $w = 1 - \frac{1}{2L} \sum |p_{model}(y|\hat{x}; \theta) - p|$ , which measures how well the model's predictions match the labels. The weight of the magnitude bin of each sample is subsequently updated to  $m_i = \rho m_i + (1 - \rho)w$ , where  $\rho = 0.99$  is a fixed exponential decay hyperparameter. In this paper, CTAugment is also divided into strong Aug and weak Aug. As mentioned in ReMixMatch, the exponential decay hyperparameter  $\rho$  does not significantly affect the results, but depth and threshold have significant effects on the results, and according to the experiments in ReMixmatch, depth = 2, threshold = 0.8 gives the best results. In this paper, according to the parameter values of depth and threshold provided by ReMixmatch, the value of the threshold has no significant effect on the results, but the value of the depth has a significant effect on the results, and depth = 2 and depth = 1 can produce better results.

### 3.4. Mean teacher based semi-supervised framework

Our segmentation network is based on the mean teacher architecture, which is structured by two identical models, the student model, and the teacher model. The weights of both models are randomly initialized at the beginning of training. The weights of the teacher model are set using an exponentially weighted moving average (EMA) of successive student weights:  $f_{\theta}: \theta_t' = \alpha \theta_{t-1}' + (1 - \alpha) \theta_t$ , where  $\theta_t$  is the parameter of the student model and  $\theta_t'$  is the parameter of the teacher model.  $\alpha$  is a hyperparameter of the smoothing factor to control the coverage of the EMA in the training history. According to the experience of [14], the set  $\alpha = 0.999$  achieves a great performance. Therefore,  $\alpha$  is also set to 0.999 in this paper. Each prediction sample of the teacher model can be considered as an ensemble of the current and previous versions of the student model.

### 3.5. Loss function

CAU<sup>+</sup> consists of a segmentation network and a discriminator network.

The segmentation network is trained by the following loss function:

$$L_S = L_{sup} + \lambda_{co} L_{co} + \lambda_{ent} L_{ent} + \lambda_{adv} L_{adv} \quad (1)$$

where  $L_{sup}$  is the loss of training labeled data, and  $L_{co}$ ,  $L_{ent}$ , and  $L_{adv}$  are the loss of training unlabeled data.  $\lambda_{co}$  is a hyperparameter,  $\lambda_{ent}$  and  $\lambda_{adv}$  are weight factors, which are defined by a time-dependent Gaussian warming up function:  $\lambda(t) = 0.1 \times e^{(-5(1 - \frac{t_i}{t_{total}})^2)}$  [27]. Where  $t_i$  represents the current training iteration and  $t_{total}$  is the total number of iterations.

$L_{sup}$  is defined as:

$$L_{sup} = L_{focal}(p_i, y_i) + L_{dice}(p_i, y_i) \quad (2)$$

$L_{focal}$  is the focal loss and  $L_{dice}$  is the dice loss.  $p_i$  represents the prediction and  $y_i$  represents the label of image  $x_i$ .

$L_{co}$  is defined as:

$$L_{co} = H\left(\frac{1}{2}(p_1(x) + p_2(x))\right) - \frac{1}{2}(H(p_1(x)) + H(p_2(x))) \quad (3)$$

where  $H(P)$  is the entropy of  $P$ . Define the student network as  $p_1(x) = f_1(v_1(x))$ , and the teacher network as  $p_2(x) = f_2(v_2(x))$ .  $v_1(x)$  denotes the strongly augmented data of the input student network, and  $v_2(x)$  denotes the weakly augmented data of the input teacher network. We calculate the Jensen-Shannon divergence of the student network and the teacher network, which is used to make their predictions for the unlabeled data close.

Derived from Shannon entropy [28],  $L_{ent}$  is defined as:

$$L_{ent}(x_t) = \sum E_{x_t}^{(h,w)} \quad (4)$$

where  $x_t$  represents the input image, and  $E_{x_t}$  is an entropy map consisting of independent pixel-level entropies in the normalized range [0,1]. We can encourage the model to make more confident predictions on unlabeled data by entropy minimization [29].

$L_{adv}$  is defined as:

$$L_{adv} = -\sum_{h,w} \log(D(S(X_n))^{(h,w)}) \quad (5)$$

$D(\cdot)$  indicates the discriminator network. With this loss, we train the segmentation network to deceive the discriminator by maximizing the probability of the prediction results that are generated from the ground truth distribution.

The discriminator network is trained by cross entropy loss, it is defined as:

$$L_D = -\sum_{k=1}^n (p_k * \log(q_k)) \quad (6)$$

where  $p_k$  is the expectation of the prediction value and  $q_k$  is the expectation of the true value. The purpose of  $L_D$  is to train a discriminator network for adversarial training. The goal of the discriminator network is to distinguish whether the input is either a ground truth labeled image or a probabilistic map generated by a segmentation network.

#### 4. Experiment and analysis

We experimented CAU and CAU<sup>+</sup> on an ACDC dataset. Our methods are compared with five existing semi-supervised algorithms in the cases of 10, 15, 20, 35, 50% and 1–5% of the labeled data, respectively.

The loss function used in the fully supervised algorithm is the same as those used in semi-supervised algorithms for the labeled data, and the labeled images are randomly selected from the dataset. The base model used in all our experiments is a classical and effective model in the field of medical image segmentation, U-Net. We use the cosine learning rate strategy  $lr = 0.05 \times (1.0 + \cos(\text{iter}_{num} - 1)) \times \frac{\pi}{\text{max\_iterations}}$ . The optimizer uses stochastic gradient descent (SGD) with a learning rate of 0.03. The batch size is 8, and the total number of iterations is 30,000. The training process slices the 3D images (total number of slices is 1562) for 2D segmentation and the predictions are generated slice by slice and stacked into a 3D volume.

The number of slices with labeled data used for different percentages of the experiments is shown in Table 1.



**Table 1.** Percentage of labeled/unlabeled data.

Selected slices/Unselected slices	Percentage
16/1546	1%
34/1528	2%
47/1515	3%
66/1496	4%
88/1474	5%
162/1400	10%
276/1286	15%
312/1250	20%
543/1019	35%
781/781	50%

#### 4.1. Dataset

In this paper, all experiments and comparisons are based on the public benchmark dataset ACDC.

The ACDC dataset was created from real clinical examinations obtained at the University Hospital of Dijon and it has a larger scope than previous cardiac datasets because it includes expert manual segmentation results for the right and left ventricles and myocardial epicardial contours. The 200 MR images with annotated short-axis cardiac images from 100 patients make the ACDC dataset a study material for clinical and algorithmic studies, and the dataset contains the left ventricle (LV), myocardium (Myo), and right ventricle (RV) and their corresponding segmentation masks. Given the large intervals between short-axis slices and the potential for interslice shifts due to respiratory motion, the ACDC dataset is more suitable for 2D segmentation than conventional cardiac images that must be segmented in 3D [30].

We selected 20% of the total dataset as the test set, 10% of the remaining data as the validation set and 90% as the training set. We crop all training images to the same size  $256 \times 256$ , normalize their intensities to the range  $[0,1]$ , and randomly disrupt them before feeding them into the network for training.

#### 4.2. Metric

We use five standard metrics to evaluate the performance of CAU<sup>+</sup>, including DSC, ASD,  $HD_{95}$ , JC and RAVD.

##### DSC

DSC is an ensemble similarity measure function, usually used to calculate the similarity of two samples, and takes the value of  $[0,1]$ . The closer it is to 1, the better the result is. It is defined as:

$$Dice = \frac{2TP}{FP+2TP+FN} \quad (7)$$

Where  $TP$  is true positive,  $FP$  is false positive, and  $FN$  is false negative.

##### ASD

ASD is a measure of the distance between two surfaces. It is defined as the average of a list of distances between each point on one surface and the nearest point on the other surface. It is defined as:

$$ASD = \frac{1}{|S(A)+S(B)|} (\sum_{a \in S(A)} \min_{b \in S(B)} \|a - b\| + \sum_{b \in S(B)} \min_{a \in S(A)} \|b - a\|) \quad (8)$$

Where  $S(A)$  and  $S(B)$  represents the set of surface voxels of  $A$  and  $B$ .  $\sum \min \|\cdot\|$  represents the Euclidean distance of any voxel to  $S(\cdot)$ .

#### HD<sub>95</sub>

The Hausdorff distance is a measure of the distance between two sets of points. It is defined as the maximum distance from one set to the nearest point in another set. 95% Hausdorff is the 95<sup>th</sup> percentile of the ordered distance measure and is more stable for smaller outliers. It is defined as:

$$h(A, B) = \max_{a \in A} (\min_{b \in B} d(a, b)) \quad (9)$$

where  $a$  is the point of set  $A$ ,  $b$  is the point of set  $B$ , and  $d(a, b)$  is the Euclidean distance between  $a$  and  $b$ .

#### JC

JC is used to measure the similarity between finite sample sets and is defined as the size of the intersection set divided by the size of the union set. It is in the range of [0%,100%]. The higher the percentage, the more similar the two sample sets are. It is defined as:

$$J(A, B) = \frac{A \cap B}{A \cup B} \quad (10)$$

where  $A \cap B$  represents the intersection of set  $A$  and set  $B$ , and  $A \cup B$  represents the union of set  $A$  and set  $B$ .

#### RAVD

RAVD is a metric used in medical imaging to evaluate the accuracy of segmentation algorithms. There is no fixed upper or lower limit for its value range. The closer the value is to 0, the closer the segmentation result is to the reference standard. It is defined as:

$$RAVD = \frac{V_r - V_s}{V_r} \quad (11)$$

where  $V_r$  represents the number of voxels in the reference standard and  $V_s$  represents the number of voxels in the segmentation result.

### 4.3. CAU<sup>+</sup>

Because the segmentation network is based on the teacher-student model, the input of the segmentation prediction data in the discriminator is different from that of AdvSemiSeg, except for the input of the ground truth label. We found that the segmentation prediction with labeled data trained only with the adversarial loss function  $L_{adv}$  is poor, which is certain because the amount of data is inherently small. Additionally, we found that it is not good to include both labeled and unlabeled data in the adversarial loss training; therefore, the adversarial loss function  $L_{adv}$  training only trains the segmentation prediction of unlabeled data.

**Table 2.** 10% data volume case of ACDC.

Method	Mean		
	DSC	ASD	$HD_{95}$
<i>CTA_ONLY</i>	88.40	5.53	1.60
<i>DAN_ONLY</i>	88.08	6.52	1.92
<i>CAU<sup>+</sup>_01</i>	88.95	4.90	1.40
<i>CAU<sup>+</sup>_02</i>	88.49	5.16	1.58
<i>CAU<sup>+</sup>_03</i>	86.89	6.03	1.85
<i>CAU<sup>+</sup>_04</i>	86.55	5.63	1.81
<i>CAU<sup>+</sup>_05</i>	87.25	7.99	2.34
<i>CAU<sup>+</sup>_06</i>	83.70	18.00	4.85
<i>CAU</i>	88.05	5.61	1.63

Table 2 is an experiment for the 10% data volume case of the ACDC dataset and serves two purposes:

- 1) To prove the validity of  $CAU^+$ ; and
- 2) Find the effective combination of training  $CAU^+$ .

*CTA\_ONLY* and *DAN\_ONLY* represent experiments with only CTAugment and adversarial training added on top of *CAU*, respectively.

The rest of the  $CAU^+_{0X}$  ( $X = 1,2,3,4,5,6$ ) experiments are the experiments with CTAugment and adversarial training, and the parameters of CTAugment are set to a depth = 2 and a threshold = 0.85. The detail design of  $CAU^+_{0X}$  experiments is shown in Table 3.

**Table 3.** The detail design of  $CAU^+_{0X}$ .

Method	$L_{adv}$ trains weakly augmented data	$L_{adv}$ trains all unlabeled data	Discriminator trains all data	Unlabeled loss in CAU retains $L_{co}$	Unlabeled loss in CAU retains $L_{ent}$	Unlabeled loss in CAU retains $L_{co} + L_{ent}$
<i>CAU<sup>+</sup>_01</i>	✓		✓	✓		
<i>CAU<sup>+</sup>_02</i>	✓		✓			✓
<i>CAU<sup>+</sup>_03</i>	✓		✓		✓	
<i>CAU<sup>+</sup>_04</i>		✓	✓	✓		
<i>CAU<sup>+</sup>_05</i>		✓	✓			✓
<i>CAU<sup>+</sup>_06</i>		✓	✓		✓	

Those marked in red in Table 2 indicate the best value in a metric, and those marked in green indicate the second-best value in a metric.

$CAU^+$  with CTAugment and adversarial training is better than *CAU* alone with CTAugment or adversarial training and is better than the original *CAU*.

According to the table, we can know that adversarial loss  $L_{adv}$  only trains weakly augmented data better than training all unlabeled data. The reason may be because the segmentation network uses the teacher-student model, and the strongly augmented data is into the student model, while the weakly augmented data is into the teacher model. The teacher model is an average of the continuous student

model, in which it theoretically learns more useful and correct semantic information and training the strongly augmented data may mislead the model.

We found little difference between the effects of  $CAU^+_{01}$  and  $CAU^+_{02}$ , so the following experiments were conducted for both combinations.

Table 4 shows the experiments for  $CAU^+_{01}$  and  $CAU^+_{02}$  at 10, 15, 20, 35, and 50% and at the extremes of 1–5% with labeled data volume.

$CAU^+_{0X\sim}$  ( $X = 1, 2$ ) indicates the experiments with the parameter depth = 1, threshold = 0.85 for CTAugment. The ones marked in red in the table indicates the best value in a metric among the four experiments for a certain amount of data.

According to the table, it can be seen that:

1) depth = 1 is better for the case of 10–50% of data volume, i.e., the case with relatively more labeled data. While depth = 2 is better for the extreme case of 1–5% of data volume; and

2)  $L_{CO} + L_{ent} + L_{adv}$  performs better for the extreme case of 1–5% data volume and  $L_{CO} + L_{adv}$  performs better for the case of 10–50% data volume.

**Table 4.** Experiment for  $CAU^+_{01}$  and  $CAU^+_{02}$ .

	Method	$CAU^+_{01}$	$CAU^+_{01\sim}$	$CAU^+_{02}$	$CAU^+_{02\sim}$
1% label	DSC	74.11	71.00	75.45	68.24
	ASD	22.19	26.34	15.02	41.1
	$HD_{95}$	6.21	6.69	4.53	10.98
	RAVD	0.27	-0.82	0.26	-2.41
	JC	60.91	58.77	62.43	55.64
2% label	DSC	83.11	74.52	82.79	78.39
	ASD	7.93	15.34	8.46	13.74
	$HD_{95}$	2.06	3.14	2.18	3.36
	RAVD	0.06	-5.84	0.04	-2.49
	JC	72.41	64.73	72.06	67.59
3% label	DSC	85.80	84.85	85.56	85.81
	ASD	8.92	7.71	9.58	6.25
	$HD_{95}$	2.62	2.11	2.50	1.72
	RAVD	0.11	0.10	0.11	0.09
	JC	76.11	74.75	75.68	76.06
4% label	DSC	86.89	86.12	86.93	86.33
	ASD	6.39	7.17	6.14	6.61
	$HD_{95}$	1.68	1.96	1.55	1.73
	RAVD	0.04	0.06	0.03	0.05
	JC	77.53	76.45	77.64	76.77
5% label	DSC	88.15	88.13	88.62	87.75
	ASD	5.49	5.73	4.90	5.46
	$HD_{95}$	1.60	1.59	1.41	1.61
	RAVD	0.09	0.09	0.08	0.09
	JC	79.66	79.58	80.36	79.05

*Continued on next page*

	<b>Method</b>	$CAU^+_{01}$	$CAU^+_{01\sim}$	$CAU^+_{02}$	$CAU^+_{02\sim}$
10% label	DSC	88.95	89.15	88.49	89.03
	ASD	4.90	4.78	5.16	4.98
	$HD_{95}$	1.40	1.25	1.58	1.38
	RAVD	0.07	0.05	0.08	0.05
	JC	80.77	81.09	80.12	80.90
15% label	DSC	89.54	88.64	88.85	88.70
	ASD	4.42	4.75	4.87	4.83
	$HD_{95}$	1.29	1.60	1.45	1.46
	RAVD	0.08	0.10	0.10	0.08
	JC	81.83	80.52	80.79	80.64
20% label	DSC	89.78	90.15	89.41	90.09
	ASD	4.43	3.96	4.30	4.09
	$HD_{95}$	1.01	0.85	1.01	1.16
	RAVD	0.01	0.04	0.02	0.04
	JC	82.00	83.01	81.45	82.59
35% label	DSC	90.00	91.07	90.45	91.14
	ASD	4.48	4.73	3.96	3.34
	$HD_{95}$	1.04	1.18	0.88	0.83
	RAVD	0.05	0.04	0.03	0.04
	JC	82.44	84.20	83.13	84.21
50% label	DSC	90.14	90.76	90.43	91.13
	ASD	4.25	3.90	4.00	3.42
	$HD_{95}$	1.16	1.19	0.91	0.97
	RAVD	0.08	0.03	0.03	0.04
	JC	82.77	83.08	83.08	84.29

## 5. Results

In Table 5, we recorded the total training time and the average time per epoch for each CAU,  $CAU^+$ , and other semi-supervised methods for 50% of the data. These are compared with the total time and average epoch time of the fully supervised method using all data.

Due to the time required for CTAugment to learn the augmentation strategy, the total time for  $CAU^+$  is relatively longer.

Our methods take slightly more time than the other semi-supervised and fully supervised algorithms in terms of total time and average epoch time, but because the trained models are offline, they are ready to use after training. Therefore, the small increase in time is worth the increase in accuracy.

The results of our method  $CAU^+$  on the ACDC dataset with 1–5%, 10, 15, 20, 35 and 50% data volumes are listed in Table 6. The ones marked in red indicate the best value of a metric in a certain data volume, and the ones marked in green indicate the second-best value.  $CAU^+$  has some improvements compared to CAU for most different data volumes, and performs much better than fully supervised and other semi-supervised algorithms with the same volume of data for extreme data amounts from 1–5%. Moreover, Dice, ASD,  $HD_{95}$ , JC, and RAVD indices perform better than fully

supervised algorithm using all data for 35% and 50% of data amounts.

**Table 5.** Total training time and the average time per epoch.

Method	Total Time	Average time per epoch
100% Fully	245 m 41 s	87.42
CAU	267 m 20 s	81.47
CAU <sup>+</sup>	312 m 3 s	56.71
Pseudo	234 m 59 s	82.85
EM	239 m 40 s	47.87
MT	240 m 14 s	46.48
DAN	267 m 42 s	50.50
FixMatch	276 m 29 s	55.10

**Table 6.** Result of CAU<sup>+</sup> on ACDC.

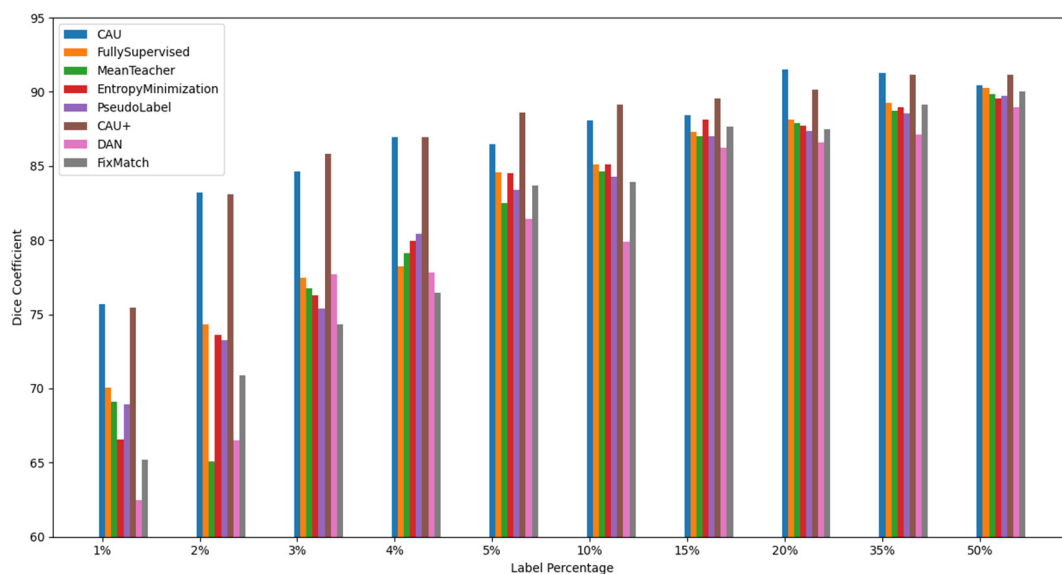
	Method	Pseudo	EM	MT	DAN	FixMatch	Fully	CAU	CAU <sup>+</sup>	100%Fully
1% label	DSC	68.93	66.57	69.11	62.48	65.18	70.05	75.71	75.45	90.73
	ASD	30.83	35.91	40.53	56.65	39.36	42.17	15.68	15.02	3.55
	HD <sub>95</sub>	8.52	10.83	11.48	15.96	12.44	12.39	3.81	4.53	0.97
	RAVD	-0.08	-0.88	0.49	0.32	-0.07	0.16	-0.04	0.26	0.06
	JC	55.66	53.27	49.31	47.21	50.99	56.58	63.00	62.43	83.55
2% label	DSC	73.25	73.60	65.10	66.52	70.87	74.30	83.21	83.11	90.73
	ASD	16.66	21.32	43.46	39.46	20.54	16.11	11.60	7.93	3.55
	HD <sub>95</sub>	5.16	5.60	12.20	12.17	6.40	3.92	3.31	2.06	0.97
	RAVD	-0.02	-0.03	0.23	0.16	-0.03	-0.10	0.13	0.06	0.06
	JC	61.09	61.00	63.99	61.88	58.20	61.83	72.31	72.41	83.55
3% label	DSC	75.38	76.26	76.76	77.72	74.30	77.48	84.64	85.81	90.73
	ASD	21.27	20.75	27.46	24.65	18.69	19.63	8.59	6.25	3.55
	HD <sub>95</sub>	5.75	5.97	7.59	5.85	5.80	5.58	2.31	1.72	0.97
	RAVD	-0.17	0.03	0.07	0.13	0.07	-0.13	0.11	0.09	0.06
	JC	59.34	64.43	64.36	65.53	61.94	63.49	74.32	76.06	83.55
4% label	DSC	80.44	79.96	79.11	77.83	76.47	78.23	86.92	86.93	90.73
	ASD	12.45	16.71	19.76	19.84	13.18	19.00	6.95	6.14	3.55
	HD <sub>95</sub>	3.45	3.89	6.32	5.37	3.53	5.25	2.11	1.55	0.97
	RAVD	-0.05	-0.04	0.05	-0.08	-0.05	-0.83	0.03	0.03	0.06
	JC	67.64	68.92	65.00	62.98	64.88	66.65	77.45	77.64	83.55
5% label	DSC	83.37	84.51	82.49	81.42	83.70	84.57	86.44	88.62	90.73
	ASD	13.96	11.32	13.92	16.45	14.77	11.54	10.54	4.90	3.55
	HD <sub>95</sub>	3.51	2.68	3.89	4.76	4.11	3.30	3.29	1.41	0.97
	RAVD	0.11	0.06	0.07	0.09	0.10	0.04	0.13	0.08	0.06
	JC	72.77	74.21	71.52	70.81	73.15	74.34	77.12	80.36	83.55
10% label	DSC	84.26	85.13	84.61	79.87	83.90	85.11	88.05	89.15	90.73
	ASD	8.55	9.38	7.56	14.69	11.67	7.00	5.61	4.78	3.55

*Continued on next page*

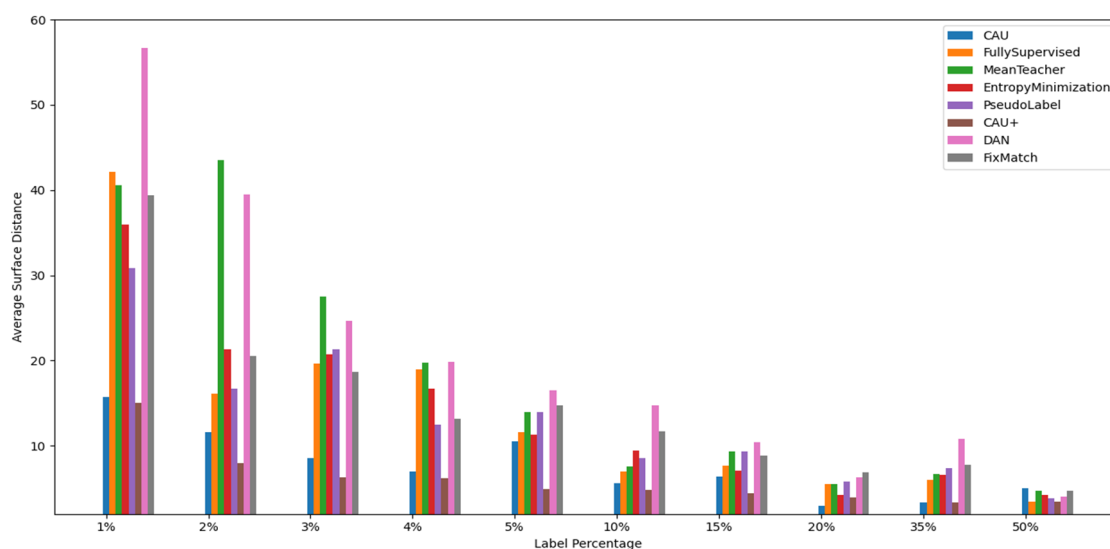
	<b>Method</b>	Pseudo	EM	MT	DAN	FixMatch	Fully	CAU	CAU <sup>+</sup>	100%Fully
15% label	<i>HD</i> <sub>95</sub>	2.74	2.35	2.24	4.00	3.54	2.17	1.63	1.25	0.97
	RAVD	0.07	0.05	0.06	0.02	0.12	0.07	0.11	0.05	0.06
	JC	73.91	75.27	74.60	68.36	73.27	75.28	79.34	81.09	83.55
	DSC	87.00	88.11	87.02	86.26	87.64	87.29	88.43	89.54	90.73
	ASD	9.29	7.10	9.37	10.40	8.80	7.62	6.36	4.42	3.55
	<i>HD</i> <sub>95</sub>	2.60	1.81	2.97	3.20	2.36	2.15	2.10	1.29	0.97
20% label	RAVD	0.10	0.06	0.12	0.09	0.06	0.07	0.11	0.08	0.06
	JC	77.89	79.41	78.05	76.76	78.73	78.30	80.21	81.83	83.55
	DSC	87.34	87.72	87.87	86.60	87.45	88.10	91.53	90.15	90.73
	ASD	5.81	4.20	5.54	6.24	6.91	5.48	2.99	3.96	3.55
	<i>HD</i> <sub>95</sub>	1.68	1.25	1.60	1.85	2.10	1.58	0.87	0.85	0.97
	RAVD	0.07	0.01	0.09	0.08	0.07	0.06	0.01	0.04	0.06
35% label	JC	78.29	79.01	79.16	77.18	78.50	79.48	84.68	83.01	83.55
	DSC	88.52	88.97	88.73	87.15	89.14	89.24	91.27	91.14	90.73
	ASD	7.34	6.54	6.65	10.79	7.79	5.97	3.30	3.34	3.55
	<i>HD</i> <sub>95</sub>	2.21	1.76	2.01	2.70	2.28	1.61	0.85	0.83	0.97
	RAVD	0.10	0.05	0.08	0.08	0.06	0.06	0.05	0.04	0.06
	JC	80.23	80.79	80.38	77.99	81.09	81.23	84.33	84.21	83.55
50% label	DSC	89.70	89.55	89.82	88.95	90.00	90.26	90.45	91.13	90.73
	ASD	3.79	4.20	4.76	4.03	4.75	3.48	5.00	3.42	3.55
	<i>HD</i> <sub>95</sub>	0.99	1.25	1.22	1.28	1.37	0.98	1.35	0.97	0.97
	RAVD	0.04	0.11	0.07	0.09	0.06	0.06	0.06	0.04	0.06
	JC	81.87	81.75	82.08	80.76	82.38	82.83	83.11	84.29	83.55

As shown in Figure 2, the bar chart demonstrates the Dice metrics for different algorithms with a different number of labeled images on the ACDC dataset. The blue color is the original CAU and the brown color is the CAU<sup>+</sup> with the improvements to it proposed in this paper. As can be seen, our method CAU<sup>+</sup> performs particularly well at the extremes of data volume (i.e., 1–5%). CAU<sup>+</sup> improves compared to CAU in almost all cases with different data volumes. Moreover, the adversarial learning method DAN has lower indices than others in almost all cases. This indicates that adversarial learning alone does not work well in cardiac image segmentation, demonstrating the effectiveness of CAU<sup>+</sup> in introducing adversarial learning into the original CAU.

As Figure 3 shows, the bar chart presents the ASD metrics for different algorithms on the ACDC dataset with different numbers of labeled images. The blue color is the original CAU and the brown color is the CAU<sup>+</sup> with its improvement proposed in this paper. It can be seen that CAU<sup>+</sup> has lower indices than the fully supervised algorithm and other semi-supervised algorithms that use the same volume of data for all different data sizes; the indices are improved compared to CAU for almost all different data sizes, in particular, CAU<sup>+</sup> improves the problem that the original CAU has higher indices than the fully supervised algorithm that use the same volume of data for 50% of the time. Moreover, the adversarial learning method DAN and mean teacher outperform the fully supervised algorithm with the same volume of data in almost all cases, demonstrating that neither outperform the fully supervised algorithm with the same volume of data in almost all cases, demonstrating that neither method alone works well for cardiac image segmentation and validating the effectiveness of CAU<sup>+</sup>.



**Figure 2.** Dice on ACDC.



**Figure 3.** ASD on ACDC.

As is shown in Figure 4, the bar chart presents the  $HD_{95}$  metrics for different algorithms on the ACDC dataset with a different number of labeled images. The blue color is the original  $CAU$  and the brown color is the  $CAU^+$  with its improvement proposed in this paper. It can be seen that  $CAU^+$  has a lower index than the fully supervised algorithm and other existing semi-supervised algorithms that use the same volume of data for all different data volumes; this has a larger improvement compared to  $CAU$  for 2–15% of the data volume, especially  $CAU^+$  improves the problem that  $CAU$  exceeds the fully supervised algorithm with the same volume of data in 50% of the cases. Moreover, in the extreme case (i.e., 1–5%), almost all other existing semi-supervised algorithms outperform the fully supervised algorithm using the same volume of data, while  $CAU^+$  performs better than all algorithms, verifying the effectiveness of  $CAU^+$  for cardiac image segmentation in the case of extremely small amount of data.



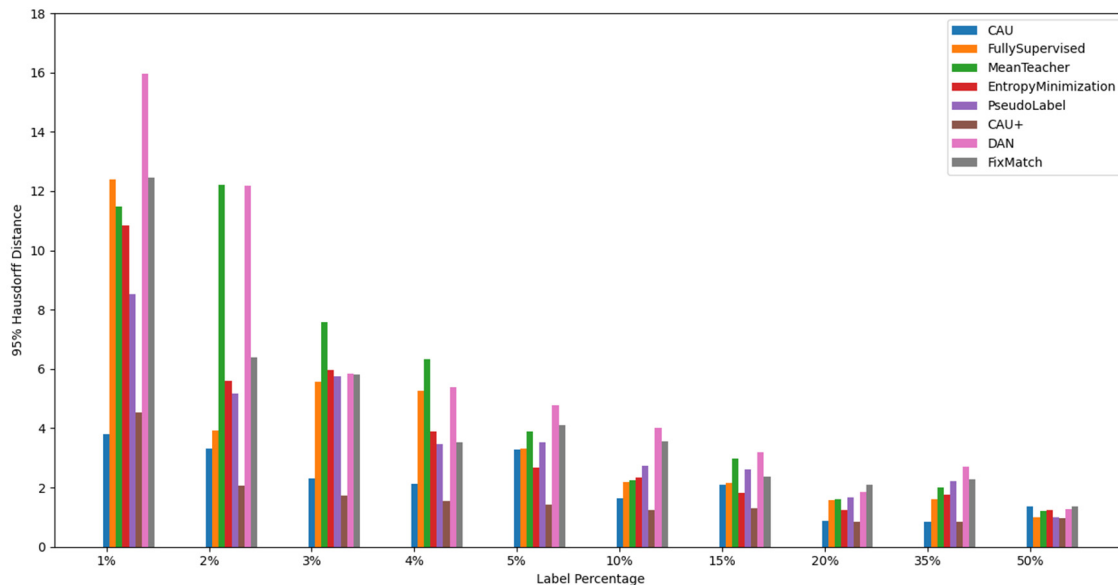


Figure 4.  $HD_{95}$  on ACDC.

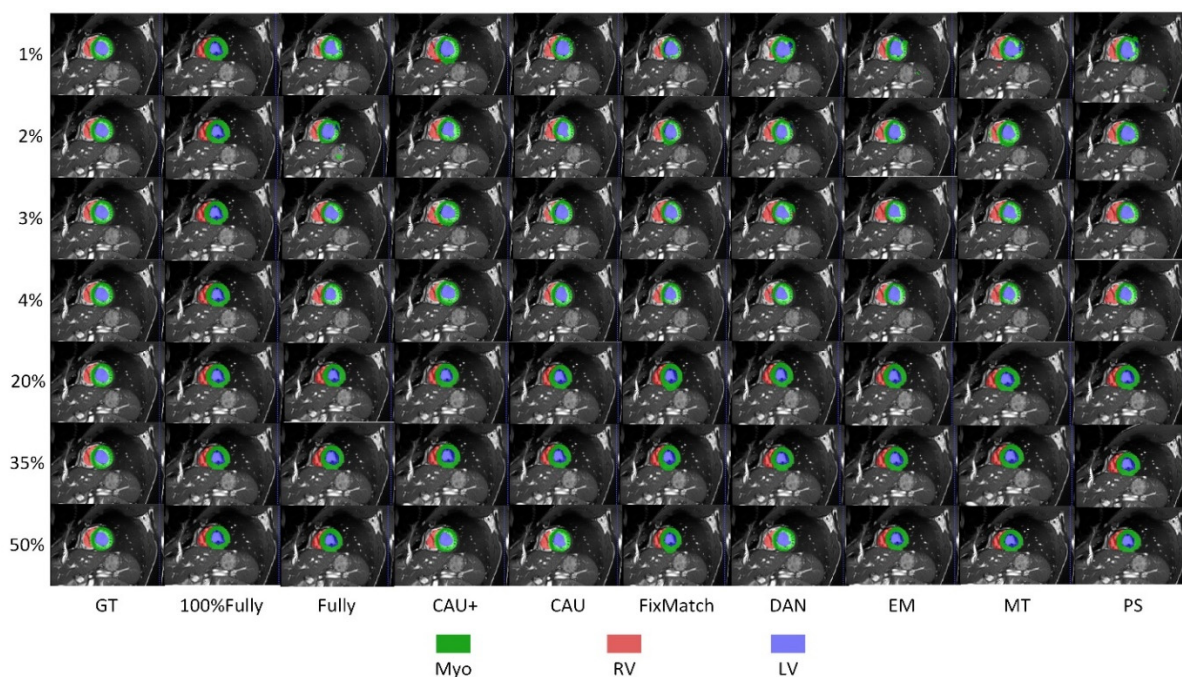


Figure 5. Comparison of the segmentation effect on ACDC.

As shown in Figure 5, the segmentation results generated by different methods on the ACDC dataset are plotted. The first column is the ground truth, the second column is fully supervised that use all labeled images, the third column is fully supervised with various data volumes, and the fourth column is our improved algorithm for  $CAU$  (i.e.,  $CAU^+$ , the fifth, sixth, seventh, eighth, ninth and tenth columns are  $CAU$ , FixMatch, adversarial learning, entropy minimization, mean teacher and pseudo label [13], respectively). The segmented areas, colored blue, green and red in the Figure 5, are the left ventricle, myocardium, and right ventricle. It can be seen that  $CAU^+$  and  $CAU$  are closer to

the ground truth in the extreme case of data volume (1–4%) than fully supervised and other semi-supervised methods with the same data volume, and the observations also show the effectiveness of our method compared to fully supervised methods using all fully labeled images.

In fact, our method can also be used to segment images from other imaging techniques, such as optical coherence tomography (OCT) [31–34]. OCT is a non-invasive imaging technique that provides structural and functional imaging of retina with high spatial and temporal resolution. In the future, we will try to apply our methods to the segmentation of images from other imaging techniques, which we believe will lead to further advances in medical image segmentation.

## 6. Conclusions

In this paper, we propose a semi-supervised training method CAU<sup>+</sup> which is suitable for cardiac image segmentation. The whole is divided into two parts: the segmentation network and the discriminator network. The segmentation network is based on the teacher student model. Labeled image is sent to the student model. Unlabeled image is processed by CTAugment. Then, the strongly augmented samples are sent to the student model and the weakly augmented samples are sent to the teacher model. The loss function adopts a hybrid loss function, which mixes the supervised loss for labeled data with unsupervised loss for unlabeled data. Adversarial learning is also introduced to facilitate semi-supervised learning of unlabeled data through confidence maps generated by discriminators. We validate CAU and CAU<sup>+</sup> on the ACDC dataset. Experiments show that CAU<sup>+</sup> has improved over the original CAU in most experiments with the same amount of data (up to 1.17 higher DSC, up to 5.64 lower ASD, up to 1.94 lower  $HD_{95}$ , up to 3.24 improvement in JC and up to 0.06 improvement in RAVD). And CAU<sup>+</sup> improves DSC by up to 18.01, JC by up to 16.72, RAVD by up to 0.8, ASD and  $HD_{95}$  reduced by more than 50% than the latest semi-supervised learning methods. It also outperforms fully supervised algorithm using all labeled data with 35% and 50% labeled data.

## Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

## Acknowledgments

The paper is sponsored by Shanghai Pujiang Program with grant number 21PJD026.

## Conflict of interest

The authors declare there is no conflict of interest.

## References

1. C. Chen, C. Qin, H. Qiu, G. Tarroni, J. Duan, W. Bai, et al., Deep learning for cardiac image segmentation: A review, *Front. Cardiovasc. Med.*, **7** (2020), 25. <https://doi.org/10.3389/fcvm.2020.00025>

2. C. A. Miller, P. Jordan, A. Borg, R. Argyle, D. Clark, K. Pearce, et al., Quantification of left ventricular indices from SSFP cine imaging: Impact of real-world variability in analysis methodology and utility of geometric modeling, *J. Magn. Reson. Imag.*, **37** (2013), 1213–1222. <https://doi.org/10.1002/jmri.23892>
3. S. Queirós, D. Barbosa, B. Heyde, P. Morais, J. L. Vilaça, D. Friboulet, et al., Fast automatic myocardial segmentation in 4D cine CMR datasets, *Med. Image Anal.*, **18** (2014), 1115–1131. <https://doi.org/10.1016/j.media.2014.06.001>
4. D. H. N. Nham, M. N. Trinh, T. T. Tran, V. T. Pham, T. T. Tran, A modified FCN-based method for Left Ventricle endocardium and epicardium segmentation with new block modules, in *2021 8th NAFOSTED Conference on Information and Computer Science (NICS)*, (2021), 392–397. <https://doi.org/10.1109/NICS54270.2021.9701571>
5. Z. F. Shaaf, M. M. A. Jamil, R. Ambar, A. A. Alattab, A. A. Yahya, Y. Asiri, Automatic left ventricle segmentation from short-axis cardiac MRI images based on fully convolutional neural network, *Diagnostics*, **12** (2022), 414. <https://doi.org/10.3390/diagnostics12020414>
6. P. Daudé, P. Ancel, S. C. Gouny, A. Jacquier, F. Kober, A. Dutour, et al., Deep-learning segmentation of epicardial adipose tissue using four-chamber cardiac magnetic resonance imaging, *Diagnostics*, **12** (2022), 126. <https://doi.org/10.3390/diagnostics12010126>
7. Z. F. Shaaf, M. M. A. Jamil, R. Ambar, A. A. Alattab, A. A. Yahya, Y. Asiri, Automatic left ventricle segmentation from short-axis cardiac MRI images based on fully convolutional neural network, *Diagnostics*, **12** (2022), 414. <https://doi.org/10.3390/diagnostics12020414>
8. Z. Fu, J. Zhang, R. Luo, Y. Sun, D. Deng, L. Xia. TF-Unet: An automatic cardiac MRI image segmentation method, *Math. Biosci. Eng.*, **19** (2022), 5207–5222. <https://doi.org/10.3934/mbe.2022244>
9. D. Abdelraouf, M. Essam, M. Elattar, Light-weight localization and scale-independent multi-gate UNET segmentation of left and right ventricles in MRI images, *Cardiovasc. Eng. Tech.*, **13** (2022), 393–406. <https://doi.org/10.1007/s13239-021-00591-2>
10. Z. Liu, X. He, Y. Lu, Combining UNet 3+ and transformer for left ventricle segmentation via signed distance and focal loss, *Appl. Sci.*, **12** (2022), 9208. <https://doi.org/10.3390/app12189208>
11. W. Cheng, J. Jiao, CAU: A consensus model of augmented unlabeled data for medical image segmentation, in *2022 7th International Conference on Image, Vision and Computing (ICIVC)*, (2022), 368–374. <https://doi.org/10.1109/ICIVC55077.2022.9886218>
12. W. Hung, Y. Tsai, Y. Liou, Y. Lin, M. Yang, Adversarial learning for semi-supervised semantic segmentation, preprint, arXiv:1802.07934. <https://doi.org/10.48550/arXiv.1802.07934>
13. D. H. Lee, Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks, in *ICML 2013 Workshop: Challenges in Representation Learning (WREPL)*, 2013.
14. A. Tarvainen, H. Valpola, Mean teachers are better role models: Weightaveraged consistency targets improve semi-supervised deep learning results, *Adv. Neural Inf. Process. Syst.*, **30** (2017).
15. D. Berthelot, N. Carlini, I. Goodfellow, N. Papernot, A. Oliver, C. Raffel, MixMatch: A holistic approach to semi-supervised learning, *Adv. Neural Inf. Process. Syst.*, **32** (2019).
16. D. Berthelot, N. Carlini, E. D. Cubuk, A. Kurakin, K. Sohn, H. Zhang, et al., ReMixMatch: Semi-supervised learning with distribution alignment and augmentation anchoring, preprint, arXiv:1911.09785. <https://doi.org/10.48550/arXiv.1911.09785>

17. K. Sohn, D. Berthelot, C. Li, Z. Zhang, N. Carlini, E. D. Cubuk, et al., FixMatch: Simplifying semi-supervised learning with consistency and confidence, *Adv. Neural Inf. Process. Syst.*, **33** (2020), 596–608.
18. E. D. Cubuk, B. Zoph, D. Mane, V. Vasudevan, Q. V. Le, AutoAugment: Learning augmentation policies from data, preprint, arXiv:1805.09501. <https://doi.org/10.48550/arXiv.1805.09501>
19. E. D. Cubuk, B. Zoph, J. Shlens, Q. V. Le, Randaugment: Practical automated data augmentation with a reduced search space, in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, (2020), 702–703. <https://doi.org/10.1109/CVPRW50498.2020.00359>
20. G. Chen, J. Ru, Y. Zhou, I. Rekik, Z. Pan, X. Liu, et al., Mtans: Multi-scale mean teacher combined adversarial network with shape-aware embedding for semi-supervised brain lesion segmentation, *NeuroImage*, **244** (2021), 118568. <https://doi.org/10.1016/j.neuroimage.2021.118568>
21. Y. Zhang, L. Yang, J. Chen, M. Fredericksen, D. P. Hughes, D. Z. Chen, Deep adversarial networks for biomedical image segmentation utilizing unannotated images, in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, (2017), 408–416. [https://doi.org/10.1007/978-3-319-66179-7\\_47](https://doi.org/10.1007/978-3-319-66179-7_47)
22. D. Zhai, B. Hu, X. Gong, H. Zou, J. Luo, ASS-GAN: Asymmetric semi-supervised GAN for breast ultrasound image segmentation, *Neurocomputing*, **493** (2022), 204–216. <https://doi.org/10.1016/j.neucom.2022.04.021>
23. K. Shen, H. Quan, J. Han, M. Wu, URO-GAN: An untrustworthy region optimization approach for adipose tissue segmentation based on adversarial learning, *Appl. Intell.*, **52** (2022), 10247–10269. <https://doi.org/10.1007/s10489-021-02976-1>
24. C. Xu, Y. Wang, D. Zhang, L. Han, Y. Zhang, J. Chen, et al., BMAnet: Boundary mining with adversarial learning for semi-supervised 2D myocardial infarction segmentation, *IEEE J. Biomed. Health Inf.*, **27** (2023), 87–96. <https://doi.org/10.1109/JBHI.2022.3215536>
25. O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*, (2015), 234–241. [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)
26. X. Luo, J. Chen, T. Song, G. Wang, Semi-supervised medical image segmentation through dual-task consistency, in *Proceedings of the AAAI Conference on Artificial Intelligence*, **35** (2021), 8801–8809. <https://doi.org/10.1609/aaai.v35i10.17066>
27. C. E. Shannon, A mathematical theory of communication, *SIGMOBILE Mob. Comput. Commun. Rev.*, **5** (2001), 3–55. <https://doi.org/10.1145/584091.584093>
28. Y. Grandvalet, Y. Bengio, Semi-supervised learning by entropy minimization, *Adv. Neural Inf. Process. Syst.*, **2004** (2004), 17.
29. W. Bai, O. Oktay, M. Sinclair, H. Suzuki, M. Rajchl, G. Tarroni, et al., Semi-supervised learning for network-based cardiac mr image segmentation, in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2017*, (2017), 253–260. [https://doi.org/10.1007/978-3-319-66185-8\\_29](https://doi.org/10.1007/978-3-319-66185-8_29)
30. R. K. Meleppat, M. V. Matham, L. K. Seah, Optical frequency domain imaging with a rapidly swept laser in the 1300nm bio-imaging window, in *International Conference on Optical and Photonic Engineering (icOPEN 2015)*, (2015), 721–729. <https://doi.org/10.1117/12.2190530>

31. K. M. Ratheesh, L. K. Seah, V. M. Murukeshan, Spectral phase-based automatic calibration scheme for swept source-based optical coherence tomography systems, *Phys. Med. Biol.*, **61** (2016), 7652. <https://doi.org/10.1088/0031-9155/61/21/7652>
32. R. K. Meleppat, M. V. Matham, L. K. Seah, An efficient phase analysis-based wavenumber linearization scheme for swept source optical coherence tomography systems, *Laser Phys. Lett.*, **12** (2015), 055601. <https://doi.org/10.1088/1612-2011/12/5/055601>
33. R. K. Meleppat, P. Prabhathan, S. L. Keey, M. V. Matham, Plasmon resonant silica-coated silver nanoplates as contrast agents for optical coherence tomography, **12** (2016), 1929–1937. <https://doi.org/10.1166/jbn.2016.2297>



AIMS Press

©2023 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)