

Adaptive Bidirectional Displacement for Semi-Supervised Medical Image Segmentation

Hanyang Chi¹ Jian Pang¹ Bingfeng Zhang^{1*} Weifeng Liu¹

¹China University of Petroleum (East China)

{chihanyang, jianpang}@s.upc.edu.cn, {bingfeng.zhang, liuwf}@upc.edu.cn

Abstract

Consistency learning is a central strategy to tackle unlabeled data in semi-supervised medical image segmentation (SSMIS), which enforces the model to produce consistent predictions under the perturbation. However, most current approaches solely focus on utilizing a specific single perturbation, which can only cope with limited cases, while employing multiple perturbations simultaneously is hard to guarantee the quality of consistency learning. In this paper, we propose an Adaptive Bidirectional Displacement (ABD) approach to solve the above challenge. Specifically, we first design a bidirectional patch displacement based on reliable prediction confidence for unlabeled data to generate new samples, which can effectively suppress uncontrollable regions and still retain the influence of input perturbations. Meanwhile, to enforce the model to learn the potentially uncontrollable content, a bidirectional displacement operation with inverse confidence is proposed for the labeled images, which generates samples with more unreliable information to facilitate model learning. Extensive experiments show that ABD achieves new state-of-the-art performances for SSMIS, significantly improving different baselines. Source code is available at <https://github.com/chy-upc/ABD>.

1. Introduction

Medical image segmentation derives from computer tomography (CT) or magnetic resonance imaging (MRI), which is crucial for various clinical applications [34, 45]. Obtaining a large medical dataset with precise annotation to train segmentation models is challenging, as reliable annotations can only be provided by experts, which constrains the development of medical image segmentation algorithms and poses substantial challenges for further research and implementation [31, 43]. To mitigate the burden of manual annotation and address these challenges, semi-supervised medical im-

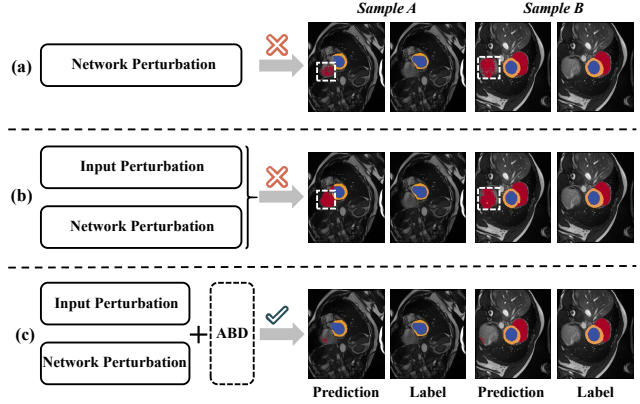


Figure 1. Illustration of prediction results. (a) Using only network perturbation; (b) Combining network perturbation with input perturbation; (c) Incorporating ABD on top of the two perturbations. Using a single perturbation has limitations while using multiple perturbations makes it uncontrollable. Introducing ABD greatly alleviates the issue and allows the model to perform significantly better. The white dashed boxes highlight the regions with wrong predictions.

age segmentation (SSMIS) [2, 28, 35, 42] is emerging as a practical approach to encourage segmentation models to learn from readily available unlabelled data in conjunction with limited labeled examples.

Most recent approaches in SSMIS employ consistency learning [9, 11, 15, 44] to make the decision boundary of the learned model located within the low-density boundary [12]. By ensuring consistent features or predictions under diverse perturbations, this strategy becomes one of the most effective solutions for learning from unlabelled data. According to the perturbation differences, consistency learning approaches can be divided into three categories: 1) Input perturbations, which mainly produce different inputs to the same model [5, 29, 41], *e.g.*, weak and strong data augmentation for the given image. 2) Feature perturbations [17, 25, 47], which mainly include feature noise, feature dropout, and context masking. 3) Network perturbations [8, 13], which focus on using different network archi-

*Corresponding author.

tectures for the same input. To ensure the stability of consistency learning, most previous approaches [16, 32, 38, 51] solely utilize one of the above perturbations, restricting the performance of the consistency learning and leading to imprecise decision boundary since the specific single perturbation can only handle limited cases, as shown in Fig. 1(a).

Utilizing mixed or multiple perturbations presents a direct solution to solve the above problem. However, once added multiple perturbations, the consistency learning process is easily out-of-control, leading to restricted learning quality. For example, Cross Pseudo Supervision (CPS) [8] is a widely-used technique in consistency learning that primarily applies network perturbation to produce two discriminate predictions for further consistency learning. If we directly add input perturbation by using weak and strong augmentation to the input training data, consistency learning will be ineffective. As shown in Fig. 1(b), when the original input is replaced with weak and strong augmentation inputs, the CPS model incorrectly classifies the background as the foreground with consistency learning mechanism, indicating decreased performance when mixed perturbations are introduced.

To tackle the aforementioned challenges, we propose **Adaptive Bidirectional Displacement (ABD)** for SSMIS, as shown in Fig. 1(c). Specifically, for each unlabelled image, the input perturbation is firstly applied to produce a pair of input images, *e.g.*, weak augmentation image and strong augmentation image. Then we generate two confidence matrices (confidence rank maps) based on the predictions from the above two perturbed images. The confidence matrix assesses the certainty of predicted pixels belonging to various categories, thereby reflecting the model’s reliability to different perturbations. Then, for any augmented image, its region with the lowest confidence rank is displaced with the region from the other augmented image that has the most similar output distribution with highly confident scores. We refer to this as an adaptive bidirectional displacement with reliable confidence (ABD-R). In this way, the newly generated image can remove the uncontrolled region and obtain complementary and approximate semantic information from another augmented image, which ensures consistent predictions of the model across different perturbations. Meanwhile, to enforce the model to learn those potentially uncontrollable regions, we incorporate inverse confidence for labeled data as an additional adaptive bidirectional displacement (ABD-I). For any labeled augmented image, the image regions with the highest confidence scores are displaced with the regions from another augmented image having the lowest confidence scores. This operation will strengthen the model to tackle uncontrollable regions. Combining these two strategies, our approach performs a novel input perturbation method, which can be directly applied to existing consistency learning approaches. Extensive

experiments show that ABD achieves new state-of-the-art (SOTA) performances for SSMIS, significantly improving different baseline performances.

We summarize our main contributions as follows:

- We observed that the combination of different perturbations leads to instability in consistency learning. To address this issue, we propose Adaptive Bidirectional Displacement to enable the generation of semantically complementary data by replacing model inadaptible regions with credible regions, which assists the model in effectively correcting erroneous predictions and enhances the quality of consistency learning.
- To take full advantage of labeled data, we propose an enhanced Adaptive Bidirectional Displacement that incorporates inverse confidence for labeled data to enforce the model to tackle those potentially uncontrollable regions.
- The proposed method can be easily plug-and-play, which can be embedded into different approaches and enhance their performance. Extensive experiments are conducted to validate the feasibility of the method, resulting in significant improvements compared to previous SOTA approaches on different datasets.

2. Related Work

2.1. Consistency Learning in Semi-Supervised Medical Image Segmentation

Consistency learning has been widely used in recent approaches for SSMIS. It aims to improve the performance of models by promoting consistent predictions for unlabeled data under different perturbations. According to the perturbation differences, consistency learning has three categories: input perturbation, feature perturbation, and network perturbation. Input perturbation is achieved through producing different inputs. For instance, Huang *et al.* [11] introduced cutout content loss and slice misalignment as perturbations in the input. In contrast, ST++ [39] involved strong augmentation on unlabeled images and directly utilizes the augmented samples for re-training. PseudoSeg [51] was similar to FixMatch [30] in leveraging pseudo-labels generated from weakly augmented images to supervise the predictions of strongly augmented images. BCP [2] encouraged the mixing of labeled and unlabeled images on the input level, enabling unlabeled data to learn comprehensive and general semantic information from labeled data in both directions. Meanwhile, feature perturbation and network perturbation are achieved by producing different features or outputs for the same input. Specifically, for feature perturbation, Mismatch [38] introduced morphological feature perturbation, which is based on classic morphological operations. For network perturbation, CPS [8] encouraged consistent predictions from different initialized networks with an input image. Mean Teacher [32] utilized

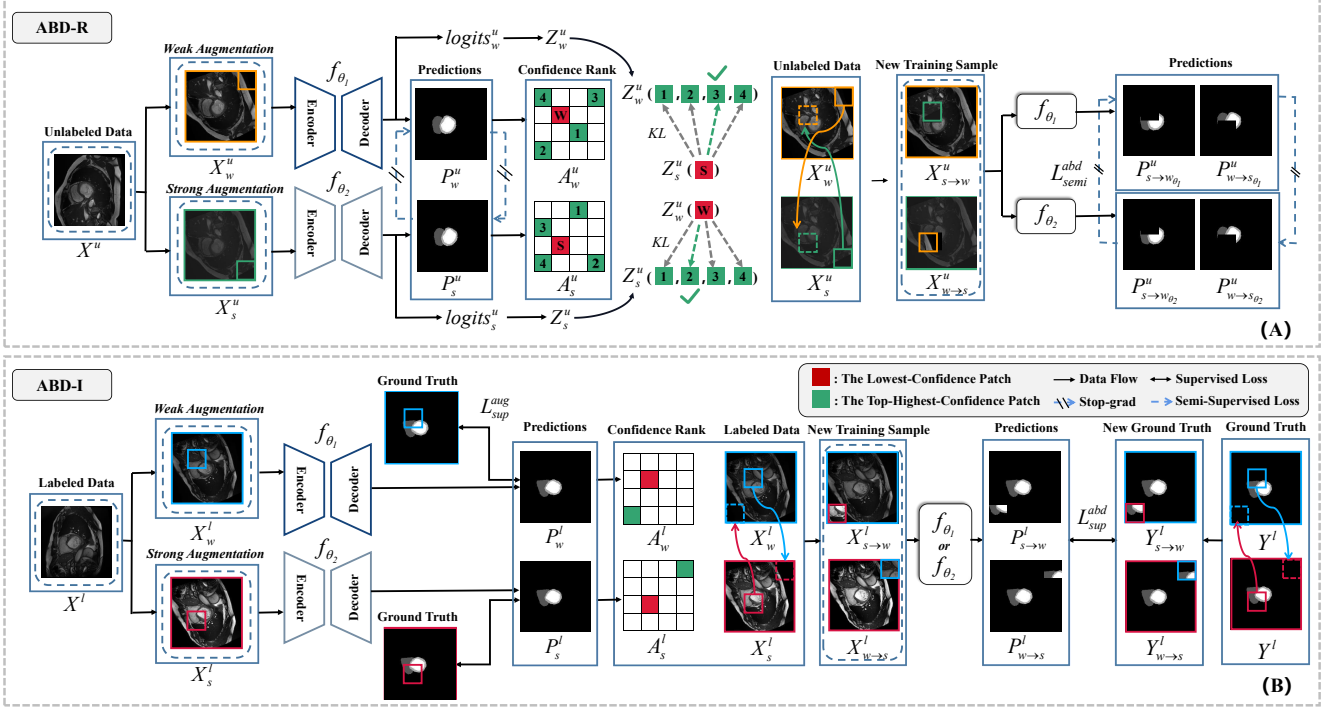


Figure 2. Overview of our adaptive bidirectional displacement framework. (A) For the unlabeled data, one image is subjected to weak and strong augmentations, resulting in two images that are separately input to two networks for cross-supervision. Then, based on the A_w^u , Z_w^u , A_s^u , and Z_s^u , the patches in the images are bidirectionally displaced, resulting in the formation of new samples $X_{s \rightarrow w}^u$ and $X_{w \rightarrow s}^u$. These new samples are further fed into the networks for cross-supervision. (B) For the labeled images, they are also subjected to both weak and strong augmentations, and their predictions are supervised by the labels. Afterward, based on the A_w^l and A_s^l , inverse bidirectional patch displacement is performed on the images, resulting in the generation of new samples $X_{s \rightarrow w}^l$ and $X_{w \rightarrow s}^l$. Similarly, the labels undergo the same operation, leading to the creation of new labels $Y_{s \rightarrow w}^l$ and $Y_{w \rightarrow s}^l$. The new samples are then fed into the network, and their predictions are supervised by the new labels. Note that ABD-R and ABD-I are two parallel modules during training.

the exponential moving average (EMA) technique to transfer semantic knowledge from the student network to the teacher network. However, existing methods only guarantee the consistency of predictions under a single perturbation, which greatly limits the scalability of consistency learning.

2.2. Consistency Learning in Other Tasks

Consistency learning [18, 46] is a major technique to tackle unlabeled data in semi-supervised learning. Several methods are proposed with consistency learning in various domains, including image segmentation [26, 40], domain generalization [49, 50], image classification [33, 48], pre-train language model [10], etc. For image segmentation, Revisiting UniMatch [40] proposed a unified approach using dual-stream perturbations for semi-supervised semantic segmentation by combining input and feature perturbations. Additionally, ViewCo [26] introduced text-to-view consistency modeling, incorporating additional text to learn comprehensive segmentation masks. In domain generalization, Zhou *et al.* [50] exploited the latent uncertainty information of the unlabeled samples to design an uncertainty-guided consistency

loss. StyleMatch [49] enforced prediction consistency between images from one domain and their style-transferred counterparts. In image classification, SimMatch [48] applied consistency regularization on both the semantic level and instance level to generate high-quality and reliable targets. ICT [33] introduced the interpolation consistency training which encourages the prediction at an interpolation of unlabeled points to be consistent with the interpolation of the predictions at those points. In addition, in the pre-train language model, GALAXY [10] applied consistency regularization on all data to minimize the bi-directional KL divergence between model predictions.

3. Method

3.1. Problem Setting

In the semi-supervised segmentation, we utilize a labeled dataset $\mathcal{D}^l = \{(X_i^l, Y_i^l)\}_{i=1}^N$ along with an unlabeled dataset $\mathcal{D}^u = \{X_i^u\}_{i=N+1}^{M+N}$. Here, X_i^l represents a labeled image and Y_i^l represents its corresponding label. Similarly,

X_i^u denotes an unlabeled image. For convenience, **i will be omitted in the following part.** It is worth emphasizing that the number of labeled images is significantly larger than the number of unlabeled images, *i.e.*, $N \gg M$.

3.2. Overview

The overall framework of our approach is shown in Fig. 2, which can be divided into the following steps:

1. For the input labeled image, we first apply input perturbation to generate two different samples, *e.g.*, a weak augmentation image and a strong augmentation image, both of which are input to the model with network perturbation: the weak augmentation image is input to one network and the strong augmentation image is input to the other network to generate two corresponding predictions, using the provided label as supervision.
2. For an unlabeled image, two predictions are obtained following the above pipeline. Then using our proposed ABD-R, the corresponding confidence matrices are used to perform bidirectional displacement between two augmented unlabeled images to generate new input samples.
3. Meanwhile, to enforce learning from the potentially uncontrollable regions, the ABD-I strategy based on inverse confidence is applied to labeled data to generate two new samples with the corresponding labels.
4. Finally, all generated new samples are input to the model to produce the corresponding predictions. The predictions of unlabeled data are used for cross-supervision, and the predictions of labeled data are supervised by the newly generated labels.

3.3. Adaptive Bidirectional Displacement with Reliable Confidence

Under various perturbations, the segmentation model produces unreliable predictions for unlabeled data. Imposing alignment using these predictions can render consistent learning ineffective. A direct solution is to remove the regions related to unreliable prediction. In practice, we first generate the confidence matrix based on the model's predictions. The confidence matrix measures the confidence of each predicted pixel belonging to different categories, it reflects the model's reliability to different perturbations. Utilizing the confidence matrix, our ABD-R can generate a new training sample that exhibits more reliable regions.

Suppose $X_w^u \in \mathbb{R}^{3 \times H \times W}$ is an unlabeled medical image with weak augmentation. $X_s^u \in \mathbb{R}^{3 \times H \times W}$ is the same unlabeled medical image but with strong augmentation. After passing the model, we generate their corresponding outputs:

$$\text{logits}_w^u = f_{\theta_1}(X_w^u), \text{logits}_s^u = f_{\theta_2}(X_s^u), \quad (1)$$

where f_{θ_1} and f_{θ_2} are two networks, which usually have different architecture or initialization to build the network

perturbation. logits_w^u and logits_s^u are the logits outputs that correspond to X_w^u and X_s^u , respectively. Applying softmax to these logits yields the corresponding prediction probability scores:

$$P_w^u = \text{softmax}(\text{logits}_w^u), P_s^u = \text{softmax}(\text{logits}_s^u), \quad (2)$$

where P_w^u is the prediction of X_w^u and P_s^u is the prediction of X_s^u . After upsampling them to the same height and width with input, we can generate $P_w^u \in \mathbb{R}^{C \times H \times W}$ and $P_s^u \in \mathbb{R}^{C \times H \times W}$. C is the number of classes.

Then we divide X_w^u into K patches, each with a size of $k \times k$, denoting as $X_w^u = \{X_w^{u,j}\}_{j=1}^K$, where $X_w^{u,j} \in \mathbb{R}^{k \times k}$ and $K = (\frac{H}{k}) \times (\frac{W}{k})$. For any patch $X_w^{u,j}$ (j is the index of the patch), the corresponding logits outputs are represented as $\text{logits}_w^{u,j} \in \mathbb{R}^{C \times k \times k}$ and predictions are represented as $P_w^{u,j} \in \mathbb{R}^{C \times k \times k}$.

Then, for each patch of X_w^u , its logits scores is:

$$Z_{w,c}^{u,j} = \frac{\sum_{m=1}^{k \times k} \text{logits}_{w,c}^{u,j}(m)}{|k \times k|}, \quad (3)$$

where $Z_{w,c}^{u,j}$ represents the average logits score of the j -th patch for the class c and $c \in \{1, 2, \dots, C\}$, $\text{logits}_{w,c}^{u,j}(m)$ is the logit score of the class c for the m -th pixel in the j -th patch. We regard $Z_w^{u,j}$ as the output distribution for the j -th patch of X_w^u .

Meanwhile, the corresponding confidence score for the j -th patch is computed as follows:

$$A_w^{u,j} = \frac{\sum_{m=1}^{k \times k} \max_{c \in C} (P_{w,c}^{u,j}(m))}{|k \times k|}, \quad (4)$$

where j is the index of the corresponding patch, $P_{w,c}^{u,j}(m)$ is the probability of the class c for the m -th pixel in the patch. $\max(\cdot)$ is a maximum operator to select the highest confidence score for each pixel. $A_w^{u,j}$ is the average confidence score for the j -th patch, which measures the reliability of the corresponding patch.

Then the index of the patch with the lowest confidence for X_w^u is computed as:

$$\text{ind}_{w-\min}^u = \underset{j \in K}{\text{argmin}} (A_w^{u,j}). \quad (5)$$

Simultaneously, using the confidence map A_w^u , the top n highest confidence patches in X_w^u are selected, and the corresponding index set is represented as:

$$\text{Ind}_{w-\text{top}}^u = \{\text{ind}_{w-\max^1}^u, \text{ind}_{w-\max^2}^u, \dots, \text{ind}_{w-\max^n}^u\}, \quad (6)$$

where $\text{Ind}_{w-\text{top}}^u$ is an index set that includes the indices of the selected top n highest confidence patches. $\text{ind}_{w-\max^1}^u$

is the selected indices, for example, ind_{w-max}^u is the index of the patch that has the highest (top 1) confidence score.

Similarly, following the above operations, we divide X_s^u into K patches, the corresponding logits and confidence score are represented as $Z_w^{u,j}$ and $A_s^{u,j}$, respectively. After processing from Eq. (5) to Eq. (6), the indices of the top-highest and the lowest confidence for X_s^u are generated, representing as ind_{s-top}^u and ind_{s-min}^u , respectively. And $\text{Ind}_{s-top}^u = \{\text{ind}_{s-max}^u, \text{ind}_{s-max}^u, \dots, \text{ind}_{s-max}^u\}$.

To achieve controlled effects of the mixed perturbations and enhance semantic coherence of the displacement operation, we select the most similar patch from the obtained index sets (Ind_{w-top}^u and Ind_{s-top}^u) by calculating the difference of the output distribution between the selected top- n highest confidence patches in one image and the lowest confidence patch in the other image:

$$\begin{aligned} \text{ind}_{w-max}^s &= \underset{i \in n}{\text{argmin}} \left(kl_{div}(Z_w^{u, \text{ind}_{w-max}^u}, Z_s^{u, \text{ind}_{s-min}^u}) \right) \\ \text{ind}_{s-max}^s &= \underset{i \in n}{\text{argmin}} \left(kl_{div}(Z_s^{u, \text{ind}_{s-max}^u}, Z_w^{u, \text{ind}_{w-min}^u}) \right), \end{aligned} \quad (7)$$

where kl_{div} is to compute the KL divergence [14], *e.g.*, $kl_{div}(Z_w^*, Z_s^*) = \text{softmax}(Z_w^*) \log \frac{\text{softmax}(Z_w^*)}{\text{softmax}(Z_s^*)}$, which reflects the difference in the output distribution. ind_{w-max}^s and ind_{s-max}^s represent the indices of final selected patches.

Finally, the bidirectional displacement operation is performed for X_s^u and X_w^u . The patch of one augmented image with the lowest confidence score is displaced with the final selected patch from the other augmented image:

$$X_{s \rightarrow w}^{u,j} = \begin{cases} X_s^{u, \text{ind}_{s-max}^u}, & \text{if } j = \text{ind}_{w-min}^u \\ X_w^{u,j}, & \text{else} \end{cases}, \quad (8)$$

$$X_{w \rightarrow s}^{u,j} = \begin{cases} X_w^{u, \text{ind}_{w-max}^u}, & \text{if } j = \text{ind}_{s-min}^u \\ X_s^{u,j}, & \text{else} \end{cases}. \quad (9)$$

After removing the patching operation and reshaping them to the image, two new samples $X_{s \rightarrow w}^u$ and $X_{w \rightarrow s}^u$ are generated, which are then input to two networks f_{θ_1} and f_{θ_2} for cross-supervision, as shown in Eq. (18).

This approach allows the two networks to adapt to each other's input perturbations, resulting in higher quality and more reliable pseudo-labels. As a cross-supervised process for the unlabeled data, it also reduces the likelihood of one model making erroneous predictions, thereby preventing the degradation of the other model's training process.

3.4. Adaptive Bidirectional Displacement with Inverse Confidence

In the above ABD-R module, unlabeled images are manipulated to create new samples by replacing the potentially

uncontrollable patches with new patches, it does not directly learn from the original regions. To enforce the model to learn from these potentially uncontrollable regions, we propose Adaptive Bidirectional Displacement with Inverse Confidence (ABD-I) for the labeled images to strengthen the learning for uncontrollable regions.

Specifically, for a labeled image X^l , we perform a strong augmentation to obtain a sample X_s^l and a weak augmentation to obtain a sample X_w^l . Let X_w^l as an example, we execute Eq. (1) to Eq. (6) on X_w^l to get its region index, *i.e.*, ind_{w-max}^l and ind_{w-min}^l . Similarly, we repeat the above operation for the strong augmented image X_s^l to get ind_{s-max}^l and ind_{s-min}^l . Note that ind_{w-max}^l and ind_{s-max}^l are the indices of the patches that have the highest (top 1) confidence score, we directly use them rather than selecting from candidates since these regions are the most potentially controllable region and we do not need to consider the semantic coherence with the given annotations. ind_{w-min}^l and ind_{s-min}^l share the same definition with Eq. (5).

Then ABD-I is used, the region of one image with the highest confidence scores is displaced with the region from the other image having the lowest confidence scores:

$$X_{s \rightarrow w}^{l,j} = \begin{cases} X_s^{l, \text{ind}_{s-min}^l}, & \text{if } j = \text{ind}_{w-max}^l \\ X_w^{l,j}, & \text{else} \end{cases}, \quad (10)$$

$$X_{w \rightarrow s}^{l,j} = \begin{cases} X_w^{l, \text{ind}_{w-min}^l}, & \text{if } j = \text{ind}_{s-max}^l \\ X_s^{l,j}, & \text{else} \end{cases}. \quad (11)$$

Using Eq. (10) and Eq. (11), the new samples $X_{s \rightarrow w}^l$ and $X_{w \rightarrow s}^l$ are generated. To effectively supervise the displaced samples, the same displacement is applied to the original label. Suppose Y^l is the label of images X_s^l and X_w^l , we also divide Y^l into K patches, denoting as $Y^{l,j} = \{Y^{l,j}\}_{j=1}^K$. The corresponding displacement operation is expressed as:

$$Y_{s \rightarrow w}^{l,j} = \begin{cases} Y^{l, \text{ind}_{s-min}^l}, & \text{if } j = \text{ind}_{w-max}^l \\ Y^{l,j}, & \text{else} \end{cases}, \quad (12)$$

$$Y_{w \rightarrow s}^{l,j} = \begin{cases} Y^{l, \text{ind}_{w-min}^l}, & \text{if } j = \text{ind}_{s-max}^l \\ Y^{l,j}, & \text{else} \end{cases}. \quad (13)$$

Finally $Y_{s \rightarrow w}^l$ and $Y_{w \rightarrow s}^l$ are used to supervise the generated samples $X_{s \rightarrow w}^l$ and $X_{w \rightarrow s}^l$, as shown in Eq. (16).

3.5. Loss Functions

The overall loss function consists of two parts, the supervised loss \mathcal{L}_{sup} for labeled data and the unsupervised loss \mathcal{L}_{semi} for unlabeled data, defined as:

$$\begin{aligned} \mathcal{L} &= \mathcal{L}_{sup} + \lambda \mathcal{L}_{semi} \\ &= (\mathcal{L}_{sup}^{aug} + \mathcal{L}_{sup}^{abd}) + \lambda (\mathcal{L}_{semi}^{aug} + \mathcal{L}_{semi}^{abd}) \end{aligned} \quad (14)$$

where λ is a weight that balances different losses, which gradually increase with iterations based on the Gaussian warming up: $\lambda = 0.1 \cdot e^{-5 \times (1-t/t_{\text{total}})^2}$, t is the current iteration and t_{total} is the total iteration.

The \mathcal{L}_{sup} consists of $\mathcal{L}_{\text{sup}}^{\text{aug}}$ and $\mathcal{L}_{\text{sup}}^{\text{abd}}$. $\mathcal{L}_{\text{sup}}^{\text{aug}}$ is used for the original augmented labeled images, $\mathcal{L}_{\text{sup}}^{\text{abd}}$ is used for the newly generated labeled samples.

The loss $\mathcal{L}_{\text{sup}}^{\text{aug}}$ is computed as:

$$\mathcal{L}_{\text{sup}}^{\text{aug}} = \frac{1}{2} \times (\mathcal{L}_{\text{ce}}(f_{\theta_1}(X_w^l), Y^l) + \mathcal{L}_{\text{dice}}(f_{\theta_1}(X_w^l), Y^l)) + \frac{1}{2} \times (\mathcal{L}_{\text{ce}}(f_{\theta_2}(X_s^l), Y^l) + \mathcal{L}_{\text{dice}}(f_{\theta_2}(X_s^l), Y^l)) \quad (15)$$

where \mathcal{L}_{ce} , $\mathcal{L}_{\text{dice}}$ are the cross-entropy loss and dice loss [24], respectively. Y^l is the label.

The loss $\mathcal{L}_{\text{sup}}^{\text{abd}}$ is computed as:

$$\mathcal{L}_{\text{sup}}^{\text{abd}} = \frac{1}{2} \times (\mathcal{L}_{\text{ce}}(f_{\theta_1}(X_{s \rightarrow w}^l), Y_{s \rightarrow w}^l) + \mathcal{L}_{\text{dice}}(f_{\theta_1}(X_{s \rightarrow w}^l), Y_{s \rightarrow w}^l)) + \frac{1}{2} \times (\mathcal{L}_{\text{ce}}(f_{\theta_2}(X_{w \rightarrow s}^l), Y_{w \rightarrow s}^l) + \mathcal{L}_{\text{dice}}(f_{\theta_2}(X_{w \rightarrow s}^l), Y_{w \rightarrow s}^l)) \quad (16)$$

The unsupervised loss $\mathcal{L}_{\text{semi}}$ consists of $\mathcal{L}_{\text{semi}}^{\text{aug}}$ and $\mathcal{L}_{\text{semi}}^{\text{abd}}$. $\mathcal{L}_{\text{semi}}^{\text{aug}}$ is used for the original augmented unlabeled images, $\mathcal{L}_{\text{semi}}^{\text{abd}}$ is for the newly generated unlabeled images.

The loss function $\mathcal{L}_{\text{semi}}^{\text{aug}}$ is computed as follows:

$$\mathcal{L}_{\text{semi}}^{\text{aug}} = \mathcal{L}_{\text{dice}}(f_{\theta_1}(X_w^u), \argmax_{c \in C}(P_{s,c}^u)) + \mathcal{L}_{\text{dice}}(f_{\theta_2}(X_s^u), \argmax_{c \in C}(P_{w,c}^u)) \quad (17)$$

where $P_{s,c}^u$ and $P_{w,c}^u$ are the predictions in Eq (2).

The loss function $\mathcal{L}_{\text{semi}}^{\text{abd}}$ is defined as:

$$\mathcal{L}_{\text{semi}}^{\text{abd}} = \mathcal{L}_{\text{dice}}(f_{\theta_1}(X_{s \rightarrow w}^u), \argmax_{c \in C}(P_{s \rightarrow w_{\theta_2},c}^u)) + \mathcal{L}_{\text{dice}}(f_{\theta_2}(X_{s \rightarrow w}^u), \argmax_{c \in C}(P_{s \rightarrow w_{\theta_1},c}^u)) + \mathcal{L}_{\text{dice}}(f_{\theta_1}(X_{w \rightarrow s}^u), \argmax_{c \in C}(P_{w \rightarrow s_{\theta_2},c}^u)) + \mathcal{L}_{\text{dice}}(f_{\theta_2}(X_{w \rightarrow s}^u), \argmax_{c \in C}(P_{w \rightarrow s_{\theta_1},c}^u)) \quad (18)$$

where $P_{s \rightarrow w_{\theta_1}}^u = f_{\theta_1}(X_{s \rightarrow w}^u)$ and $P_{s \rightarrow w_{\theta_2}}^u = f_{\theta_2}(X_{s \rightarrow w}^u)$, which are the predictions from f_{θ_1} and f_{θ_2} , respectively. Similarly, $P_{w \rightarrow s_{\theta_1}}^u$ and $P_{w \rightarrow s_{\theta_2}}^u$ are obtained using the same operation.

4. Experiments

4.1. Dataset and Evaluation Metrics

ACDC dataset: The ACDC dataset [4] comprises 200 annotated short-axis cardiac cine-MR images from a cohort

of 100 patients with four classes. 2D segmentation is more common than direct 3D segmentation [1]. The data split remains 70 patients' scans for training, 10 for validation, and 20 for testing. Following previous approaches, four evaluation metrics are chosen: *Dice Similarity Coefficient* (DSC), *Jaccard*, *95% Hausdorff Distance* (95HD) and *Average Surface Distance* (ASD).

PROMISE12 dataset: The PROMISE12 dataset [19] was made available for the MICCAI 2012 prostate segmentation challenge. MRI of 50 patients with various diseases was acquired at different locations. All 3D scans are converted into 2D slices. Following previous approaches, DSC and ASD are used as evaluation metrics.

4.2. Implementation Details

We evaluate our approach on two baselines: Cross Teaching [23] and BCP [2].

Cross Teaching [23] is a cross-supervision framework, which utilizes two networks to perform the network perturbation: f_{θ_1} and f_{θ_2} , representing UNet [27] and Swin-UNet [6], respectively. We add the input perturbation using weak and strong data augmentation to provide two kinds of inputs. For weak data augmentation, random rotation and flipping are used. For strong data augmentations, colorjitter, blur [7] and Cutout [39] are used. During training, all inputs are cropped to 224×224 and divided into $K = 16$ patches. The networks are trained with a batch size of 16, including 8 labeled slices and 8 unlabeled slices.

BCP [2] is a mean-teacher framework, which applies weak and strong augmentation to perform input perturbation. We add the network perturbation by providing two student networks with different initializations. Our ABD is used on these two student networks following the above design. During training, all inputs are cropped to 256×256 and divided into $K = 16$ patches.

All other settings follow the default settings with original Cross Teaching [23] and BCP [2].

4.3. Comparison with State-of-the-Art Methods

ACDA dataset: Table 1 shows the average performance on the ACDC dataset using 5% and 10% labeled ratios. For the 5% label ratio, our ABD achieves state-of-the-art performance compared to recent methods, surpassing the baseline Cross Teaching by a large margin. BCP is one of the newly released methods, it achieved high performance on the ACDC dataset. By incorporating our method into BCP, i.e., **Ours-ABD** (BCP), it has a noticeable improvement, which highlights the flexibility and scalability of our approach. Specifically, we observe a 20.75% DSC improvement compared to Cross Teaching and a 1.37% DSC improvement compared to BCP. Considering the 10% labeled ratio, our ABD outperforms SCP-Net [44] in all evaluation metrics and achieves a new state-of-the-art performance. It

Table 1. Comparisons with other methods on the ACDC test set, “Ours-ABD (Cross Teaching)” and “Ours-ABD (BCP)” represent the baseline is Cross Teaching [23] and BCP [2] respectively.

Method	Scans used		Metrics			
	Labeled	Unlabeled	DSC↑	Jaccard↑	95HD↓	ASD↓
U-Net (MICCAI’2015) [27]	3(5%)	0	47.83	37.01	31.16	12.62
	7(10%)	0	79.41	68.11	9.35	2.70
	70(All)	0	91.44	84.59	4.30	0.99
DTC (AAAI’2021) [21]	3(5%)	67(95%)	56.90	45.67	23.36	7.39
URPC (MICCAI’2021) [22]			55.87	44.64	13.60	3.74
MC-Net (MICCAI’2021) [36]			62.85	52.29	7.62	2.33
SS-Net (MICCAI’2022) [37]			65.83	55.38	6.67	2.28
SCP-Net (MICCAI’2023) [44]			87.27	-	-	2.65
Cross Teaching (Reported) (MIDL’2022) [23]			65.60	-	16.2	-
BCP (CVPR’2023) [2]			87.59	78.67	1.90	0.67
Ours-ABD (Cross Teaching)			86.35	76.73	4.12	1.22
Ours-ABD (BCP)			88.96	80.70	1.57	0.52
DTC (AAAI’2021) [21]	7(10%)	63(90%)	84.29	73.92	12.81	4.01
URPC (MICCAI’2021) [22]			83.10	72.41	4.84	1.53
MC-Net (MICCAI’2021) [36]			86.44	77.04	5.50	1.84
SS-Net (MICCAI’2022) [37]			86.78	77.67	6.07	1.40
SCP-Net (MICCAI’2023) [44]			89.69	-	-	0.73
PLGCL (CVPR’2023) [3]			89.1	-	4.98	1.80
Cross Teaching (Reported) (MIDL’2022) [23]			86.40	-	8.60	-
Cross Teaching (Reproduced)			86.45	77.02	6.30	1.86
BCP (CVPR’2023) [2]			88.84	80.62	3.98	1.17
Ours-ABD (Cross Teaching)			88.52	79.97	5.06	1.43
Ours-ABD (BCP)			89.81	81.95	1.46	0.49

Table 2. Comparisons with state-of-the-art semi-supervised segmentation methods on the PROMISE12 test set.

Method	Scans used		Metrics	
	Labeled	Unlabeled	DSC↑	ASD↓
U-Net [27]	7(20%)	0	60.88	13.87
	35(100%)	0	84.76	1.58
CCT [25]	7(20%)	28(80%)	71.43	16.61
URPC [22]			63.23	4.33
SS-Net [37]			62.31	4.36
SLC-Net [20]			68.31	4.69
SCP-Net [44]			77.06	3.52
Ours-ABD	3(10%)	32(90%)	82.06	1.33
Ours-ABD			81.81	1.46

demonstrates that ABD mitigates the uncontrolled effects caused by the mixture of multiple perturbations and validates the mixed perturbations could expand the upper bound of consistency learning.

Fig. 3 gives some qualitative results, our method effectively suppresses areas that are inaccurately segmented by

Cross Teaching and BCP, while accurately segmenting the foregrounds that are ignored by the two methods.

PROMISE12 dataset: Following SS-Net [37], we conduct experiments with 20% labeled data. We compare ABD with CCT [25], URPC [22], SS-Net [37], SLC-Net [20] and SCP-Net [44]. As shown in Table 2, Ours-ABD surpasses all other approaches. Compared to the recently proposed SCP-Net, ABD brings a 5.0% DSC increase.

4.4. Ablation Studies

We select Cross Teaching [23] as the baseline. All experiments are conducted on the ACDC dataset with 10% labeled data unless otherwise stated.

Effectiveness of each module in ABD: Table 3 demonstrates the effectiveness of each module in ABD by progressively adding ABD-R and ABD-I. It can be observed that the introduction of input perturbation (IP) decreases the performance of the baseline, indicating that the mixed perturbation easily leads to out-of-control. Incorporating ABD-R and ABD-I brings increased performance. By incorporating both ABD-R and ABD-I into the baseline, the model has the best result, reaching 88.52% DSC and surpassing

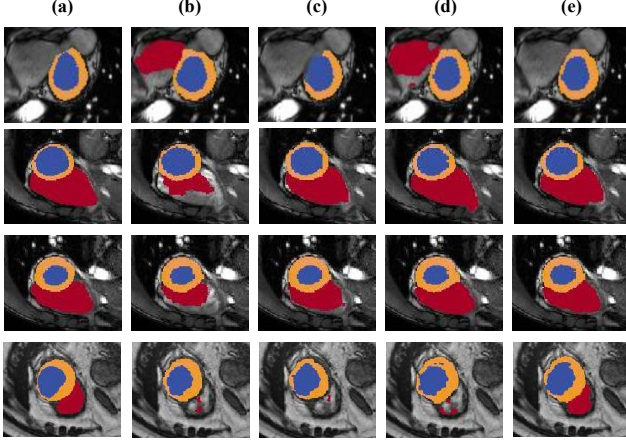


Figure 3. Visualization of segmentation results on ACDC dataset with 10% labeled data. (a) Ground-truth. (b) Cross Teaching results. (c) **Ours**-ABD (Cross Teaching) results. (d) BCP results. (e) **Ours**-ABD (BCP) results. Best viewed in color on the screen.

Table 3. Effectiveness of ABD-R and ABD-I modules. “base” means baseline is the Cross Teaching. “IP” means adding perturbation to the input.

Base	IP	ABD-R	ABD-I	DSC↑	Jaccard↑	95HD↓	ASD↓
✓				86.45	77.02	6.30	1.86
✓	✓			86.25	76.69	5.44	1.72
✓	✓	✓		87.42	78.37	5.23	1.68
✓	✓		✓	87.20	78.07	6.06	1.96
✓	✓	✓	✓	88.52	79.97	5.06	1.43

the baseline in 2.07%. The improvement indicates that the two modules are complementary which is consistent with the design targets. ABD-R enhances the upper limit of consistency learning for mixed perturbations, and ABD-I enables the model to learn potentially uncontrollable regions, thereby the combination can learn more semantics.

Displacement Strategies: Table 4 illustrates the impact of different displacement strategies. The default displacement strategy (denoted as “Reliable”) in our ABD-R module is displacing the region with the top-highest confidence from one image with the region of lowest confidence from the other image. “Random” refers to randomly selecting patches for displacement. “Same” means the displacement strategy is performed by displacing the region that has the highest confidence from one image with the corresponding region from the other image. “Same+Reliable” represents that in each iteration, we randomly select a strategy from the “Same” or “Reliable”. The “Same” generates samples that have more uncontrolled regions and yields unsatisfactory results. While the “Same+Reliable” alleviates the unfavorable results, it is still not as effective as our approach.

Selection of Patch Number: Table 5 indicates the impact of the patch number. K represents the number of patches. An optimal result is achieved when $K = 16$.

Table 4. Influence of the displacement strategies in ABD-R.

Strategy	DSC↑	Jaccard↑	95HD↓	ASD↓
Random	86.55	77.07	6.13	1.74
Same	87.22	78.04	5.61	1.50
Same+Reliable	87.38	78.06	4.56	1.69
Reliable	87.42	78.37	5.23	1.68

Table 5. Ablation study of patch number K .

K	DSC↑	Jaccard↑	95HD↓	ASD↓
4	87.39	78.35	5.06	1.69
16	88.52	79.97	5.06	1.43
64	87.54	78.57	5.52	1.88

Table 6. Effectiveness of the strong data augmentation.

Cutout	Colorjitter	Blur	DSC↑	Jaccard↑	95HD↓	ASD↓
✓			88.23	79.53	5.90	1.40
	✓		88.03	79.34	7.15	1.76
		✓	87.76	78.76	7.28	1.61
✓	✓		88.52	79.97	5.06	1.43
✓	✓	✓	87.83	79.02	6.14	1.87

Choice of Strong Augmentation: Table 6 illustrates the influence of three different augmentation methods: colorjitter, blur, and cutout. Through comparison, combining colorjitter and cutout produces better performance.

5. Conclusion

In this paper, we proposed an adaptive bidirectional displacement (ABD) for semi-supervised medical image segmentation. Our key idea is to mitigate the constraints of mixed perturbations on consistency learning, thereby enhancing the upper limit of consistency learning. To achieve this, we designed two novel modules in our ABD: an ABD-R module reduces the uncontrolled regions in unlabeled samples and captures comprehensive semantic information from input perturbations, and an ABD-I module enhances the learning capacity to uncontrollable regions in labeled samples to compensate for the deficiencies of ABD-R. With the cooperation of two modules, our method achieves state-of-the-art performance and is easily embedded into different methods. In the future, we will design a patch adaptive displacement strategy to tackle more complicated cases.

Acknowledge: This work was supported by National Natural Science Foundation of China (No. 62301613), the Taishan Scholar Program of Shandong (No. tsqn202306130), the Shandong Natural Science Foundation (No. ZR2023QF046), Qingdao Postdoctoral Applied Research Project (No. QDBSH20230102091) and Independent Innovation Research Project of China University of Petroleum (East China) (No. 22CX06060A).

References

- [1] Wenjia Bai, Ozan Oktay, Matthew Sinclair, Hideaki Suzuki, Martin Rajchl, Giacomo Tarroni, Ben Glocker, Andrew King, Paul M Matthews, and Daniel Rueckert. Semi-supervised learning for network-based cardiac mr image segmentation. In *Medical Image Computing and Computer-Assisted Intervention*, pages 253–260. Springer, 2017. 6
- [2] Yunhao Bai, Duowen Chen, Qingli Li, Wei Shen, and Yan Wang. Bidirectional copy-paste for semi-supervised medical image segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11514–11524, 2023. 1, 2, 6, 7
- [3] Hritam Basak and Zhaozheng Yin. Pseudo-label guided contrastive learning for semi-supervised medical image segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19786–19797, 2023. 7
- [4] Olivier Bernard, Alain Lalonde, Clement Zotti, Frederick Cervenansky, Xin Yang, Pheng-Ann Heng, Irem Cetin, Karim Lekadir, Oscar Camara, Miguel Angel Gonzalez Ballester, et al. Deep learning techniques for automatic mri cardiac multi-structures segmentation and diagnosis: is the problem solved? *IEEE transactions on medical imaging*, 37(11):2514–2525, 2018. 6
- [5] David Berthelot, Nicholas Carlini, Ian Goodfellow, Nicolas Papernot, Avital Oliver, and Colin A Raffel. Mixmatch: A holistic approach to semi-supervised learning. *Advances in neural information processing systems*, 32, 2019. 1
- [6] Hu Cao, Yueyue Wang, Joy Chen, Dongsheng Jiang, Xiaopeng Zhang, Qi Tian, and Manning Wang. Swin-unet: Unet-like pure transformer for medical image segmentation. In *European conference on computer vision*, pages 205–218. Springer, 2022. 6
- [7] Xinlei Chen, Haoqi Fan, Ross Girshick, and Kaiming He. Improved baselines with momentum contrastive learning. *arXiv preprint arXiv:2003.04297*, 2020. 6
- [8] Xiaokang Chen, Yuhui Yuan, Gang Zeng, and Jingdong Wang. Semi-supervised semantic segmentation with cross pseudo supervision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2613–2622, 2021. 1, 2
- [9] Jie Du, Xiaoci Zhang, Peng Liu, and Tianfu Wang. Coarse-refined consistency learning using pixel-level features for semi-supervised medical image segmentation. *IEEE Journal of Biomedical and Health Informatics*, 2023. 1
- [10] Wanwei He, Yinpei Dai, Yinhe Zheng, Yuchuan Wu, Zheng Cao, Dermot Liu, Peng Jiang, Min Yang, Fei Huang, Luo Si, et al. Galaxy: A generative pre-trained model for task-oriented dialog with semi-supervised learning and explicit policy injection. In *Proceedings of the AAAI conference on artificial intelligence*, number 10, pages 10749–10757, 2022. 3
- [11] Wei Huang, Chang Chen, Zhiwei Xiong, Yueyi Zhang, Xuejin Chen, Xiaoyan Sun, and Feng Wu. Semi-supervised neuron segmentation via reinforced consistency learning. *IEEE Transactions on Medical Imaging*, 41(11):3016–3028, 2022. 1, 2
- [12] Rushi Jiao, Yichi Zhang, Le Ding, Rong Cai, and Jicong Zhang. Learning with limited annotations: a survey on deep semi-supervised learning for medical image segmentation. *arXiv preprint arXiv:2207.14191*, 2022. 1
- [13] Zhanghan Ke, Di Qiu, Kaican Li, Qiong Yan, and Rynson WH Lau. Guided collaborative training for pixel-wise semi-supervised learning. In *Computer Vision—ECCV*, pages 429–445. Springer, 2020. 1
- [14] Solomon Kullback and Richard A Leibler. On information and sufficiency. *The annals of mathematical statistics*, 22(1): 79–86, 1951. 5
- [15] Xin Lai, Zhuotao Tian, Li Jiang, Shu Liu, Hengshuang Zhao, Liwei Wang, and Jiaya Jia. Semi-supervised semantic segmentation with directional context-aware consistency. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1205–1214, 2021. 1
- [16] Xiaomeng Li, Lequan Yu, Hao Chen, Chi-Wing Fu, Lei Xing, and Pheng-Ann Heng. Transformation-consistent self-ensembling model for semisupervised medical image segmentation. *IEEE Transactions on Neural Networks and Learning Systems*, 32(2):523–534, 2020. 2
- [17] Yanwen Li, Luyang Luo, Huangjing Lin, Hao Chen, and Pheng-Ann Heng. Dual-consistency semi-supervised learning with uncertainty quantification for covid-19 lesion segmentation from ct images. In *Medical Image Computing and Computer Assisted Intervention*, pages 199–209. Springer, 2021. 1
- [18] Chen Liang, Wenguan Wang, Jiaxu Miao, and Yi Yang. Logic-induced diagnostic reasoning for semi-supervised semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 16197–16208, 2023. 3
- [19] Geert Litjens, Robert Toth, Wendy Van De Ven, Caroline Hoeks, Sjoerd Kerkstra, Bram Van Ginneken, Graham Vincent, Gwenael Guillard, Neil Birbeck, Jindang Zhang, et al. Evaluation of prostate segmentation algorithms for mri: the promise12 challenge. *Medical image analysis*, 18(2):359–373, 2014. 6
- [20] Jinhua Liu, Christian Desrosiers, and Yuanfeng Zhou. Semi-supervised medical image segmentation using cross-model pseudo-supervision with shape awareness and local context constraints. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 140–150. Springer, 2022. 7
- [21] Xiangde Luo, Jieneng Chen, Tao Song, and Guotai Wang. Semi-supervised medical image segmentation through dual-task consistency. In *Proceedings of the AAAI conference on artificial intelligence*, number 10, pages 8801–8809, 2021. 7
- [22] Xiangde Luo, Wenjun Liao, Jieneng Chen, Tao Song, Yinan Chen, Shichuan Zhang, Nianyong Chen, Guotai Wang, and Shaoting Zhang. Efficient semi-supervised gross target volume of nasopharyngeal carcinoma segmentation via uncertainty rectified pyramid consistency. In *Medical Image Computing and Computer Assisted Intervention*, pages 318–329. Springer, 2021. 7
- [23] Xiangde Luo, Minhao Hu, Tao Song, Guotai Wang, and Shaoting Zhang. Semi-supervised medical image segmentation via cross teaching between cnn and transformer. In

- International Conference on Medical Imaging with Deep Learning*, pages 820–833. PMLR, 2022. 6, 7
- [24] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 fourth international conference on 3D vision (3DV)*, pages 565–571. Ieee, 2016. 6
- [25] Yassine Ouali, Céline Hudelot, and Myriam Tami. Semi-supervised semantic segmentation with cross-consistency training. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12674–12684, 2020. 1, 7
- [26] Pengzhen Ren, Changlin Li, Hang Xu, Yi Zhu, Guangrun Wang, Jianzhuang Liu, Xiaojun Chang, and Xiaodan Liang. Viewco: Discovering text-supervised segmentation masks via multi-view semantic consistency. *arXiv preprint arXiv:2302.10307*, 2023. 3
- [27] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention*, pages 234–241. Springer, 2015. 6, 7
- [28] Suman Sedai, Bhavna Antony, Ravneet Rai, Katie Jones, Hiroshi Ishikawa, Joel Schuman, Wollstein Gadi, and Rahil Garnavi. Uncertainty guided semi-supervised segmentation of retinal layers in oct images. In *Medical Image Computing and Computer Assisted Intervention*, pages 282–290. Springer, 2019. 1
- [29] Yucheng Shu, Hengbo Li, Bin Xiao, Xiuli Bi, and Weisheng Li. Cross-mix monitoring for medical image segmentation with limited supervision. *IEEE Transactions on Multimedia*, 2022. 1
- [30] Kihyuk Sohn, David Berthelot, Nicholas Carlini, Zizhao Zhang, Han Zhang, Colin A Raffel, Ekin Dogus Cubuk, Alexey Kurakin, and Chun-Liang Li. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. *Advances in neural information processing systems*, 33:596–608, 2020. 2
- [31] Nima Tajbakhsh, Laura Jeyaseelan, Qian Li, Jeffrey N Chiang, Zhihao Wu, and Xiaowei Ding. Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation. *Medical Image Analysis*, 63:101693, 2020. 1
- [32] Antti Tarvainen and Harri Valpola. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. *Advances in neural information processing systems*, 30, 2017. 2
- [33] Vikas Verma, Kenji Kawaguchi, Alex Lamb, Juho Kannala, Arno Solin, Yoshua Bengio, and David Lopez-Paz. Interpolation consistency training for semi-supervised learning. *Neural Networks*, 145:90–106, 2022. 3
- [34] Yan Wang, Yuyin Zhou, Wei Shen, Seyoun Park, Elliot K Fishman, and Alan L Yuille. Abdominal multi-organ segmentation with organ-attention networks and statistical fusion. *Medical image analysis*, 55:88–102, 2019. 1
- [35] Huisi Wu, Zhaoze Wang, Youyi Song, Lin Yang, and Jing Qin. Cross-patch dense contrastive learning for semi-supervised segmentation of cellular nuclei in histopathologic images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11666–11675, 2022. 1
- [36] Yicheng Wu, Minfeng Xu, Zongyuan Ge, Jianfei Cai, and Lei Zhang. Semi-supervised left atrium segmentation with mutual consistency training. In *Medical Image Computing and Computer Assisted Intervention*, pages 297–306. Springer, 2021. 7
- [37] Yicheng Wu, Zhonghua Wu, Qianyi Wu, Zongyuan Ge, and Jianfei Cai. Exploring smoothness and class-separation for semi-supervised medical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 34–43. Springer, 2022. 7
- [38] Mou-Cheng Xu, Yu-Kun Zhou, Chen Jin, Stefano B Blumberg, Frederick J Wilson, Marius deGroot, Daniel C Alexander, Neil P Oxtoby, and Joseph Jacob. Learning morphological feature perturbations for calibrated semi-supervised segmentation. In *International Conference on Medical Imaging with Deep Learning*, pages 1413–1429. PMLR, 2022. 2
- [39] Lihe Yang, Wei Zhuo, Lei Qi, Yinghuan Shi, and Yang Gao. St++: Make self-training work better for semi-supervised semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4268–4277, 2022. 2, 6
- [40] Lihe Yang, Lei Qi, Litong Feng, Wayne Zhang, and Yinghuan Shi. Revisiting weak-to-strong consistency in semi-supervised semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7236–7246, 2023. 3
- [41] Xiaosu Yang, Jiya Tian, Yaping Wan, Mingzhi Chen, Lingna Chen, and Junxi Chen. Semi-supervised medical image segmentation via cross-guidance and feature-level consistency dual regularization schemes. *Medical Physics*, 2023. 1
- [42] Shuo Zhang, Jiaojiao Zhang, Biao Tian, Thomas Lukasiewicz, and Zhenghua Xu. Multi-modal contrastive mutual learning and pseudo-label re-learning for semi-supervised medical image segmentation. *Medical Image Analysis*, 83:102656, 2023. 1
- [43] Yichi Zhang, Rushi Jiao, Qingcheng Liao, Dongyang Li, and Jicong Zhang. Uncertainty-guided mutual consistency learning for semi-supervised medical image segmentation. *Artificial Intelligence in Medicine*, 138:102476, 2023. 1
- [44] Zhenxi Zhang, Ran Ran, Chunna Tian, Heng Zhou, Xin Li, Fan Yang, and Zhicheng Jiao. Self-aware and cross-sample prototypical learning for semi-supervised medical image segmentation. *arXiv preprint arXiv:2305.16214*, 2023. 1, 6, 7
- [45] Xiangyu Zhao, Zengxin Qi, Sheng Wang, Qian Wang, Xuehai Wu, Ying Mao, and Lichi Zhang. Rcps: Rectified contrastive pseudo supervision for semi-supervised medical image segmentation. *arXiv preprint arXiv:2301.05500*, 2023. 1
- [46] Zhen Zhao, Lihe Yang, Sifan Long, Jimin Pi, Luping Zhou, and Jingdong Wang. Augmentation matters: A simple-yet-effective approach to semi-supervised semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Com-*

- puter Vision and Pattern Recognition*, pages 11350–11359, 2023. [3](#)
- [47] Ke Zheng, Junhai Xu, and Jianguo Wei. Double noise mean teacher self-ensembling model for semi-supervised tumor segmentation. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1446–1450. IEEE, 2022. [1](#)
 - [48] Mingkai Zheng, Shan You, Lang Huang, Fei Wang, Chen Qian, and Chang Xu. Simmatch: Semi-supervised learning with similarity matching. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14471–14481, 2022. [3](#)
 - [49] Kaiyang Zhou, Chen Change Loy, and Ziwei Liu. Semi-supervised domain generalization with stochastic stylematch. *International Journal of Computer Vision*, pages 1–11, 2023. [3](#)
 - [50] Qianyu Zhou, Zhengyang Feng, Qiqi Gu, Guangliang Cheng, Xuequan Lu, Jianping Shi, and Lizhuang Ma. Uncertainty-aware consistency regularization for cross-domain semantic segmentation. *Computer Vision and Image Understanding*, 221:103448, 2022. [3](#)
 - [51] Yuliang Zou, Zizhao Zhang, Han Zhang, Chun-Liang Li, Xiao Bian, Jia-Bin Huang, and Tomas Pfister. Pseudoseg: Designing pseudo labels for semantic segmentation. *arXiv preprint arXiv:2010.09713*, 2020. [2](#)