Medical Image Analysis 101 (2025) 103450

Contents lists available at ScienceDirect

Medical Image Analysis

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

Semi-supervised medical image segmentation via weak-to-strong perturbation consistency and edge-aware contrastive representation

Yang Yang [a], Guoying Sun [a], Tong Zhang [d], Ruixuan Wang [c,d,*], Jingyong Su [a,b,**]

[a] School of Computer Science and Technology, Harbin Institute of Technology at Shenzhen, Shenzhen, 518055, China
[b] National Key Laboratory of Smart Farm Technologies and Systems, Harbin, 150001, China
[c] School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou, 510006, China
[d] Department of Network Intelligence, Peng Cheng Laboratory, Shenzhen, 518055, China

ARTICLE INFO

Dataset link: https://github.com/youngyzzZ/SSL-w2sPC

Keywords:
Semi-supervised learning
Contrastive learning
Uncertainty estimation
Image segmentation

ABSTRACT

Despite that supervised learning has demonstrated impressive accuracy in medical image segmentation, its reliance on large labeled datasets poses a challenge due to the effort and expertise required for data acquisition. Semi-supervised learning has emerged as a potential solution. However, it tends to yield satisfactory segmentation performance in the central region of the foreground, but struggles in the edge region. In this paper, we propose an innovative framework that effectively leverages unlabeled data to improve segmentation performance, especially in edge regions. Our proposed framework includes two novel designs. Firstly, we introduce a weak-to-strong perturbation strategy with corresponding feature-perturbed consistency loss to efficiently utilize unlabeled data and guide our framework in learning reliable regions. Secondly, we propose an edge-aware contrastive loss that utilizes uncertainty to select positive pairs, thereby learning discriminative pixel-level features in the edge regions using unlabeled data. In this way, the model minimizes the discrepancy of multiple predictions and improves representation ability, ultimately aiming at impressive performance on both primary and edge regions. We conducted a comparative analysis of the segmentation results on the publicly available BraTS2020 dataset, LA dataset, and the 2017 ACDC dataset. Through extensive quantification and visualization experiments under three standard semi-supervised settings, we demonstrate the effectiveness of our approach and set a new state-of-the-art for semi-supervised medical image segmentation. Our code is released publicly at https://github.com/youngyzzZ/SSL-w2sPC.

1. Introduction

Automated semantic segmentation plays a critical role in medical image analysis, which has demonstrated exceptional performance in diverse segmentation tasks, reaching the forefront of the field (Yang and Farsiu, 2023; Lu et al., 2023; Gupta et al., 2022; Huang et al., 2022; Chen et al., 2023b; Zhuang et al., 2023). Despite the importance of fully supervised learning approaches in achieving satisfactory performance in semantic segmentation, their reliance on abundant and accurate annotations introduces challenges. Acquiring a large-scale dataset with pixel-wise annotations is often difficult due to its high cost and time-consuming nature. Semi-supervised learning presents a potential solution to overcome the challenge imposed by annotation scarcity in medical image analysis. These approaches leverage a combination of a limited number of labeled samples and a sufficiently large set of unlabeled samples to effectively train models.

The prevailing semi-supervised learning methods in deep learning mainly include consistency regularization (Rasmus et al., 2015; Tarvainen and Valpola, 2017; Li et al., 2022) and self-training with pseudo-labels (Fan et al., 2020; Bai et al., 2017). These approaches can be summarized by the utilization of limited annotated data to train an initial network, which is subsequently employed to generate pseudo-labels for unlabeled data, and appropriate constraints are applied to these pseudo-labels to extract the abundant semantic information embedded in the unlabeled data. This strategy utilizes both unlabeled data and the limited availability of strongly labeled samples to train the segmentation network.

These approaches suffer from one limitation. They impose appropriate constraints (e.g. consistency regularization) on multiple predictions following the introduction of image-level perturbations. However, image-level perturbation requires careful augmentation design and provides a limited diversity of augmented data (Li et al., 2021a). In

* Corresponding author at: School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou, 510006, China.
** Corresponding author at: School of Computer Science and Technology, Harbin Institute of Technology at Shenzhen, Shenzhen, 518055, China.
E-mail addresses: wangruix5@mail.sysu.edu.cn (R. Wang), sujingyong@hit.edu.cn (J. Su).

https://doi.org/10.1016/j.media.2024.103450
Received 23 February 2024; Received in revised form 2 December 2024; Accepted 26 December 2024
Available online 6 January 2025

contrast, feature-level perturbations facilitate a more extensive exploration within the feature space, enabling the model to better leverage unlabeled data (Yang et al., 2023a). As a result, a noticeable phenomenon arises wherein the majority of semi-supervised frameworks demonstrate the capability to accurately segment the primary body region of the object but fall short of satisfactory in edge regions.

To overcome this limitation, many existing consistency-based methods generate multiple predictions through the design of network architectures such as multiple decoders, pyramid structures, or MC dropout to better exploit unlabeled data, but these methods overlook the importance of meaningful perturbations at the feature level, which can improve the reliability and performance of model outputs. Inspired by recent studies (Ouali et al., 2020; Miyato et al., 2018; Wu et al., 2022), which indicate that random variations in features can lead to inconsistent and inaccurate predictions, we further observe that the introduction of perturbations on unlabeled data induces substantial changes in the edge regions. Based on this phenomenon, we propose a novel semi-supervised learning framework that combines weak-to-strong perturbation for consistency regularization and edge-aware contrastive learning for effective exploitation of unlabeled data to achieve accurate segmentation. Specifically, we design a shared encoder and multiple decoders, each equipped with dedicated weak-to-strong perturbation modules. These weak-to-strong perturbation modules at the semantic level introduce slight variations in the predictions of the decoder. The pixel/voxel-level uncertainty associated with these predictions is then used to obtain an uncertainty-weighted aggregation label. To guide the model in learning reliable regions of predictions, a feature-perturbed consistency is imposed between the predictions generated by the decoders and the aggregation label. We also define an edge-aware contrastive loss for unlabeled data to improve performance in edge regions. This contrastive loss meticulously selects positive pairs by uncertainty ranking, directing the model's attention towards edge regions. It encourages the similarity of representations for pixels belonging to the same class semantic label in the edge region, while simultaneously ensuring their dissimilarity to the representations of pixels from different classes.

Concretely, our contributions are summarized as follows:

- We develop a weak-to-strong feature-perturbed consistency scheme that encourages the consistency of reliable regions across different predictions. This scheme consists of weak-to-strong feature-level perturbation and feature-perturbed consistency. Inspired by previous work (Yang et al., 2023b), the proposed feature perturbations incorporate additional statistical information along the channel dimension and adopt a different integration way in the model. This scheme allows our model to focus on learning reliable predictions and mitigate the negative effects of unreliable regions, leading to more effective training.
- We propose a novel edge-aware contrastive loss that effectively leverages class-discriminative representations, specifically in the edge regions. This loss function improves the discrimination between class edges by sampling positive pixels from object edges guided by uncertainty, resulting in impressive segmentation performance in edge regions.
- Extensive evaluations on both 2D and 3D datasets for lesion and organ segmentation demonstrate the superiority of our method, with new state-of-the-art performance achieved in semi-supervised segmentation.

2. Related work

This paper is mainly based on semi-supervised learning and contrastive learning. Therefore, our attention is directed towards exploring and discussing relevant literature in these two specific fields.

2.1. Semi-supervised learning

Recently, several semi-supervised learning methods (Wang et al., 2023c; Lei et al., 2023; Chen et al., 2023c; Zhao et al., 2023; Yang et al., 2023a; Chaitanya et al., 2023) have been proposed leveraging both unlabeled and labeled images during training. Depending on the training strategy, semi-supervised learning can be broadly classified into the following categories:

Self-training based: In this methodology, the model initialized with labeled data is employed to generate preliminary predictions for the unlabeled data. Then, the annotated labels from the labeled data, along with the pseudo-labels generated from the previous iteration's prediction results, are utilized as ground truths for iteratively updating the network. This iterative process ensures the progressive refinement of the network's performance, and the pseudo-label estimates are periodically updated after a few epochs of training, with the expectation that their quality will progressively improve throughout the training process. This paradigm has demonstrated remarkable improvements in the field of medical image segmentation (Lyu et al., 2022; Basak and Yin, 2023). Nevertheless, this method has a drawback. Since the initial model relies on a limited amount of labeled data for initialization, its segmentation performance is unsatisfactory. Limited segmentation ability can lead to the generation of low-quality pseudo-labels, which in turn can hinder the subsequent training of the model and prevent its performance from improving (Chapelle et al., 2009). To address this problem, some approaches have incorporated the assessment of uncertainty and confidence (Yu et al., 2019; Wang et al., 2021a; Wu et al., 2022; Luo et al., 2022; Qiao et al., 2023; V. et al., 2023) into the training process, aiming to generate higher quality pseudo-labels, thereby mitigating the adverse effects caused by low-quality pseudo-labels. Other approaches involve the utilization of teacher-student networks (Tarvainen and Valpola, 2017; Basak and Yin, 2023), typically consisting of two networks with identical architectures. In this setup, the teacher network functions by providing labels to the student network, enabling it to learn while simultaneously evaluating the quality of the pseudo-labels. The student model, on the other hand, is trained using both the ground truth from the labeled set and the pseudo-labels derived from the unlabeled set.

Consistency regularization based: Those methods (Luo et al., 2021a,b; Yu et al., 2019; Qiu et al., 2022; Chaitanya et al., 2020; Ma et al., 2022; Huang et al., 2023) rely on the assumption that the predictions from the model should remain consistent across different perturbations (e.g. data augmentation or feature perturbation). The desired result is that the network maintains consistency in its output irrespective of perturbation applied to the input image or features. To achieve this, the mean square error or Kullback–Leibler divergence between outputs obtained after applying different perturbations is commonly employed to minimize the distribution of output labels, ensuring a consistent prediction across different perturbations. For example, Yu et al. (2019) introduces a framework called the uncertainty-guided mean teacher (UA-MT) framework, which incorporates transformation consistency to improve overall performance. Li et al. (2020b) and Wang et al. (2021a) further investigate shape constraints by introducing the signed distance map and the signed distance field, respectively. The above methods and Luo et al. (2021a) achieve consistency regularization by designing special auxiliary tasks. Later, Luo et al. (2022) introduces a pyramid-prediction network for lesion segmentation, incorporating uncertainty-rectified pyramid consistency. However, it has been observed that the predictions derived from the shallow layers of the network tend to be coarse and imprecise. Wu et al. (2022) designs a mutual consistency network that leverages unannotated images by promoting consistency between the predictions of three marginally different decoders. Similar works (Ouali et al., 2020; Liu et al., 2022) yield slightly different predictions and further encourage their consistency by carefully designing the model structure. Nguyen-Duc et al. (2023)

proposes cross-adversarial local distribution consistency for the extraction of information from unlabeled data. Chen et al. (2023a) further presents decoupled consistency and begins to notice the importance of uncertainty in pseudo-labels. Yet, it only uses the uncertainty map as a threshold for selecting pseudo-labels and fails to fully exploit the information embedded within the uncertainty map. Yang et al. (2023b) suggests a feature-level perturbation with corresponding consistency in an attempt to improve edge segmentation performance. However, it neglects the importance of controlling the intensity of the perturbation and the ability of the model to learn to focus on edge regions.

Besides, several studies (Peng et al., 2021a; Zhou et al., 2020; Li et al., 2020a; Bai et al., 2023; Wang et al., 2023a) in medical image segmentation have explored different variants and combinations of mean-teacher or virtual adversarial training methods. Entropy minimization (Grandvalet and Bengio, 2004; Rizve et al., 2021; Pham et al., 2021) methods are also proposed with additional regularization to boost semi-supervised learning.

2.2. Boundary in segmentation

Boundary problem is a fundamental component of medical imaging and has a historical background (Cheng et al., 2021; Lee et al., 2020; Marin et al., 2019; Yuan et al., 2020). For example, Tsai et al. (2003) proposes an active contour model to find edge information and perform prostate segmentation in MRI images. Nain et al. (2007) presents an active contour formula integrating shape priors based on spherical wavelets. These methods focused on active contours to facilitate edge segmentation, but the edge segmentation remains a great challenge, as blurry edge regions are difficult to design with handcrafted features. Later, Chu et al. (2020) employs edge detectors to identify discontinuities and assist in segmentation. While this method is simple and efficient, it is primarily applicable to the segmentation of internal discontinuities. Peng et al. (2020) introduces circular convolution for efficient feature learning on edge regions, yet it is susceptible to mis-segmentation caused by grayscale non-uniformity and noise sensitivity. Wang et al. (2022a) introduces a boundary-aware context neural network that captures detailed information around the boundary at each stage, which suffers from the undesired complexity introduced by feature fusion. Tang et al. (2022) uses a 3D ConvNet architecture and improves feature discrimination between points across boundaries by sampling and contrasting their representations using scene contexts at multiple scales. However, this sampling strategy still has the limitation that the representation derived from shallow features (before skip connections) may potentially lose a substantial amount of detailed information, including edge information. Instead of sampling representations from different layers, our method employs representations from the last layer of different decoders, guided by estimated uncertainty. This approach enables our framework to focus on learning discriminative information in edge regions.

2.3. Contrastive learning

In recent years, there has been a noticeable emergence of powerful (dis)similarity learning methodologies that exploit contrastive loss in various computer vision tasks (Wang et al., 2022c; Peng et al., 2021b; Zhao et al., 2021; Wang et al., 2023b). Most of the previous contrastive learning methods for segmentation were mainly used in self-monitored pre-training. These methods aim to create a highly effective feature extractor that can be applied later to downstream tasks. The contrastive representation ensures the similarity of representations within positive pairs, while simultaneously promoting dissimilarity between different negative pairs. In general, positive pairs are formed by applying two distinct transformations (random augmentations) to an image. Cai et al. (2020) denotes that including a substantial number of negative examples is essential to the effectiveness and success of these methods. Zhao et al. (2022) proposes a contrastive learning strategy specifically designed to extract relational characteristics between image-level and patch-level representations. Similarly, Wang et al. (2021b) investigates the cross-image pixel contrast for semantic segmentation.

Other studies (Alonso et al., 2021; Gu et al., 2022; Zhang et al., 2023; You et al., 2022; Ma et al., 2023) have introduced semi-supervised learning frameworks that use unlabeled images and incorporate various adaptations of the contrastive loss setup. Alonso et al. (2021) integrates the pseudo-labels obtained from unlabeled images into both the cross-entropy loss and the contrastive loss. Peng et al. (2021b) proposes a self-paced strategy for contrastive learning that dynamically adjusts the importance of individual samples in the contrastive loss. Zhou et al. (2021) proposes a pixel-wise contrastive loss that primarily focuses on highly confident predicted regions belonging to the same class, exploiting the consistency between predictions of teacher-student networks. Similar to Zhou et al. (2021) with applying teacher-student network, You et al. (2022) incorporates the contrastive loss along with segmentation and additional consistency losses. Their approach specifically focuses on using the contrastive loss to learn object shape information using boundary-aware representations, which are defined based on the predicted signed distance maps generated by the teacher-student networks. Wang et al. (2022b) introduce the uncertainty map into contrastive loss to mitigate the possibility of noise sampling by removing the uncertainty region. However, it also hinders the model from learning some crucial areas, such as edge regions. Wang et al. (2023c) proposes a density-guided contrastive learning approach aimed at moving anchor features located in sparse areas toward cluster centers approximated by high-density positive keys, which improves segmentation performance. Later, Chaitanya et al. (2023) presents a novel self-training strategy based on local contrastive learning, which uses semantic label information derived from pseudo-labels. However, their proposed pixel-level contrastive learning encounters challenges in effectively learning discriminative features without a meticulous selection of positive and negative pairs. Furthermore, their method lacks a pseudo-label refinement strategy, essential for improving the quality of generated pseudo-labels.

To obtain reliable pseudo-labels and optimize the pair selection strategy, we propose weak-to-strong perturbation consistency and edge-aware contrastive learning, guided by the uncertainty, and jointly minimize the consistency and contrastive loss for improving segmentation performance.

3. Method

The objective of this study is to present a comprehensive semi-supervised learning framework that effectively utilizes unlabeled data and learns discriminative pixel-wise representations, especially in edge regions. To achieve it, we introduce a novel weak-to-strong feature-level perturbation strategy with consistency loss and edge-aware contrastive loss. These components enable the model to leverage sufficient semantic information from the unlabeled dataset during the semi-supervised learning process.

3.1. The overall semi-supervised learning framework

In this work, we present a semi-supervised learning framework, as depicted in Fig. 1, consisting of a shared encoder E and a main decoder G_1. These components together form the segmentation network for both labeled and unlabeled training data. The other perturbed decoders G_2 to G_M are also introduced for leveraging extra semantic information from the unlabeled training data. Specifically, the segmentation decoders G_2 to G_M incorporate semantic-level perturbation modules, offering different degrees of feature perturbation. To exert better control over the level of perturbation during the learning process, we adopt a weak-to-strong feature perturbation strategy, applying it on G_2 to G_M separately to generate segmentation predictions for each unlabeled

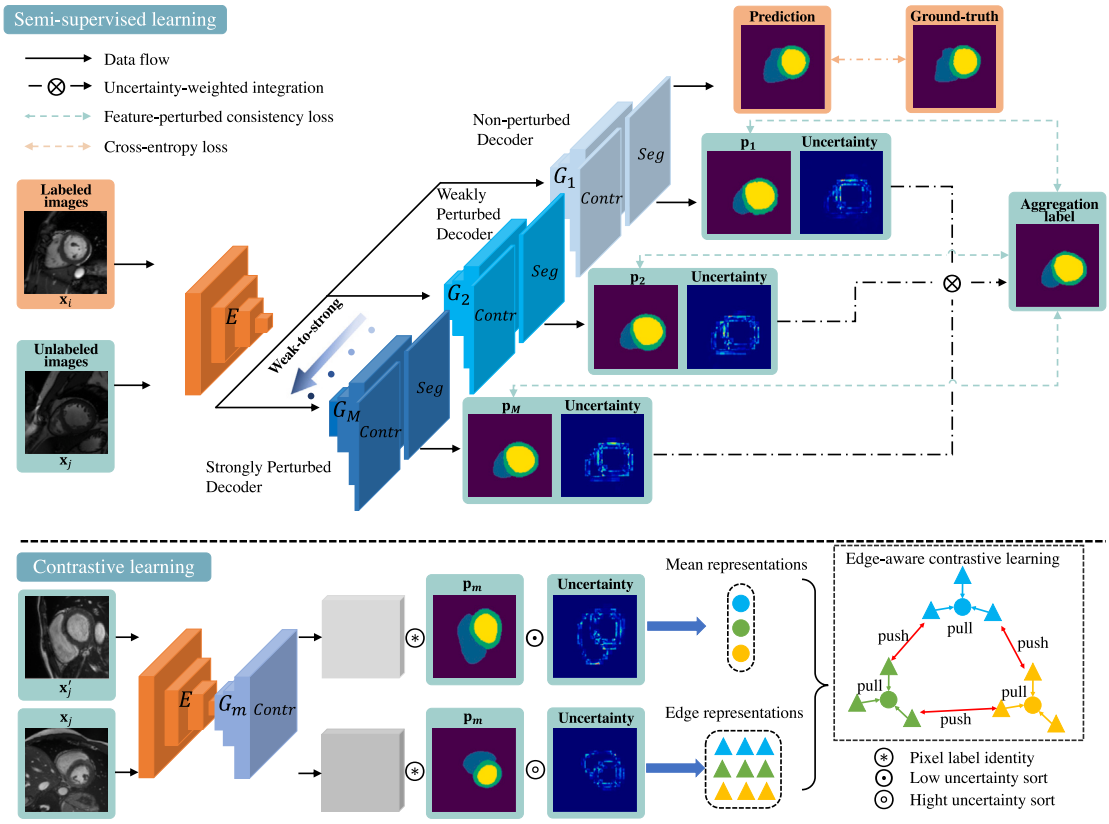

Fig. 1. Overview of our weak-to-strong perturbation and edge-aware semi-supervised segmentation framework, visualized with 2D inputs for improved clarity. In the semi-supervised learning branch ($E \circ G_m$ with *Seg* block), the labeled images $\mathbf{x}_i$ are exclusively fed into the Non-perturbed decoder for better initialization of G_1. The unlabeled images $\mathbf{x}_j$ are processed by weak-to-strong perturbed decoders, including the Non-perturbed decoder, to generate multi-predictions and the aggregation label. In the contrastive learning branch ($E \circ G_m$ with *Contr* block), predictions and uncertainty maps are employed to select discriminative features, facilitating better representation learning. Skip connection between each stage of the encoder and the corresponding stage of each decoder is omitted for simplicity.

input image. Moreover, we propose an innovative feature perturbation consistency loss, which encourages consistency between predictions generated with feature perturbation. Then, an edge-aware contrastive loss is designed to promote more accurate predictions in edge regions. These loss functions collectively aim to train an outperformance segmentation model.

3.2. Weak-to-strong feature-level perturbation

The perturbations applied to either the original images or hidden representations play a crucial role in consistency regularization methods. Image-level perturbation methods, such as FixMatch (Sohn et al., 2020) and UniMatch (Yang et al., 2023a), simultaneously perturb an image from weak to strong by two operators, *i.e.*, weak perturbation such as cropping, and strong perturbation such as color jitter. This weak-to-strong perturbation relies heavily on careful design and requires time-consuming optimization of their combinations and hyper-parameters. Furthermore, the difficulty in finely controlling the intensity of image-level perturbations limits the model's capacity to maintain multi-level consistency against a wider range of perturbations. On the other hand, feature-level perturbation methods have achieved remarkable success by constructing a more comprehensive perturbation hidden space. For example, Ouali et al. (2020) add stochastic noise at feature maps and Yang et al. (2023a) introduce dropout layer into network. However, these methods overlook the inherent domain-specific information, such as the standard deviation and mean of the feature, present in medical images. As the abstraction of features, feature statistics can capture informative characteristics of the corresponding domain (such as color, texture, and contrast), according to previous works (Huang and Belongie, 2017; Li et al., 2021b).

To this end, we propose a novel feature-level perturbation module (Fig. 2), which utilizes the feature statistics information, to introduce meaningful and reasonable perturbations at the feature level. The design of the perturbation module is guided by two fundamental considerations: (1) Different levels of perturbation within the appropriate range facilitate the exploration of the hidden space of the model, making efficient use of the unlabeled data. (2) The segmentation results obtained from unperturbed features are of substantial value. Hence, we design multiple independent feedforward streams that contain different levels of perturbation. This design allows the model to achieve targeted consistency in each stream more directly. Specifically, for each unlabeled training data $\mathbf{x}_j$, we consider the cth channel of the feature map output from a predetermined layer in the decoder G_m, denoted as $\mathbf{f}_{j,c}$. We calculate the mean $\mu_{j,c}$ and standard deviation $\sigma_{j,c}$ of all elements in $\mathbf{f}_{j,c}$. Then, $\mathbf{f}_{j,c}$ is randomly perturbed using a linear transformation of its normalized version $\bar{\mathbf{f}}_{j,c} = \frac{\mathbf{f}_{j,c} - \mu_{j,c}}{\sigma_{j,c}}$ as below,

$$\hat{\mathbf{f}}_{j,c} = \boldsymbol{\gamma}_{j,c} \cdot \bar{\mathbf{f}}_{j,c} + \boldsymbol{\beta}_{j,c} \tag{1}$$

$$\boldsymbol{\gamma}_{j,c} = \lambda \cdot \sigma_{j,c} + (1-\lambda) \cdot \sigma_c + \epsilon \cdot \phi_c \tag{2}$$

$$\boldsymbol{\beta}_{j,c} = \lambda \cdot \mu_{j,c} + (1-\lambda) \cdot \mu_c + \epsilon \cdot \psi_c, \tag{3}$$

where (σ_c, ϕ_c) are the estimated mean and standard deviation of $\{\sigma_{j,c}\}_{j=1}^{B}$, and (μ_c, ψ_c) the estimated mean and standard deviation of $\{\mu_{j,c}\}_{j=1}^{B}$ over a mini-batch of B unlabeled training images, including $\mathbf{x}_j$, during model training. $\boldsymbol{\gamma}_{j,c}$ and $\boldsymbol{\beta}_{j,c}$ are slightly perturbed version of the standard deviation $\sigma_{j,c}$ and the mean $\mu_{j,c}$, respectively. The distribution information (σ_c, ϕ_c, μ_c and ψ_c) ensure that any perturbed feature $\hat{\mathbf{f}}_{j,c}$ remains semantically meaningful. The feature-level perturbation is jointly determined by λ and ϵ. While parameter λ determines the proportion of retaining the feature's own semantic information, a scale

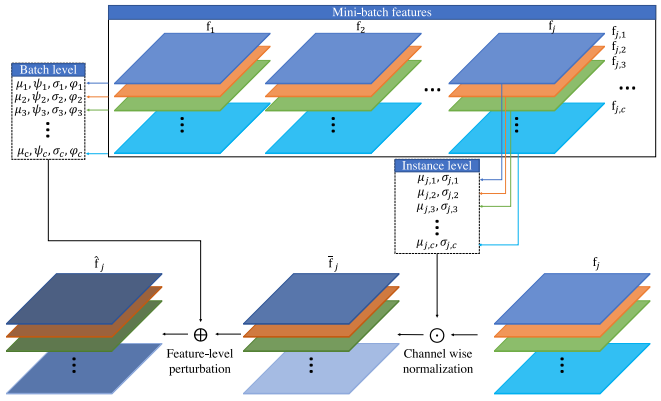

Fig. 2. The feature-level perturbation module incorporated in decoders. Means and standard deviations at both the instance and batch levels are applied to each feature map at convolutional layers for each image within a mini-batch.

variable ϵ is randomly sampled from a uniform distribution within the range of $[-\kappa, \kappa]$ to regulate the intensity of the perturbation. Therefore, ϕ_c and ψ_c play a crucial role in controlling the perturbations within a reasonable range. It is worth mentioning that the scale variable ϵ is shared across all feature channels ($\{c\}$), ensuring that all channels of the feature maps receive the same level of perturbation. The perturbed feature maps are then fed into the subsequent convolutional layer(s), which generate the segmentation probability output for the specific feature-level perturbation of the input $\mathbf{x}_j$. This perturbation is incorporated into each layer of the decoders, allowing our model to explore a wider range of hidden space. Compared to the method of Yang et al. (2023b), which employs a single decoder for segmentation, our method introduces multi-decoder architecture to enable weak-to-strong feature perturbation, leading to a wider exploration of feature space. Unlike (Yang et al., 2023b)'s method, which applies varying perturbation intensities in the final layer, our approach integrates consistent intensity perturbations, controlled by ϵ, across each layer. Additionally, our method further introduces the batch-level mean σ_c of $\{\sigma_{j,c}\}_{j=1}^{B}$ and μ_c of $\{\mu_{j,c}\}_{j=1}^{B}$ for a more comprehensive perturbation.

3.3. Feature-perturbed consistency for semi-supervised learning

In semi-supervised learning, a commonly employed strategy is to design a consistency loss on unlabeled training data. The underlying concept is to encourage the model to generate similar outputs when presented with two or more transformed versions of the input. In our approach, the segmentation output probabilities predicted by the weak-to-strong perturbed features provide a natural foundation for designing the consistency loss. In detail, let $D_u = \{\mathbf{x}_j, j = 1, \dots, J\}$ denote the unlabeled training set, which consists of J unlabeled images. Each image contains K elements (pixels or voxels). For the jth training unlabeled data $\mathbf{x}_j \in D_u$, we perform M independent perturbations on different decoders as described above, resulting in M output probability maps from the segmentation decoders G_1 to G_M. Let $\{\mathbf{p}_{j,k,m}, m = 1, \dots, M\}$ denote the M output probability vectors corresponding to the kth element of the input $\mathbf{x}_j$. Given that there are no annotations available for any image in D_u, and the potential bias introduced by the perturbations in the decoders, the results $\mathbf{p}_{j,k,m}$ obtained by the model may be unreliable. Averaging m predictions to produce the final prediction, as in previous works (Liu et al., 2022; Luo et al., 2022), is insufficient to eliminate errors in the predictions. Therefore, we employ uncertainty estimation and an uncertainty-weighted integration strategy in our framework. Specifically, the uncertainty of each pixel for the M preliminary predictions can be calculated by $\mathcal{H}(\mathbf{p}_{j,k,m})$, which represents the entropy of the discrete probability $\mathbf{p}_{j,k,m}$. Naturally, pixels with low uncertainty should have a greater contribution to the final prediction result, which can be defined as a weight map, $\boldsymbol{\alpha}_{j,k,m} = e^{-\mathcal{H}(\mathbf{p}_{j,k,m})} / \sum_{i=1}^{M} e^{-\mathcal{H}(\mathbf{p}_{j,k,i})}$, to highlight areas of higher confidence during the aggregation process. Consequently, the aggregated prediction of segmentation $\bar{\mathbf{p}}_{j,k}$ can be formulated as $\bar{\mathbf{p}}_{j,k} = \sum_{m=1}^{M} \boldsymbol{\alpha}_{j,k,m} \cdot \mathbf{p}_{j,k,m}$. The uncertainty of the aggregated prediction can be calculated by $\mathcal{H}(\bar{\mathbf{p}}_{j,k})$. Note that our network architecture incorporates multiple decoders, simultaneously performing several times uncertainty estimates for a pixel. This design reduces the potential negative impact of erroneous uncertainty estimates from single branches. Therefore, the aggregation approach helps mitigate the unreliability caused by the lack of annotations and the introduced perturbations, resulting in more robust and accurate segmentation results.

Then the consistency loss based on feature perturbations can be represented as,

$$\mathcal{L}_u = \frac{1}{J \cdot K} \sum_{j=1}^{J} \sum_{k=1}^{K} \mathcal{G}(\mathbf{p}_{j,k,1}, \dots, \mathbf{p}_{j,k,M}), \tag{4}$$

where the consistency assessment $\mathcal{G}(\cdot)$ can be any appropriate function that quantifies the similarity or consistency among all the M outputs $\{\mathbf{p}_{j,k,m}, m = 1, \dots, M\}$ for each element in the image. In our design, we take inspiration from recent research (Luo et al., 2022) and develop a consistency measurement function as,

$$\mathcal{G}(\mathbf{p}_{j,k,1}, \dots, \mathbf{p}_{j,k,M}) = \frac{1}{M} \sum_{m=1}^{M} \frac{\boldsymbol{\omega}_{j,k,m} \|\mathbf{p}_{j,k,m} - \bar{\mathbf{p}}_{j,k}\|}{\sum_{m=1}^{M} \boldsymbol{\omega}_{j,k,m}}, \tag{5}$$

where $\boldsymbol{\omega}_{j,k,m} = \exp\{-\mathcal{H}(\mathbf{p}_{j,k,m})\}$. The term $\|\mathbf{p}_{j,k,m} - \bar{\mathbf{p}}_{j,k}\|$ represents the L_p norm ($p = 1$ in this study) of the difference between a single output prediction $\mathbf{p}_{j,k,m}$ and the aggregated prediction $\bar{\mathbf{p}}_{j,k}$. Minimization of this term would encourage the predictions generated from all the M perturbed features to be similar. Besides, entropy serves as an indicator of prediction uncertainty. A higher entropy value corresponds to a lower weight $\omega_{j,k,n}$. This weighting mechanism directs the consistency measurement function to focus more on confident predictions rather than unconfident (i.e., uncertainty) ones, ensuring that the confident predictions are consistent. This is justified since uncertain predictions often occur near the edge region.

3.4. Edge-aware contrastive learning

Despite the impressive performance of many semi-supervised learning methods in segmentation tasks, they often share a common limitation. These methods are adept at identifying the main body of the target object for segmentation, but they may lack the sensitivity to accurately detect edge areas or small foreground regions. To tackle this issue, we propose an edge-aware contrastive learning loss. Our objective is to learn discriminative pixel representations based on their predictions and uncertainty. As illustrated in Fig. 1, we introduce two branches at the end of each decoder: one dedicated to the segmentation task and the other focused on contrastive learning. This design allows us to leverage the learned representations in the feature space to optimize the model for accurate segmentation of edge regions. The proposed edge-aware contrastive learning consists of two components:

Contrastive Learning: Once an image $\mathbf{x}_j$ is processed through the common network $E \circ G_m$ with the contrastive branch, the resulting feature map is denoted as $\mathbf{v}(\mathbf{x}_j)$. This feature map has dimensions of $H \times W \times D$, where H and W correspond to the dimensions of the input image, and D represents the number of channels in the feature map. We can define the set of pixel indexes that belong to a foreground class c for image $\mathbf{x}_j$ as $S_c(\mathbf{x}_j)$, where $S_c(\mathbf{x}_j)$ represents the collection of pixel indexes that are predicted to the cth class for unlabeled samples. Here, $S_c(\mathbf{x}_j)$ is defined for $1 \leqslant c \leqslant C$, where C denotes the total number of classes for segmentation. For two randomly sampled samples from D_u, $\mathbf{x}_j$ and $\mathbf{x}'_j$, the contrast loss function can be defined as,

$$\mathcal{L}_r(\mathbf{x}_j, \mathbf{x}'_j) = \frac{1}{J \cdot C} \sum_{j=1}^{J} \sum_{c=1}^{C} \frac{1}{|S_c(\mathbf{x}_j)|} \sum_{i \in S_c(\mathbf{x}_j)} \mathcal{R}(\mathbf{v}_i(\mathbf{x}_j), \bar{\mathbf{v}}^c(\mathbf{x}'_j)), \tag{6}$$

where $\mathbf{v}_i(\mathbf{x}_j)$ presents the feature vector of $\mathbf{x}_j$ at pixel index i, which is D-dimensional, and $\bar{\mathbf{v}}^c(\mathbf{x}'_j)$ denotes the mean pixel representation of $\mathbf{x}'_j$ for class c, it is formulated as,

$$\bar{v}^c(\mathbf{x}'_j) = \frac{1}{|S_c(\mathbf{x}'_j)|} \sum_{i \in S_c(\mathbf{x}'_j)} \mathbf{v}_i(\mathbf{x}'_j). \tag{7}$$

Then the contrastive loss, denoted as $\mathcal{R}(\cdot,\cdot)$, between a pixel representation feature vector and a mean pixel representation belonging to a potentially different image, is defined for a given class c as,

$$\mathcal{R}(\mathbf{v}_i(\mathbf{x}_j), \bar{\mathbf{v}}^c(\mathbf{x}'_j)) = -\log \frac{e^{sim(\mathbf{v}_i(\mathbf{x}_j), \bar{\mathbf{v}}^c(\mathbf{x}'_j))/\tau}}{\sum_{n \in C} e^{sim(\mathbf{v}_i(\mathbf{x}_j), \bar{\mathbf{v}}^n(\mathbf{x}'_j))/\tau}}, \tag{8}$$

where $sim(\mathbf{r},\mathbf{s}) = \mathbf{r}^T\mathbf{s}/\|\mathbf{r}\|\|\mathbf{s}\|$ represents the cosine similarity, which measures the similarity between two representation vectors $\mathbf{r}$ and $\mathbf{s}$. The temperature scaling factor, denoted as τ, is used to adjust the scale of the similarity measurement.

Edge-aware Pixel Selection: In the proposed pixel-wise contrastive loss Eq. (6), each pixel representation $\mathbf{v}_i(\mathbf{x}_j)$ of image $\mathbf{x}_j$ at pixel location i is designed to match the mean representation vector $\bar{\mathbf{v}}^n(\mathbf{x}'_j), n \in C$. Subsequently, we pull similar representations together for pixels where $\mathbf{v}_i(\mathbf{x}_j)$ has the same label with $\bar{\mathbf{v}}^n(\mathbf{x}'_j)$, while simultaneously pushing $\mathbf{v}_i(\mathbf{x}_j)$ away from $\bar{\mathbf{v}}^n(\mathbf{x}'_j)$ if the corresponding pixel vectors and the mean representation belong to different classes in Eq. (8). Indeed, efficient representation selection strategy is the key to contrastive learning. Randomly sampling all pixel representations from $\mathbf{x}_j$ and $\mathbf{x}'_j$ to calculate the contrast loss poses a formidable challenge in terms of computational memory. Moreover, this approach does not facilitate the learning of discriminative representations, especially edge regions. Wang et al. (2022b) attempted to reduce the possibility of noise sampling by removing high uncertainty regions in the uncertainty map to ensure the effectiveness of contrastive learning. However, they overlook the fact that regions with high uncertainty typically present at the boundaries, which are crucial for capturing significant discriminative representations. Therefore, we design a more efficient pixel selection strategy named edge-aware pixel selection. Specifically, we sort each feature vector $\mathbf{v}_i(\mathbf{x}_j)$ and $\mathbf{v}_i(\mathbf{x}'_j)$ based on its corresponding uncertainty values obtained from the segmentation branch. For the pixels in $S_c(\mathbf{x}_j)$, we initially discard a certain percentage of pixels with the highest uncertainty, as these pixels, typically presented in edge regions, may be subject to misclassification by the decoders. Then, a subset of pixels with high uncertainty, typically located at the edges, are selected to form $S_c(\mathbf{x}_j)$. With respect to the pixels in $S_c(\mathbf{x}'_j)$, we first select those with low uncertainty, such as the top sixty percent, and then randomly select the pixels from the subset to construct $S_c(\mathbf{x}'_j)$ for the purpose of computing the mean representations. This selective approach enables the model to mitigate computational memory requirements while promoting our framework to learn discriminative pixel representations, particularly in edge regions. By prioritizing pixels with high uncertainty and incorporating them into $S_c(\mathbf{x}_j)$, our model acquires powerful representation ability in challenging regions. This allows our framework to improve its ability to accurately segment edge regions, which is crucial for achieving accurate and reliable segmentation results.

Finally, the total semi-supervised learning loss function can then be designed as,

$$\mathcal{L} = \mathcal{L}_s + \lambda_1 \mathcal{L}_u + \lambda_2 \mathcal{L}_r, \tag{9}$$

where $\mathcal{L}_s$ is supervised loss (e.g., cross-entropy loss) only on the labeled training set, $\mathcal{L}_u$ and $\mathcal{L}_l$ are the feature-perturbed consistency loss and the edge-aware contrastive loss, respectively, on the unlabeled sets. λ_1 and λ_2 are two coefficients to balance the three loss terms.

4. Experiment

4.1. Datasets and evaluation

In this study, we performed evaluations of our method and performed comparisons with several related works on three publicly available datasets: the Automated Cardiac Diagnosis Challenge segmentation dataset (2017 ACDC), the whole brain tumor segmentation dataset (BraTS2020), and the Left Atrium dataset from Atrial Segmentation Challenge (LA). It is important to note that the ACDC dataset consists of 2D slices, while the BraTS and LA datasets consist of 3D volumes. We referred to and followed image preprocessing as in previous works (Chaitanya et al., 2019; Chen et al., 2020; Hooper et al., 2020), where all images were bias corrected using N4 (Tustison et al., 2010) algorithm. Random affine transformation (i.e., scaling, rotation, translation) and random global intensity transformation (brightness and contrast) were also applied.

2017 ACDC dataset: The ACDC (Bernard et al., 2018) segmentation dataset consists of 100 short-axis MR-cine T1 3D volumes of cardiac anatomy. These volumes were acquired using both 1.5T and 3T scanners, with a spatial resolution ranging from 1.37 to 1.68 mm^2/pixel. This dataset comprises five categories of images, i.e., Normal (NOR), previous myocardial infarction (MINF), dilated cardiomyopathy (DCM), hypertrophic cardiomyopathy (HCM), and abnormal right ventricle (RV). Each image in the dataset requires segmentation of three categories: the right ventricle (RV), left ventricle (LV) cavities, and the myocardium (specifically the epicardial contour). We randomly select 75, 5 and 20 subjects for training, validation and testing respectively. The five categorizations (NOR, MINF, DCM, HCM, RV) are guaranteed to be included when constituting the labeled dataset and validation set. For pre-processing, the intensity of each scan is rescaled to $[0, 1]$.

BraTS2020 dataset: The BraTS2020 (Menze et al., 2014) dataset contains 496 subjects, each containing four modalities (FLAIR, T1, T1ce and T2) with an isotropic 1 mm^3 resolution. In this study, the FLAIR modality is adopted for semi-supervised segmentation of the whole tumors. The dataset was randomly divided into training, validation, and test sets, consisting of 380, 26, and 90 scans, respectively. As part of the preprocessing step, each instance was normalized using its channel-wise means and standard deviations, followed by intensity rescaling to the range of $[0, 1]$.

LA dataset: The dataset (Xiong et al., 2021) comprises 100 3D gadolinium-enhanced MR imaging (GE-MRI) scans along with left atrium segmentation masks. The scans have an isotropic resolution of $0.625 \times 0.625 \times 0.625$ mm^3. We divided the dataset of 100 scans into 80 scans for training and 20 scans for evaluation. Furthermore, to focus on the heart region, all scans were cropped, centering around this area. Additionally, we applied normalization to ensure that the data had a zero mean and unit variance.

Baselines: To demonstrate the superior performance of our proposed method in semi-supervised learning, we performed a comprehensive comparison of lesion and organ segmentation performance between our method and several state-of-the-art approaches. The evaluated methods included nnU-Net (Isensee et al., 2021), SASSnet (Li et al., 2020b), UAMT (Yu et al., 2019), Tri-U-MT (Wang et al., 2021a), DTC (Luo et al., 2021a), CoraNet (Shi et al., 2021), SPCL (Peng et al., 2021b), MC-Net+ (Wu et al., 2022), URPC (Luo et al., 2022), PLCT (Chaitanya et al., 2023), DGCL (Wang et al., 2023c), CAML (Gao et al., 2023), DCNet (Chen et al., 2023a) and SFPC (Yang et al., 2023b), where PLCT and DGCL are based on contrastive learning. Among these methods, only nnU-Net was trained in a fully supervised manner, serving as a performance upper bound. Following previous works (Wang et al., 2021a; Luo et al., 2022; Yu et al., 2019), we employ the top-performing model identified through validation set evaluation for inference on the ACDC and BraTS2020 datasets, while opting for the model from the final epoch for inference on the LA dataset. To ensure the reliability and consistency of our experimental results, we

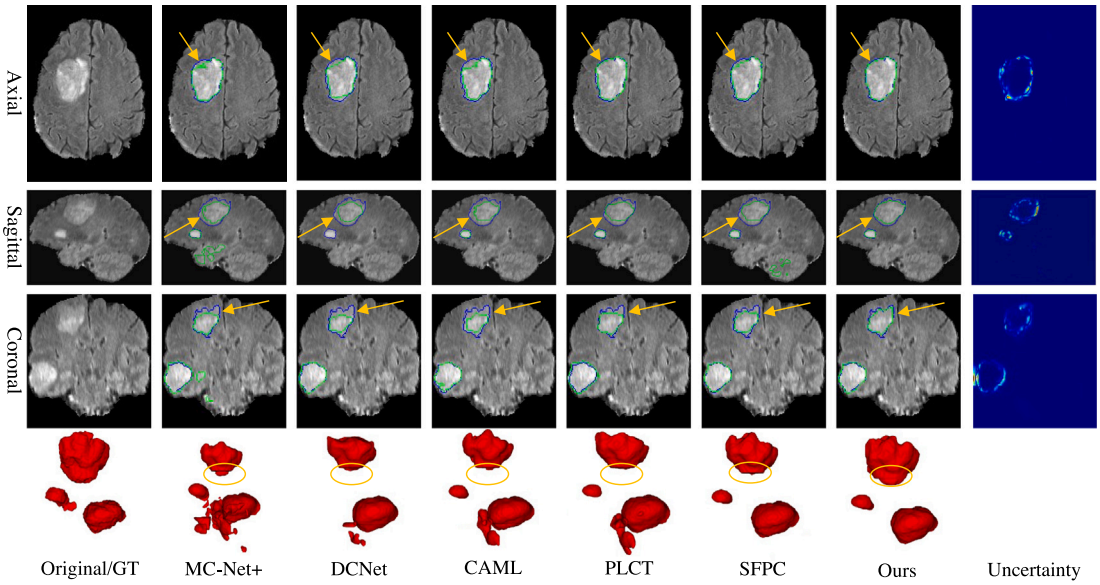

Fig. 3. Visual comparisons between the proposed method and strong baseline methods (second to sixth column) on one image from BraTS2020 dataset. During training, 5% of the training samples were annotated. Green and blue contours denote prediction and ground-truth edges, respectively. Fourth row (in red): view of the 3D segmentation lesions. Last column: pixel-wise uncertainty of the aggregation prediction using entropy.

conducted all experiments with five different cross-validations and recorded the average results, standard deviation, and p-value.

Evaluation metrics: To quantitatively evaluate the performance of our framework, we employed four commonly used evaluation metrics: the Dice Similarity Coefficient (DSC), the Jaccard Index (Jaccard), the 95% Hausdorff Distance (95HD), and the Average Surface Distance (ASD).

4.2. Implementation details

We re-implemented all compared methods and conducted the experiments under an identical environment (Hardware: Intel(R) Xeon(R) Gold 5220R CPU@2.20 GHz, NVIDIA GeForce RTX 3090 GPU; Software: PyTorch 1.13.0, CUDA 12.2, and Python 3.8.0). The backbone segmentation network is UNet (Ronneberger et al., 2015) and 3D-UNet (Çiçek et al., 2016) respectively. The contrastive branch is composed of two 1×1 convolution layers that generate feature maps with dimensions $H \times W \times D$, where D is set to 16. The network was optimized using the SGD optimizer with a weight decay of 1e-4 and momentum of 0.9. Training was performed for 6000 iterations, starting with an initial learning rate of 0.01 that decayed by a factor of 0.1 every 2500 iterations. The batch size was set to 8, consisting of 4 labeled images and 4 unlabeled images. For 3D volumes, we randomly cropped sub-volumes of size $112 \times 112 \times 112$ as input to the network. For 2D slices, we resized the input to 256×256. Data augmentations, including random cropping, flipping, and rotation, were employed to prevent overfitting. To obtain the final segmentation results for 3D volumes, we utilized a sliding window strategy. For subtle perturbations at the semantic level, we set the perturbation parameter κ to 0.2, and λ to 0.9. The weight λ_1 was determined by a time-dependent Gaussian warming up function, as elaborately designed in previous studies (Tarvainen and Valpola, 2017; Yu et al., 2019). It balanced the weight between supervised and unsupervised learning in a stable manner. The function was defined as $\lambda(t) = \omega_{max} \cdot e^{-5(1-\frac{t}{t_{max}})^2}$, where ω_{max} represents the final regularization weight, t is the current training round and t_{max} denotes the maximum training round. Based on the relevant study (Yu et al., 2019), we set ω_{max} to 0.1 for all experiments. In the construction of $S_c(\mathbf{x}_j)$, 3% of pixels with the highest uncertainty were initially discard. The coefficient λ_2 was empirically set to 0.5 to achieve the desired balance between supervised and unsupervised learning.

4.3. Performance on the BraTS2020 dataset

Quantitative results obtained on the BraTS2020 dataset using different proportions of labeled samples in the training set are presented in Table 1 (left half). Our proposed method consistently outperforms all the compared semi-supervised methods, achieving the highest DSC values of 81.82%, 85.98%, 87.79% and the highest Jaccard values of 70.36%, 75.63%, 78.40% in three different scenarios. Moreover, our method shows the lowest 95HD values of 12.19, 9.97, 8.18, and the lowest ASD values of 3.47, 2.64, 2.01, respectively. Compared to the strongest baseline method, SFPC, with only 5% labeled samples, our proposed method achieves absolute improvements of 1.06% in DSC, 1.18% in Jaccard, 2.68 in 95HD, and 0.55 in ASD. Furthermore, when the proportion of labeled data is increased to 20%, our model achieves comparable results to the nn-UNet model trained with 100% labeled data, with a DSC of 87.79% compared to the upper-bound model's score of 89.58%. As depicted in Fig. 3, our approach (seventh column) presents superior accuracy in identifying edge regions (indicated by yellow arrows and ellipses in the 2D and 3D views, respectively), outperforming other baseline methods (second to sixth column) on the BraTS2020 dataset. The uncertainty in the pixel-level predictions obtained from our method (last column) effectively highlights the challenging areas for segmentation, revealing that the uncertain regions are located mainly along the edge of the lesions.

4.4. Performance on the LA dataset

Similar results were obtained on the LA dataset. As shown in Table 1 (middle half), our method outperforms all strong semi-supervised baselines on all four evaluation metrics. In particular, when only 5% of the training images are annotated, our method has a clear superiority. While PLCT achieves the best performance among the existing methods with a DSC of 87.63% and a Jaccard index of 78.69%, our method outperforms it with a gain of 1.39% in DSC and 1.14% in Jaccard. In addition, our method slightly outperforms all existing methods when 10% and 20% of the training images are annotated. Furthermore, our method demonstrates even greater superiority when trained with a smaller number of labeled samples, highlighting its capability to effectively leverage unlabeled scans for performance improvement. Fig. 4 provides a visual comparison based on 5% labeled data, illustrating two

Table 1
Quantitative comparisons with other state-of-the-art methods on the BraTS2020, LA and 2017 ACDC datasets. ↑ indicates that larger values are better and ↓ indicates that smaller values are better.

Method	% scans used		BraTS2020 (3D)				LA (3D)				2017 ACDC (2D)			
	Labeled	Unlabeled	DSC (%) ↑	Jaccard (%) ↑	95HD (mm) ↓	ASD (mm) ↓	DSC (%) ↑	Jaccard (%) ↑	95HD (mm) ↓	ASD (mm) ↓	DSC (%) ↑	Jaccard (%) ↑	95HD (mm) ↓	ASD (mm) ↓
UAMT (Yu et al., 2019)	5	95	49.46± 2.51*	38.46± 1.86*	19.57± 3.28*	6.54± 0.86*	78.69± 1.26*	65.56± 1.31*	27.69± 2.84*	8.04± 1.29*	51.23± 1.96*	41.82± 1.62*	17.13± 2.82*	7.76± 2.01*
SASSNet (Li et al., 2020b)			51.82± 1.74*	43.93± 1.42*	23.47± 2.83*	7.47± 1.09*	80.04± 1.12*	67.36± 1.03*	25.29± 3.16*	7.05± 1.32*	58.47± 1.74*	47.04± 2.02*	18.04± 3.63*	7.31± 1.53*
Tri-U-MT (Wang et al., 2021a)			53.95± 1.97*	44.33± 2.18*	19.68± 3.06*	7.29± 0.84*	80.47± 1.19*	67.07± 1.07*	23.07± 3.03*	7.62± 1.02*	59.15± 2.01*	47.37± 1.82*	17.37± 2.77*	7.34± 1.31*
DTC (Luo et al., 2021a)			56.72± 2.04*	45.78± 1.67*	17.38± 4.31*	6.28± 1.22*	80.68± 1.06*	68.35± 1.64*	22.97± 2.57*	7.07± 0.71*	57.09± 1.57*	45.61± 1.23*	20.63± 2.61*	7.05± 1.94*
CoraNet (Shi et al., 2021)			57.97± 1.83*	46.40± 1.64*	19.52± 2.80*	5.83± 0.85*	80.94± 1.35*	68.42± 1.26*	19.54± 2.10*	5.09± 1.32*	59.91± 2.08*	48.37± 1.75*	15.53± 2.23*	5.96± 1.42*
SPCL (Peng et al., 2021b)			78.73± 1.54*	67.90± 1.29*	16.26± 1.68*	4.47± 1.08*	87.36± 0.95*	78.21± 0.91*	13.06± 1.14*	3.37± 0.96*	81.82± 1.24	70.62± 1.04	5.96± 1.62	2.21± 0.29
MC-Net+ (Wu et al., 2022)			58.91± 1.47*	47.24± 1.36*	20.82± 3.35*	7.14± 1.12*	83.95± 1.64*	72.45± 1.47*	14.99± 2.28*	3.24± 1.36*	63.47± 1.75*	53.13± 1.41*	7.38± 1.68*	2.37± 0.32*
URPC (Luo et al., 2022)			60.48± 2.01*	50.69± 1.99*	18.21± 3.27*	7.12± 0.95*	82.67± 0.83*	70.66± 0.77*	16.37± 1.24*	3.80± 0.75*	62.57± 1.18*	52.75± 1.36*	7.79± 1.85*	2.64± 0.36*
PLCT (Chaitanya et al., 2023)			65.74± 2.17*	55.40± 1.85*	16.61± 3.04*	6.85± 1.39*	87.63± 0.71*	78.69± 0.62*	12.28± 0.89*	2.62± 0.77*	78.42± 1.45*	67.43± 1.25*	6.54± 1.62*	2.48± 0.24*
DGCL (Wang et al., 2023c)			80.21± 0.75*	68.86± 0.63*	14.91± 1.53*	4.63± 1.16*	87.47± 0.97*	78.37± 0.87*	12.64± 1.14*	2.71± 0.93*	80.57± 1.12*	68.74± 0.96*	6.04± 1.73	2.17± 0.30
CAML (Gao et al., 2023)			77.86± 0.96*	66.42± 1.37*	15.21± 1.74*	5.10± 1.12*	87.42± 0.86*	78.30± 0.78*	12.63± 0.97*	3.47± 0.72*	79.04± 0.83*	68.45± 0.97*	6.28± 1.79*	2.24± 0.26
DCNet (Chen et al., 2023a)			78.52± 1.21*	67.81± 1.07*	17.37± 1.48*	4.32± 0.96*	86.56± 0.74*	75.55± 0.75*	11.44± 0.91*	2.54± 0.83*	71.57± 1.58*	61.12± 1.19*	8.37± 1.92*	4.08± 0.84*
SFPC (Yang et al., 2023b)			80.76± 0.74	69.18± 0.83	14.87± 1.92*	4.02± 0.75*	86.81± 0.65*	78.49± 0.63*	11.70± 0.74*	2.88± 0.69*	80.52± 1.03*	68.73± 0.88*	6.08± 1.47	2.14± 0.22
Ours			**81.82± 0.77**	**70.36± 0.71**	**12.19± 1.31**	**3.47± 0.51**	**89.02± 0.51**	**79.83± 0.57**	**10.23± 0.71**	**2.18± 0.55**	**82.39± 0.82**	**71.16± 0.75**	**5.67± 1.21**	**2.01± 0.17**
UAMT (Yu et al., 2019)	10	90	81.04± 1.46*	68.88± 1.57*	17.27± 3.35*	6.25± 1.63*	85.69± 0.68*	75.42± 0.57*	16.24± 2.23*	4.45± 0.47*	81.86± 1.25*	71.07± 1.43*	12.92± 1.68*	3.49± 0.64*
SASSNet (Li et al., 2020b)			82.36± 2.08*	71.03± 2.35*	14.80± 3.72*	4.11± 1.54*	85.73± 0.72*	75.78± 0.69*	14.24± 2.58*	4.62± 0.93*	84.61± 1.97*	74.53± 1.78*	6.02± 1.54*	1.71± 0.35
Tri-U-MT (Wang et al., 2021a)			82.83± 1.35*	71.52± 1.21*	15.19± 2.86*	3.57± 1.30	85.68± 0.83*	75.74± 0.75*	14.07± 0.62*	4.29± 0.27*	84.06± 1.69*	74.32± 1.77*	7.41± 1.63*	2.59± 0.51*
DTC (Luo et al., 2021a)			81.98± 2.41*	70.41± 2.73*	16.27± 3.62*	3.62± 1.71	84.86± 1.37*	74.19± 1.14*	13.25± 0.56*	3.28± 0.44*	82.91± 1.65*	71.61± 1.81*	8.69± 1.84*	3.04± 0.59*
CoraNet (Shi et al., 2021)			81.38± 1.68*	70.01± 1.83*	13.94± 2.72*	3.95± 1.26*	83.60± 1.73*	72.14± 1.26*	17.06± 1.58*	4.07± 0.52*	84.56± 1.53*	74.41± 1.49*	6.11± 1.15*	2.35± 0.44*
SPCL (Peng et al., 2021b)			84.65± 1.16	73.91± 1.19*	12.24± 1.47*	3.28± 0.42	89.17± 0.79	80.93± 0.64	7.68± 0.81	2.51± 0.33	87.57± 1.15	78.63± 0.89	4.87± 0.79	1.31± 0.27
MC-Net+ (Wu et al., 2022)			83.93± 1.73*	72.34± 1.69*	13.52± 2.74*	3.37± 1.13	87.83± 1.31*	78.45± 1.46*	10.49± 0.94*	2.78± 0.53*	86.78± 1.41*	77.31± 1.27*	6.92± 0.95*	2.04± 0.37*
URPC (Luo et al., 2022)			84.23± 1.41	72.37± 1.26*	11.52± 1.79	3.26± 1.14	84.52± 0.29*	72.63± 0.32*	11.26± 0.69*	2.99± 0.36*	85.18± 0.98*	74.65± 0.83*	5.01± 0.79	1.52± 0.26
PLCT (Chaitanya et al., 2023)			83.66± 1.82*	71.99± 1.67*	13.68± 1.29*	3.59± 1.02	89.41± 0.63	81.00± 0.85	7.34± 0.72	2.68± 0.24*	86.83± 1.17*	77.04± 0.83*	6.62± 0.86*	2.27± 0.42*
DGCL (Wang et al., 2023c)			84.02± 1.24*	72.16± 1.07*	12.98± 1.28*	3.02± 0.96	89.68± 0.64	81.37± 0.59	7.91± 0.68	2.46± 0.31	87.74± 1.06	78.82± 1.22	4.74± 0.73	1.56± 0.24
CAML (Gao et al., 2023)			84.34± 1.03	73.84± 0.92*	12.02± 1.84*	3.31± 0.58	89.53± 0.62	81.04± 0.71	9.72± 0.84*	2.64± 0.25	87.67± 0.83	78.70± 0.91	4.97± 0.62	1.35± 0.17
DCNet (Chen et al., 2023a)			83.39± 0.97*	71.94± 0.88*	11.93± 1.24*	3.50± 0.33	87.96± 0.76*	78.66± 0.69*	8.84± 0.78*	2.99± 0.29*	87.81± 0.88	78.96± 0.94	4.84± 0.81	1.23± 0.21
SFPC (Yang et al., 2023b)			85.01± 0.89	74.67± 1.14*	10.73± 1.36	3.03± 0.31	89.59± 0.57	81.26± 0.61	7.56± 0.69	2.81± 0.32*	87.76± 0.92	78.94± 0.83	4.90± 0.74	1.28± 0.23
Ours			**85.98± 0.75**	**75.63± 0.62**	**9.97± 1.22**	**2.64± 0.27**	**90.23± 0.53**	**81.52± 0.51**	**7.16± 0.48**	**1.95± 0.15**	**88.92± 0.64**	**79.65± 0.55**	**4.32± 0.47**	**1.19± 0.18**
UAMT (Yu et al., 2019)	20	80	84.95± 1.32*	74.71± 1.25*	12.18± 2.89*	2.45± 0.31	88.36± 0.41*	79.23± 0.83*	9.76± 1.53*	3.07± 0.69*	85.96± 0.97*	76.96± 0.83*	9.17± 1.25*	1.42± 0.22
SASSNet (Li et al., 2020b)			84.74± 2.23*	74.04± 2.17*	9.38± 2.57	2.61± 0.44	88.31± 0.59*	79.25± 0.57*	9.47± 1.74*	3.48± 1.18*	87.18± 1.27*	77.51± 1.44*	5.28± 0.86*	2.47± 0.45*
Tri-U-MT (Wang et al., 2021a)			85.11± 1.48*	74.71± 1.29*	8.80± 2.37	3.11± 0.50*	88.19± 0.67*	79.12± 0.53*	8.35± 0.71*	3.19± 0.48*	87.32± 1.14*	78.24± 0.99*	5.53± 0.79*	1.56± 0.36
DTC (Luo et al., 2021a)			84.89± 2.04*	74.61± 1.83*	12.67± 2.97*	3.42± 0.48*	88.06± 0.42*	78.71± 0.49*	10.13± 0.94*	2.62± 0.47*	86.43± 0.82*	77.12± 0.94*	6.24± 0.73*	2.27± 0.38*
CoraNet (Shi et al., 2021)			84.46± 1.53*	73.84± 1.72*	9.03± 2.26	2.60± 0.38	87.95± 0.69*	78.60± 0.85*	11.04± 0.68*	3.47± 0.82*	86.51± 1.29*	77.21± 1.17*	6.40± 0.98*	2.16± 0.46*
SPCL (Peng et al., 2021b)			85.92± 0.86*	75.49± 1.08*	9.74± 1.29	2.79± 0.41	90.19± 0.52	82.55± 0.54*	7.28± 0.39*	2.31± 0.28	88.65± 1.06*	79.32± 0.89*	4.94± 0.52	1.87± 0.24
MC-Net+ (Wu et al., 2022)			85.40± 1.13*	75.09± 0.94*	9.68± 1.62	2.98± 0.39*	90.15± 0.44*	82.53± 0.48*	6.56± 0.43	2.16± 0.25	88.46± 0.96*	79.23± 1.02*	5.73± 0.83*	1.71± 0.42
URPC (Luo et al., 2022)			85.81± 0.97*	75.44± 0.86*	8.86± 1.24	2.52± 0.32	89.97± 0.38*	81.30± 0.52*	9.33± 0.46*	3.46± 0.24*	87.48± 0.83*	78.55± 0.77*	5.13± 0.62*	1.56± 0.40
PLCT (Chaitanya et al., 2023)			85.53± 1.06*	75.29± 0.91*	8.64± 1.22	2.89± 0.24*	90.12± 0.21*	82.46± 0.39*	7.08± 0.53*	2.44± 0.20*	88.41± 0.74*	79.21± 0.82*	5.76± 0.59*	2.08± 0.51*
DGCL (Wang et al., 2023c)			85.79± 1.12*	75.43± 1.24*	8.39± 1.36	2.51± 0.26	90.31± 0.47	82.87± 0.61	6.81± 0.72	2.03± 0.23	88.75± 0.53*	80.19± 0.47	5.28± 0.48*	1.79± 0.33
CAML (Gao et al., 2023)			86.12± 0.89*	75.66± 0.94*	9.53± 1.41	3.62± 0.57*	90.64± 0.38	83.12± 0.42	6.34± 0.31	1.93± 0.30	89.22± 0.66	80.62± 0.72	5.16± 0.41*	1.51± 0.26
DCNet (Chen et al., 2023a)			86.20± 0.73*	75.79± 0.84*	8.46± 1.07	2.71± 0.32	90.09± 0.44*	81.68± 0.48*	8.25± 0.42*	2.64± 0.35*	89.42± 0.60	80.69± 0.58	4.83± 0.25	1.12± 0.21
SFPC (Yang et al., 2023b)			86.48± 0.78	76.54± 0.83*	8.61± 0.94	2.33± 0.29	90.63± 0.39	82.99± 0.45	6.65± 0.57	2.09± 0.12	89.30± 0.47	80.64± 0.54	5.02± 0.46*	1.43± 0.27
Ours			**87.79± 0.69**	**78.40± 0.72**	**8.18± 0.86**	**2.01± 0.23**	**91.05± 0.35**	**83.45± 0.38**	**5.94± 0.36**	**1.67± 0.14**	**90.23± 0.39**	**81.24± 0.41**	**4.01± 0.22**	**1.06± 0.19**
nn-UNet	100	0	89.58± 0.45	81.33± 0.53	7.92± 0.47	1.78± 0.14	93.08± 0.11	85.27± 0.36	4.07± 0.33	1.22± 0.20	92.71± 0.31	85.35± 0.27	1.57± 0.15	0.46± 0.12

* Means our method is significantly better than the compared method with $p < 0.05$ via paired t-test.

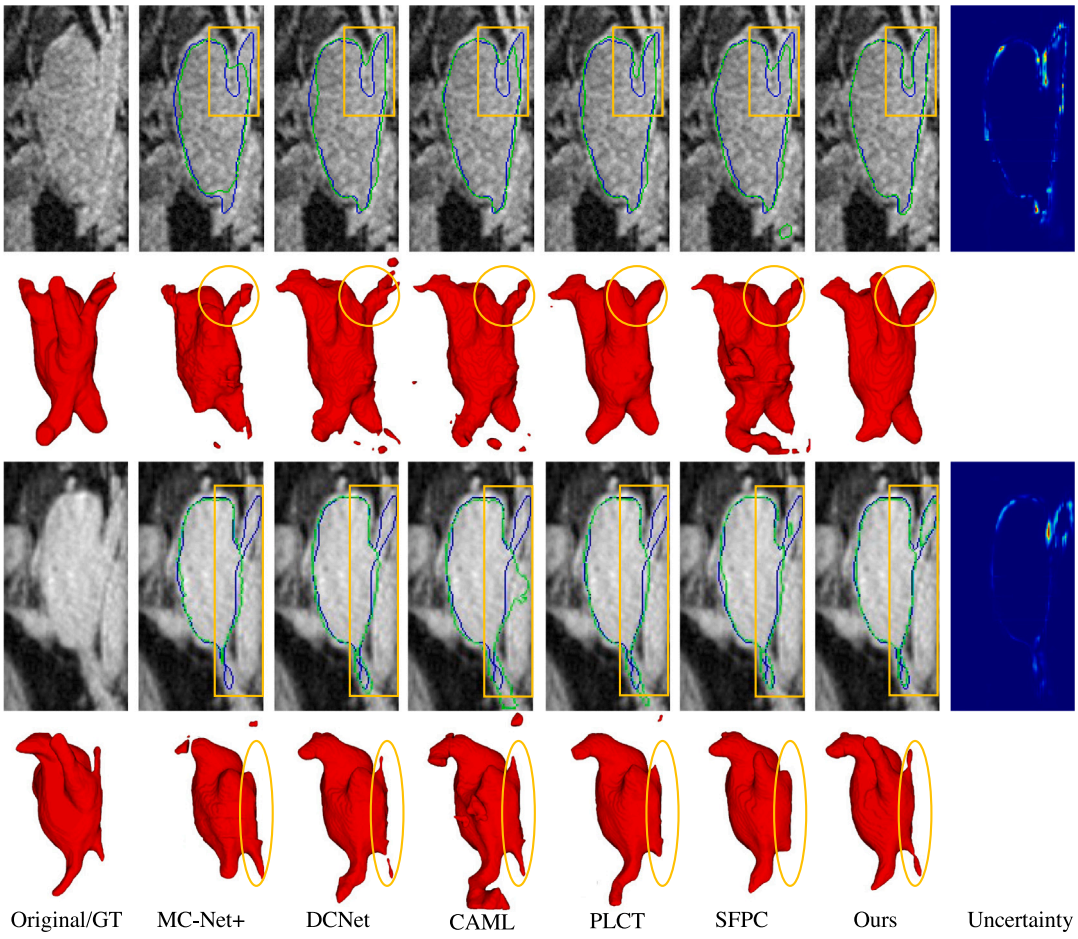

Fig. 4. Visual comparisons between the proposed method and strong baseline methods (second to sixth column) on two representative images from LA dataset. During training, 5% of training samples were annotated. Green and blue contours denote prediction and ground-truth edges, respectively. Second and fourth row (in red): view of the segmentation 3D organ. Last column: pixel-wise uncertainty of the aggregation prediction using entropy.

representative segmentation results from our method (seventh column) and the strong baselines (second to sixth column). This visual comparison further confirms the superior segmentation performance of our method, particularly in accurately delineating region edges (indicated by yellow rectangles and ellipses in the 2D and 3D views, respectively).

4.5. Performance on the 2017 ACDC dataset

To assess the scalability of our method, we conducted further evaluations on the 2D multi-class organ segmentation task on the 2017 ACDC dataset. The corresponding results of our method and other semi-supervised methods on the test dataset are displayed in Table 1 (right half), where our method demonstrates significant performance superiority compared to the other methods when trained with only 5% labeled data. While SPCL and DGCL achieve the top two performances among the other methods with a DSC of 81.82% and 80.57%, our method achieves an absolute improvement of 0.57% and 1.82%. Even when the proportion of annotated training data is increased to 20%, our method maintains a slight margin of superiority over the other existing methods. Similar to the visualization results obtained on the BraTS2020 dataset, our method shows the ability to accurately segment organ edges without the need for post-processing modules or shape-related constraints. Fig. 5 highlights the accurate segmentation of challenging areas by our method.

Overall, from both quantitative and qualitative results on three datasets, our method presents superiority compared to other SOTA methods for semi-supervised medical image segmentation with different ratios of annotated training data. Meanwhile, it demonstrates

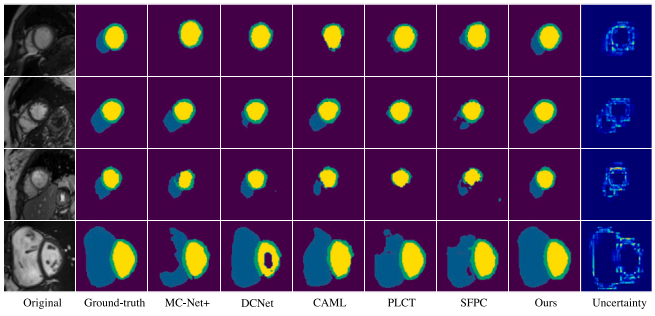

Fig. 5. Visual comparisons between the proposed method (last two columns) and strong baselines (third to seventh column) on four representative images from 2017 ACDC dataset. 5% training samples were annotated for model training. Second to eighth column: different colors indicate different types of segmented regions. Last column: pixel-wise uncertainty of the aggregation prediction using entropy.

that introducing meaningful feature-level perturbations to explore a broader range of hidden space can lead to improved semi-supervised segmentation performance on all datasets. Our method is not limited to a specific backbone network or segmentation task. Experimental results have demonstrated that our framework can be applied to various medical tasks (e.g. lesion or organ segmentation) in either 2D or 3D segmentation.

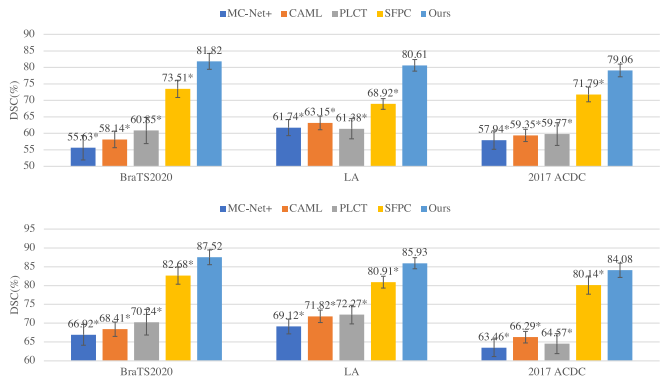

Fig. 6. Performance on edge regions from our method and strong baselines on BraTS2020, LA 2017 ACDC datasets, with 5% (upper) and 10% (lower) labeled images used for training. $*$ means our method is significantly better than the compared method with $p < 0.05$ via paired t-test.

4.6. Comparison on edge region

Some studies (Baumgartner et al., 2019; Kohl et al., 2018) have indicated the discrepancies among experts in ground truth annotations, especially around the edges regions. Therefore, accurate segmentation of edges depends heavily on the expert ground truth employed for training and evaluation. The annotation bias between different experts and the ambiguity present in edge regions pose significant challenges to the accurate segmentation and evaluation of such areas. While evaluating the segmentation performance, it is important to note that the seemingly small difference in DSC between our method and the strong baselines may be due to the fact that DSC is a metric that considers the entire segmentation regions, and the current strong baselines already demonstrate satisfactory performance in segmenting the main region of interest. However, accurate segmentation of edge regions remains a challenge for these methods. To compare the segmentation performance of different methods in edge areas relatively fairly and objectively, a 10-pixel-wide band is defined for the assessment. The pixel band is wide enough to encompass the slightly inter-expert discrepancies in the edge region, particularly when the boundaries are distinct. For both the BraTS2020, LA and 2017 ACDC datasets, our method outperformed the best baseline (SFPC) by 8.31%, 11.69% and 7.27% in DSC with 5% labeled images used for training, respectively (Fig. 6). This significant improvement is attributed to the edge-aware contrastive loss function, which enables the model to learn more discriminative representations in edge areas.

5. Discussion

In this section, we further discuss the impact of parameters and strategies in the proposed framework on model segmentation performance.

5.1. Hyper-parameter analysis

5.1.1. Effects of decoder numbers M

We investigate the sensitivity of M. The M controls the number of decoders and corresponding predictions, which plays a vital role in stable training and uncertainty estimation. As demonstrated in Fig. 7, M is set from 2 to 9 for our method, respectively. We can see that the M can also improve the segmentation performance. Note that the feature-level perturbation module has been introduced from G_2 to G_M at this stage. We observed that increasing the number of decoders can slightly improve the segmentation performance of the model when M is small, and our method is not sensitive to the number of M when it is greater than 4 on the three datasets.

5.1.2. Effects of contrastive loss weight λ_2

While the coefficient λ_1 in the loss function is automatically adjusted during training iterations, the selection of the other coefficient λ_2 can indeed impact the performance of our method. To investigate this, we conducted a hyperparameter sensitivity experiment on the BraTS2020, LA, and 2017 ACDC datasets, specifically focusing on the effects of λ_2 in balancing different losses (see Fig. 8). Here, a smaller λ would lead to decreased performance since insufficient contrastive training would result in inaccurate edge regions of the outputs generated by the decoders. On the other hand, a larger value of λ has no discernible effect. Therefore, in this paper, the weight λ is set to 0.5 to make optimal use of unlabeled data.

5.1.3. Effects of weak-to-strong perturbation

To achieve weak-to-strong perturbations, our method incorporates a sensible perturbation module that applies perturbations to different segmentation branches. The magnitude of the perturbation is adjusted using the hyperparameter κ. Specifically, we divide the perturbation range into equal parts based on the number of divided branches, with $M = 4$ in this section. Each part corresponds to a branch, and the perturbation is applied accordingly, gradually increasing in strength from one branch to the next. In Table 2 (lower half), we present the DSC performance of our method trained with different degrees of perturbation by varying κ on the BraTS2020 dataset. The results indicate that, in each semi-supervised setting, the DSC performance is comparable across different κ values, suggesting that our method is relatively robust to changes in the hyperparameter κ. Here, a larger κ would cause excessive perturbation to the features of the middle layer, leading to the loss of structural information and inaccurate segmentation results. On the other hand, a smaller κ may not introduce enough perturbation, resulting in overly consistent outputs from the decoders, which hinders the model explore a wider range of hidden space and perform accurate segmentation of edge regions. Therefore, we have adopted a perturbation coefficient of $\kappa = 0.2$, which generates different and meaningful outputs on all datasets. To further validate the effectiveness of the perturbation strategy that gradually increases in strength from weak to strong, we conducted a comparison by setting the same perturbation range for all perturbation branches, as illustrated in Table 2 (upper half). In this comparison, we used a perturbation range of $\kappa = 0.2$ for each perturbation branch on the BraTS2020 dataset. The results demonstrate our proposed weak-to-strong strategy achieves certain performance improvements.

5.1.4. Effects of pixel representation $|S_c(\mathbf{x}_j)|$

Here, we investigate the effects of the positive pairs selection strategy. In Table 3, the results using the uncertainty ranking selection strategy are always higher than the random selection strategy, which confirms the effectiveness of our proposed selection strategy. Besides, the number of selected pixel representations $|S_c(\mathbf{x}_j)|$, which is sorted by uncertainty, to match its class mean representation is varied between the values of 3, 6, and 10. In Table 3 (lower half), there is a marginal improvement in performance when the number of selected pixels is increased from 3 to 6. However, when the number of selected pixels is further increased from 6 to 10, there is almost no noticeable pattern in performance. This observation suggests that the performance of the model remains stable as the number of positive pixel representations sampled per class per image is varied. The final performance of the model is not significantly affected by these changes.

5.2. Computing complexity

The backbone networks of most baseline methods employ three prominent medical image segmentation frameworks, *i.e.*, U-Net (Ronneberger et al., 2015), V-Net (Milletari et al., 2016) and 3D-UNet (Çiçek et al., 2016). Compared with 3D-Unet, V-Net incorporates the element-wise horizontal residual function, resulting in increased

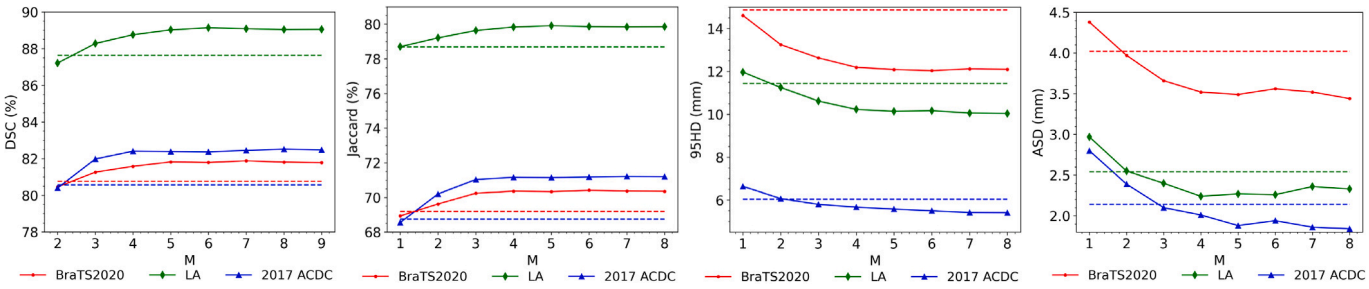

Fig. 7. Performance of our method on BraTS2020, LA and 2017 ACDC datasets with different decoder number of M, where 5% labeled images were used for model training. Dashed lines represent the performance of the strongest baselines.

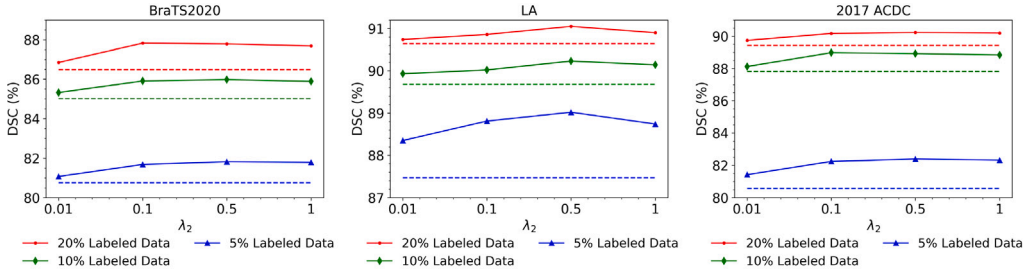

Fig. 8. Performance of our method on the BraTS2020, LA and 2017 ACDC datasets with different contrastive loss weights λ_2, where 5%, 10% and 20% labeled images were used for model training, respectively. Dashed lines represent the performance of the strongest baselines.

Table 2
Effects analysis of the weak-to-strong perturbation strategy on the BraTS2020 and 2017 ACDC datasets with 5%, 10% and 20% labeled images for training. DSC is used as the evaluation metric to assess the performance.

Perturbation strategy	κ	BraTS2020 (3D)			ACDC (2D)		
		5% labeled	10% labeled	20% labeled	5% labeled	10% labeled	20% labeled
Same range	0.1	80.86± 0.81	84.84± 0.77	86.77± 0.69	80.96± 0.95*	87.74± 0.88	88.96± 0.34
Same range	0.2	**81.02± 0.89**	**85.16± 0.82**	**87.14± 0.76**	**81.14± 0.89**	88.01± 0.70	**89.37± 0.62**
Same range	0.3	80.67± 0.92	85.07± 1.24	87.03± 1.16	81.02± 0.92	**88.06± 0.83**	89.25± 0.48
Same range	0.4	80.54± 1.13	85.01± 1.08	86.91± 1.21	80.99± 1.04*	87.83± 1.12	89.14± 0.51
Weak-to-strong	0.1	81.34± 0.74	85.61± 0.82	87.47± 0.51	82.08± 0.68	88.54± 0.72	89.94± 0.42
Weak-to-strong	0.2	**81.82± 0.77**	**85.98± 0.75**	87.79± 0.69	**82.39± 0.82**	**88.92± 0.64**	**90.23± 0.39**
Weak-to-strong	0.3	81.75± 1.13	85.92± 1.07	**87.83± 0.74**	82.25± 0.73	88.78± 0.51	90.18± 0.44
Weak-to-strong	0.4	81.49± 1.04	85.70± 1.16	87.58± 0.98	82.16± 0.92	88.61± 0.94	90.04± 0.37

* Means our method (Weak-to-strong, $\kappa = 0.2$) is significantly better than the compared method with $p < 0.05$ via paired t-test.

Table 3
Effects analysis of the contrastive pixel selection strategy on the BraTS2020 and 2017 ACDC datasets with 5%, 10% and 20% labeled images for training. DSC is used as the evaluation metric to assess the performance.

Selection strategy	$\lvert S_c(\mathbf{x}_j)\rvert$	BraTS2020 (3D)			ACDC (2D)		
		5% labeled	10% labeled	20% labeled	5% labeled	10% labeled	20% labeled
Random sort	3	80.14± 0.92*	84.85± 0.76	**87.01± 0.65**	81.26± 0.83	87.35± 0.82	89.23± 0.47
Random sort	6	80.42± 0.79	**84.92± 0.62**	86.94± 0.73	**81.46± 0.58**	87.77± 0.67	**89.37± 0.45**
Random sort	10	**80.48± 0.83**	84.73± 0.57*	86.78± 0.68	81.32± 0.74	**87.81± 0.79**	89.19± 0.36
Uncertainty sort	3	81.63± 0.68	85.75± 0.95	87.85± 0.83	82.24± 0.65	88.72± 0.81	90.11± 0.42
Uncertainty sort	6	**81.82± 0.77**	85.98± 0.75	87.79± 0.69	**82.39± 0.82**	**88.92± 0.64**	90.23± 0.39
Uncertainty sort	10	81.68± 0.73	**86.03± 0.86**	**87.96± 0.91**	82.30± 0.91	88.86± 0.58	**90.31± 0.32**

* Means our method (Uncertainty sort, $|S_c(\mathbf{x}_j)| = 6$) is significantly better than the compared method with $p < 0.05$ via paired t-test.

model parameters and higher computational complexity. To enable the application of our method to both 2D and 3D images, we opt U-Net (Ronneberger et al., 2015) and 3D-UNet (Çiçek et al., 2016) as the backbone networks for our semi-supervised framework. Number of Parameters (Para.), Multiply ACcumulate operations (MACs) and training time are used to fairly compare the computational complexity. As illustrated in Table 4, the complexity overhead of our method is marginally greater than that of the single decoder methods based on the V-Net backbone, such as UAMT (Yu et al., 2019) and DTC (Luo et al., 2021a). However, when compared to the V-Net-based MC-Net+ (Wu et al., 2022) with multi-decoder and Tri-U-MT (Wang et al., 2021a) with teacher-student network, our method ($M = 4$ by default) has lower computational overhead while achieving superior segmentation performance. All observations suggest that our method achieves performance improvements with an appropriate increase in computational overhead.

5.3. Effects of different perturbation methods

The application of perturbations to the feature representation at a specific hidden layer is a crucial aspect of consistency training. To

Table 4
Quantitative comparisons with other state-of-the-art methods in computational complexity. M denotes the number of decoders in our method.

Method	BraTS2020 (3D)			2017 ACDC (2D)		
	Para.(M)	MACs(G)	Time(H)	Para.(M)	MACs(G)	Time(H)
UAMT (Yu et al., 2019)	9.44	47.02	2.9	1.81	2.99	1.2
SASSNet (Li et al., 2020b)	9.44	47.02	3.6	1.81	2.99	1.5
Tri-U-MT (Wang et al., 2021a)	36.52	52.15	3.2	5.13	8.17	1.3
DTC (Luo et al., 2021a)	9.44	47.05	1.8	1.81	3.02	0.9
CoraNet (Shi et al., 2021)	8.93	107.41	4.1	1.69	9.42	1.8
MC-Net+ (Wu et al., 2022)	15.25	126.35	4.4	2.58	5.32	1.8
URPC (Luo et al., 2022)	5.88	69.43	2.2	1.83	3.02	1.0
PLCT (Chaitanya et al., 2023)	8.13	101.37	3.8	1.72	7.75	1.6
DGCL (Wang et al., 2023c)	10.86	94.22	3.6	2.26	6.14	1.5
CAML (Gao et al., 2023)	20.13	112.89	4.2	3.62	13.16	1.7
DCNet (Chen et al., 2023a)	12.56	78.16	2.7	2.38	5.36	1.3
SFPC (Yang et al., 2023b)	8.24	87.27	3.1	1.81	4.76	1.4
Ours ($M = 3$)	9.21	90.65	3.2	2.02	3.68	1.3
Ours ($M = 4$)	11.25	106.47	3.6	2.23	4.19	1.4
Ours ($M = 5$)	13.29	121.64	3.9	2.46	4.71	1.2
nn-UNet	4.86	63.12	2.2	1.47	2.52	2.1

Table 5
Effects analysis of our proposed ensemble strategies on the BraTS2020 and 2017 ACDC datasets, where 5% and 10% labeled images were used for model training. WSP, ECL and FPC+ denote the Weak-to-Strong Perturbation, Edge-aware Contrastive Loss and Feature-Perturbed Consistency with uncertainty-weighted aggregation, respectively.

% scans used		Designs			BraTS2020 (3D)				2017 ACDC (2D)			
Labeled	Unlabeled	WSP	ECL	FPC+	DSC (%) ↑	Jaccard (%) ↑	95HD (mm) ↓	ASD (mm) ↓	DSC (%) ↑	Jaccard (%) ↑	95HD (mm) ↓	ASD (mm) ↓
5	95				46.82	36.96	24.16	8.27	49.37	41.52	18.04	8.26
		✓			62.44	52.04	16.72	6.81	61.53	52.91	15.47	7.42
			✓		57.51	46.19	17.74	6.43	59.47	50.62	14.36	5.69
				✓	75.43	65.35	14.86	4.95	72.47	63.06	6.85	2.82
		✓	✓		65.41	55.82	15.94	6.26	68.72	59.09	7.02	2.53
		✓		✓	80.94	69.77	13.07	3.72	80.64	68.78	5.96	2.09
			✓	✓	77.69	66.37	14.79	4.63	76.95	66.17	7.41	3.69
		✓	✓	✓	**81.82**	**70.36**	**12.19**	**3.47**	**82.39**	**71.16**	**5.67**	**2.01**
10	90				49.02	38.87	20.587	7.29	54.52	46.34	15.63	7.24
		✓			64.89	54.78	17.13	6.77	68.39	58.81	7.04	2.67
			✓		59.28	47.59	17.14	6.04	63.27	53.14	7.27	2.35
				✓	77.12	65.92	14.42	5.36	75.41	64.73	6.95	2.52
		✓	✓		71.45	60.30	15.96	5.68	73.04	62.57	6.68	2.61
		✓		✓	85.29	75.14	10.77	3.01	87.65	78.69	5.04	1.41
			✓	✓	82.37	70.72	15.89	3.49	84.16	72.37	7.74	2.65
		✓	✓	✓	**85.98**	**75.63**	**9.97**	**2.64**	**88.92**	**79.65**	**4.32**	**1.19**

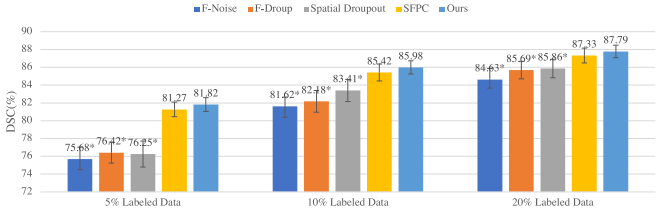

Fig. 9. Segmentation performance with different perturbation strategies on the BraTS2020 dataset, where 5%, 10% and 20% labeled images were used for model training, respectively. $*$ means our method is significantly better than the compared method with $p < 0.05$ via paired t-test.

evaluate the impact of the feature perturbation strategy adopted in our method on the segmentation performance, we compared it with four additional strategies: F-Noise (Ouali et al., 2020), F-Drop (Ouali et al., 2020), Spatial Dropout (Tompson et al., 2015), and SFPC (Yang et al., 2023b). Fig. 9 illustrates the segmentation performance achieved using these different feature perturbation strategies. The results clearly demonstrate that the perturbation strategy adopted in our method achieves the highest DSC value in three semi-supervised learning settings. The results presented above evident that our proposed weak-to-strong perturbation module enables a broader exploration of the latent space while remaining within a reasonable range.

5.4. Effects of ensemble strategies

To investigate the effectiveness of our designed module, we conduct detailed ablation studies in different model settings on BraTS2020 and 2017 ACDC with 5% and 10% labeled images for model training. The number of M is set to 4 and the feature-perturbed consistency with average aggregation is applied by default in the framework. As shown in Table 5, (1) the substantial performance improvements, with average DSC gains of 16.41%, 14.53% (compare the last and the fourth to last row in upper and lower halves respectively) on BraTS2020 dataset and 13.67%, 15.88% on 2017 ACDC dataset, are observed when employing the feature-perturbed consistency approach with uncertainty-weighted aggregation. This is because in the initial stage of model training, the segmentation results generated by G_2 to G_M present high uncertainty and are considered inaccurate. As a result, if the results of G_1 to G_M are simply averaged to obtain the aggregation label $\bar{\mathbf{p}}$, the accuracy of $\bar{\mathbf{p}}$ is seriously compromised. This hindrance prevents the improvement of model performance. However, when uncertainty-weighted aggregation stratagy is introduced, the segmentation results from G_2 to G_M with high uncertainty have minimal impact on $\bar{\mathbf{p}}$ in the initial stage. In other words, $\bar{\mathbf{p}}$ is primarily determined by the results of G_1. As the segmentation capabilities of G_2 to G_M improve during the model update process, the uncertainty of their outputs gradually decreases. Consequently, the weights assigned to their outputs in $\bar{\mathbf{p}}$ steadily increase; (2) introducing weak-to-strong perturbation, denoted by WSP, results in obvious average DSC improvement of 4.13%, 3.61% (compare the last

two rows in upper and lower halves respectively) on BraTS2020 dataset and 5.44%, 4.76% on 2017 ACDC dataset. These results demonstrate that our proposed weak-to-strong perturbation module allows for more extensive exploration within a reasonable range of the latent space, leading to better performance; (3) encouraging the framework focus on edge regions using edge-aware contrastive loss, denoted by ECL, results in average DSC improvements of 0.88%, 0.69% (compare the last and the third to last row) on BraTS2020 dataset and 1.75%, 1.27% on 2017 ACDC dataset. It is worth noting that the ECL module is designed specifically for accurate segmentation of edge regions. Its main function is to improve the segmentation performance of edge regions while WSP and FPC+ for the whole region. Given that the edge region constitutes only a minor portion of the total segmentation area, the improvement in performance achieved by ECL is comparatively less significant than that of the other two modules when considering the overall DSC metric. All the observations suggest that the contrastive learning branch we developed effectively improves the model's ability to learn more discriminative features in the edge region.

6. Conclusion

Deploying high-performance deep learning models for medical image segmentation, especially in edge regions, presents a formidable challenge due to the requirement for a large number of annotations. In this study, we propose a semi-supervised approach that addresses this issue by utilizing a substantial amount of unlabeled images alongside a limited set of annotations. Initially, we introduce a weak-to-strong perturbation module that effectively exploits the semantic information from the unlabeled data. To ensure learning from reliable regions of multiple predictions, we further develop a feature-perturbed consistency loss with an uncertainty-weighted aggregation strategy that automatically filters out unreliable regions. Additionally, we define an edge-aware contrastive loss to guide our framework to learn more discriminative representations in edge regions. By incorporating this contrastive loss, our framework achieves superior accuracy in identifying edge regions compared to other baseline methods on three datasets. We extensively evaluate our approach on three open-source datasets in both 2D and 3D, achieving the highest segmentation performance across various limited annotation scenarios. Furthermore, we demonstrate the robustness of our method to hyperparameter settings from different perspectives. However, our framework has the limitation that the number of parameters of the model increases dramatically as the number of decoders gradually increases. In the future, we aim to optimize the negative sample selection strategy and apply our proposed method to a broader range of medical image segmentation tasks.

CRediT authorship contribution statement

Yang Yang: Writing – review & editing, Writing – original draft, Software, Methodology, Investigation, Conceptualization. **Guoying Sun:** Writing – review & editing, Validation, Data curation. **Tong Zhang:** Writing – review & editing, Supervision. **Ruixuan Wang:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Jingyong Su:** Writing – review & editing, Resources, Project administration, Funding acquisition, Formal analysis, Conceptualization.

Data use declaration

Our experimental data was collected from open source datasets. The BraTS2020 can be downloaded at: https://www.med.upenn.edu/cbica/brats2020/data.html, the LA dataset is released at : https://www.cardiacatlas.org/atriaseg2018-challenge/, and the 2017 ACDC dataset is available at: https://humanheart-project.creatis.insa-lyon.fr/database/#

Declaration of competing interest

The authors affirm that they have no disclosed competing financial interests or personal relationships that could have potentially influenced the findings presented in this paper.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (grant No. 62376068, 62350710797, U24A20340, 32361143787, 62071502), by Guangdong Basic and Applied Basic Research Foundation (grant No. 2023B1515120065), by Guangdong S&T programme (grant No. 2023A0505050109), by Shenzhen Science and Technology Innovation Program (grant No. JCYJ20220818102414031), by the Guangdong Excellent Youth Team Program (grant No. 2023B1515040025), by the Major Key Project of PCL (grant No. PCL2023AS7-1). We also appreciate the efforts to share several public benchmarks (Luo et al., 2022; Yu et al., 2019; Chaitanya et al., 2023).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.media.2024.103450.

Data availability

The data used in the paper are from open source datasets, and download links have been released in the article. The code is released publicly at https://github.com/youngyzzZ/SSL-w2sPC.

References

Alonso, I., Sabater, A., Ferstl, D., Montesano, L., Murillo, A.C., 2021. Semi-supervised semantic segmentation with pixel-level contrastive learning from a class-wise memory bank. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 8219–8228.

Bai, Y., Chen, D., Li, Q., Shen, W., Wang, Y., 2023. Bidirectional copy-paste for semi-supervised medical image segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11514–11524.

Bai, W., Oktay, O., Sinclair, M., Suzuki, H., Rajchl, M., Tarroni, G., Glocker, B., King, A., Matthews, P.M., Rueckert, D., 2017. Semi-supervised learning for network-based cardiac MR image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 253–260.

Basak, H., Yin, Z., 2023. Pseudo-label guided contrastive learning for semi-supervised medical image segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 19786–19797.

Baumgartner, C.F., Tezcan, K.C., Chaitanya, K., Hötker, A.M., Muehlematter, U.J., Schawkat, K., Becker, A.S., Donati, O., Konukoglu, E., 2019. Phiseg: Capturing uncertainty in medical image segmentation. In: Medical Image Computing and Computer Assisted Intervention. pp. 119–127.

Bernard, O., Lalande, A., Zotti, C., Cervenansky, F., Yang, X., Heng, P.-A., Cetin, I., Lekadir, K., Camara, O., Ballester, M.A.G., et al., 2018. Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: is the problem solved? IEEE Trans. Med. Imaging 37 (11), 2514–2525.

Cai, Q., Wang, Y., Pan, Y., Yao, T., Mei, T., 2020. Joint contrastive learning with infinite possibilities. Adv. Neural Inf. Process. Syst. 33, 12638–12648.

Chaitanya, K., Erdil, E., Karani, N., Konukoglu, E., 2020. Contrastive learning of global and local features for medical image segmentation with limited annotations. Adv. Neural Inf. Process. Syst. 33, 12546–12558.

Chaitanya, K., Erdil, E., Karani, N., Konukoglu, E., 2023. Local contrastive loss with pseudo-label based self-training for semi-supervised medical image segmentation. Med. Image Anal. 87, 102792.

Chaitanya, K., Karani, N., Baumgartner, C.F., Becker, A., Donati, O., Konukoglu, E., 2019. Semi-supervised and task-driven data augmentation. In: Information Processing in Medical Imaging. pp. 29–41.

Chapelle, O., Scholkopf, B., Zien, A., 2009. Semi-supervised learning. IEEE Trans. Neural Netw. 20 (3), 542.

Chen, F., Fei, J., Chen, Y., Huang, C., 2023a. Decoupled consistency for semi-supervised medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 551–561.

Chen, G., Li, L., Dai, Y., Zhang, J., Yap, M.H., 2023b. AAU-net: An adaptive attention U-net for breast lesions segmentation in ultrasound images. IEEE Trans. Med. Imaging 42 (5), 1289–1300.
Chen, C., Qin, C., Qiu, H., Ouyang, C., Wang, S., Chen, L., Tarroni, G., Bai, W., Rueckert, D., 2020. Realistic adversarial data augmentation for MR image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 667–677.
Chen, C., Zhou, K., Wang, Z., Xiao, R., 2023c. Generative consistency for semi-supervised cerebrovascular segmentation from TOF-MRA. IEEE Trans. Med. Imaging 42 (2), 346–353.
Cheng, B., Girshick, R., Dollár, P., Berg, A.C., Kirillov, A., 2021. Boundary iou: Improving object-centric image segmentation evaluation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 15334–15342.
Chu, J., Chen, Y., Zhou, W., Shi, H., Cao, Y., Tu, D., Jin, R., Xu, Y., 2020. Pay more attention to discontinuity for medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, Vol. 12264. Springer, pp. 166–175.
Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O., 2016. 3D U-net: learning dense volumetric segmentation from sparse annotation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 424–432.
Fan, D.-P., Zhou, T., Ji, G.-P., Zhou, Y., Chen, G., Fu, H., Shen, J., Shao, L., 2020. Inf-net: Automatic COVID-19 lung infection segmentation from CT images. IEEE Trans. Med. Imaging 39 (8), 2626–2637.
Gao, S., Zhang, Z., Ma, J., Li, Z., Zhang, S., 2023. Correlation-aware mutual learning for semi-supervised medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 98–108.
Grandvalet, Y., Bengio, Y., 2004. Semi-supervised learning by entropy minimization. Adv. Neural Inf. Process. Syst. 17.
Gu, R., Zhang, J., Wang, G., Lei, W., Song, T., Zhang, X., Li, K., Zhang, S., 2022. Contrastive semi-supervised learning for domain adaptive segmentation across similar anatomical structures. IEEE Trans. Med. Imaging 42 (1), 245–256.
Gupta, S., Hu, X., Kaan, J., Jin, M., Mpoy, M., Chung, K., Singh, G., Saltz, M., Kurc, T., Saltz, J., et al., 2022. Learning topological interactions for multi-class medical image segmentation. In: European Conference on Computer Vision. pp. 701–718.
Hooper, S., Wornow, M., Seah, Y.H., Kellman, P., Xue, H., Sala, F., Langlotz, C., Re, C., 2020. Cut out the annotator, keep the cutout: better segmentation with weak supervision. In: International Conference on Learning Representations.
Huang, X., Belongie, S., 2017. Arbitrary style transfer in real-time with adaptive instance normalization. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 1501–1510.
Huang, Y., Kang, D., Chen, L., Zhe, X., Jia, W., Bao, L., He, X., 2022. Car: Class-aware regularizations for semantic segmentation. In: European Conference on Computer Vision. pp. 518–534.
Huang, H., Xie, S., Lin, L., Tong, R., Chen, Y.-W., Li, Y., Wang, H., Huang, Y., Zheng, Y., 2023. SemiCVT: Semi-supervised convolutional vision transformer for semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11340–11349.
Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H., 2021. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods 18 (2), 203–211.
Kohl, S., Romera-Paredes, B., Meyer, C., De Fauw, J., Ledsam, J.R., Maier-Hein, K., Eslami, S., Jimenez Rezende, D., Ronneberger, O., 2018. A probabilistic u-net for segmentation of ambiguous images. Adv. Neural Inf. Process. Syst. 31.
Lee, H.J., Kim, J.U., Lee, S., Kim, H.G., Ro, Y.M., 2020. Structure boundary preserving segmentation for medical image with ambiguous boundary. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4817–4826.
Lei, T., Zhang, D., Du, X., Wang, X., Wan, Y., Nandi, A.K., 2023. Semi-supervised medical image segmentation using adversarial consistency learning and dynamic convolution network. IEEE Trans. Med. Imaging 42 (5), 1265–1277.
Li, P., Li, D., Li, W., Gong, S., Fu, Y., Hospedales, T.M., 2021a. A simple feature augmentation for domain generalization. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 8886–8895.
Li, D., Liu, Y., Song, L., 2022. Adaptive weighted losses with distribution approximation for efficient consistency-based semi-supervised learning. IEEE Trans. Circuits Syst. Video Technol. 32 (11), 7832–7842.
Li, B., Wu, F., Lim, S.-N., Belongie, S., Weinberger, K.Q., 2021b. On feature normalization and data augmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 12383–12392.
Li, X., Yu, L., Chen, H., Fu, C.-W., Xing, L., Heng, P.-A., 2020a. Transformation-consistent self-ensembling model for semisupervised medical image segmentation. IEEE Trans. Neural Netw. Learn. Syst. 32 (2), 523–534.
Li, S., Zhang, C., He, X., 2020b. Shape-aware semi-supervised 3D semantic segmentation for medical images. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 552–561.
Liu, Y., Tian, Y., Chen, Y., Liu, F., Belagiannis, V., Carneiro, G., 2022. Perturbed and strict mean teachers for semi-supervised semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4258–4267.
Lu, C., de Geus, D., Dubbelman, G., 2023. Content-aware token sharing for efficient semantic segmentation with vision transformers. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 23631–23640.
Luo, X., Chen, J., Song, T., Wang, G., 2021a. Semi-supervised medical image segmentation through dual-task consistency. In: Proceedings of the AAAI Conference on Artificial Intelligence. (10), pp. 8801–8809.
Luo, X., Liao, W., Chen, J., Song, T., Chen, Y., Zhang, S., Chen, N., Wang, G., Zhang, S., 2021b. Efficient semi-supervised gross target volume of nasopharyngeal carcinoma segmentation via uncertainty rectified pyramid consistency. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 318–329.
Luo, X., Wang, G., Liao, W., Chen, J., Song, T., Chen, Y., Zhang, S., Metaxas, D.N., Zhang, S., 2022. Semi-supervised medical image segmentation via uncertainty rectified pyramid consistency. Med. Image Anal. 80, 102517.
Lyu, F., Ye, M., Carlsen, J.F., Erleben, K., Darkner, S., Yuen, P.C., 2022. Pseudo-label guided image synthesis for semi-supervised COVID-19 pneumonia infection segmentation. IEEE Trans. Med. Imaging 42 (3), 797–809.
Ma, Y., Shi, H., Tan, S., Tao, Y., Song, B., 2022. Consistency regularization auto-encoder network for semi-supervised process fault diagnosis. IEEE Trans. Instrum. Meas. 71, 1–15.
Ma, J., Wang, C., Liu, Y., Lin, L., Li, G., 2023. Enhanced soft label for semi-supervised semantic segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 1185–1195.
Marin, D., He, Z., Vajda, P., Chatterjee, P., Tsai, S., Yang, F., Boykov, Y., 2019. Efficient segmentation: Learning downsampling near semantic boundaries. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 2131–2141.
Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al., 2014. The multimodal brain tumor image segmentation benchmark (BRATS). IEEE Trans. Med. Imaging 34 (10), 1993–2024.
Milletari, F., Navab, N., Ahmadi, S.-A., 2016. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: International Conference on 3D Vision. pp. 565–571.
Miyato, T., Maeda, S.-i., Koyama, M., Ishii, S., 2018. Virtual adversarial training: a regularization method for supervised and semi-supervised learning. IEEE Trans. Pattern Anal. Mach. Intell. 41 (8), 1979–1993.
Nain, D., Haker, S., Bobick, A., Tannenbaum, A., 2007. Multiscale 3-d shape representation and segmentation using spherical wavelets. IEEE Trans. Med. Imaging 26 (4), 598–618.
Nguyen-Duc, T., Le, T., Bammer, R., Zhao, H., Cai, J., Phung, D., 2023. Cross-adversarial local distribution regularization for semi-supervised medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 183–194.
Ouali, Y., Hudelot, C., Tami, M., 2020. Semi-supervised semantic segmentation with cross-consistency training. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 12674–12684.
Peng, S., Jiang, W., Pi, H., Li, X., Bao, H., Zhou, X., 2020. Deep snake for real-time instance segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8533–8542.
Peng, J., Pedersoli, M., Desrosiers, C., 2021a. Boosting semi-supervised image segmentation with global and local mutual information regularization. arXiv preprint arXiv:2103.04813.
Peng, J., Wang, P., Desrosiers, C., Pedersoli, M., 2021b. Self-paced contrastive learning for semi-supervised medical image segmentation with meta-labels. Adv. Neural Inf. Process. Syst. 34, 16686–16699.
Pham, H., Dai, Z., Xie, Q., Le, Q.V., 2021. Meta pseudo labels. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11557–11568.
Qiao, P., Li, H., Song, G., Han, H., Gao, Z., Tian, Y., Liang, Y., Li, X., Zhou, S.K., Chen, J., 2023. Semi-supervised CT lesion segmentation using uncertainty-based data pairing and SwapMix. IEEE Trans. Med. Imaging.
Qiu, D., Yi, J., Peng, J., 2022. WDA-net: Weakly-supervised domain adaptive segmentation of electron microscopy. In: IEEE International Conference on Bioinformatics and Biomedicine. pp. 1132–1137.
Rasmus, A., Berglund, M., Honkala, M., Valpola, H., Raiko, T., 2015. Semi-supervised learning with ladder networks. Adv. Neural Inf. Process. Syst. 28, 3546–3554.
Rizve, M.N., Duarte, K., Rawat, Y.S., Shah, M., 2021. In defense of pseudo-labeling: An uncertainty-aware pseudo-label selection framework for semi-supervised learning. In: International Conference on Learning Representations.
Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer Assisted Intervention. pp. 234–241.
Shi, Y., Zhang, J., Ling, T., Lu, J., Zheng, Y., Yu, Q., Qi, L., Gao, Y., 2021. Inconsistency-aware uncertainty estimation for semi-supervised medical image segmentation. IEEE Trans. Med. Imaging 41 (3), 608–620.
Sohn, K., Berthelot, D., Carlini, N., Zhang, Z., Zhang, H., Raffel, C.A., Cubuk, E.D., Kurakin, A., Li, C.-L., 2020. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. Adv. Neural Inf. Process. Syst. 33, 596–608.
Tang, L., Zhan, Y., Chen, Z., Yu, B., Tao, D., 2022. Contrastive boundary learning for point cloud segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8489–8499.

Tarvainen, A., Valpola, H., 2017. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. Adv. Neural Inf. Process. Syst. 30, 1195–1204.

Tompson, J., Goroshin, R., Jain, A., LeCun, Y., Bregler, C., 2015. Efficient object localization using convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 648–656.

Tsai, A., Yezzi, A., Wells, W., Tempany, C., Tucker, D., Fan, A., Grimson, W.E., Willsky, A., 2003. A shape-based approach to the segmentation of medical imagery using level sets. IEEE Trans. Med. Imaging 22 (2), 137–154.

Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egan, A., Yushkevich, P.A., Gee, J.C., 2010. N4ITK: improved N3 bias correction. IEEE Trans. Med. Imaging 29 (6), 1310–1320.

V., S.A., Dolz, J., Lombaert, H., 2023. Anatomically-aware uncertainty for semi-supervised image segmentation. Med. Image Anal. 91, 103011.

Wang, R., Chen, S., Ji, C., Fan, J., Li, Y., 2022a. Boundary-aware context neural network for medical image segmentation. Med. Image Anal. 78, 102395.

Wang, T., Lu, J., Lai, Z., Wen, J., Kong, H., 2022b. Uncertainty-guided pixel contrastive learning for semi-supervised medical image segmentation. In: International Joint Conference on Artificial Intelligence. pp. 1444–1450.

Wang, Y., Wang, H., Shen, Y., Fei, J., Li, W., Jin, G., Wu, L., Zhao, R., Le, X., 2022c. Semi-supervised semantic segmentation using unreliable pseudo-labels. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4248–4257.

Wang, Y., Xiao, B., Bi, X., Li, W., Gao, X., 2023a. MCF: Mutual correction framework for semi-supervised medical image segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 15651–15660.

Wang, C., Xie, H., Yuan, Y., Fu, C., Yue, X., 2023b. Space engage: Collaborative space supervision for contrastive-based semi-supervised semantic segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 931–942.

Wang, K., Zhan, B., Zu, C., Wu, X., Zhou, J., Zhou, L., Wang, Y., 2021a. Tripled-uncertainty guided mean teacher model for semi-supervised medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 450–460.

Wang, X., Zhang, B., Yu, L., Xiao, J., 2023c. Hunting sparsity: Density-guided contrastive learning for semi-supervised semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3114–3123.

Wang, W., Zhou, T., Yu, F., Dai, J., Konukoglu, E., Van Gool, L., 2021b. Exploring cross-image pixel contrast for semantic segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 7303–7313.

Wu, Y., Ge, Z., Zhang, D., Xu, M., Zhang, L., Xia, Y., Cai, J., 2022. Mutual consistency learning for semi-supervised medical image segmentation. Med. Image Anal. 81, 102530.

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

Yang, Z., Farsiu, S., 2023. Directional connectivity-based segmentation of medical images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11525–11535.

Yang, L., Qi, L., Feng, L., Zhang, W., Shi, Y., 2023a. Revisiting weak-to-strong consistency in semi-supervised semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7236–7246.

Yang, Y., Wang, R., Zhang, T., Su, J., 2023b. Semi-supervised medical image segmentation via feature-perturbed consistency. In: IEEE International Conference on Bioinformatics and Biomedicine. pp. 1635–1642.

You, C., Zhou, Y., Zhao, R., Staib, L., Duncan, J.S., 2022. Simcvd: Simple contrastive voxel-wise representation distillation for semi-supervised medical image segmentation. IEEE Trans. Med. Imaging 41 (9), 2228–2237.

Yu, L., Wang, S., Li, X., Fu, C.-W., Heng, P.-A., 2019. Uncertainty-aware self-ensembling model for semi-supervised 3D left atrium segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 605–613.

Yuan, J., Deng, Z., Wang, S., Luo, Z., 2020. Multi receptive field network for semantic segmentation. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 1894–1903.

Zhang, Y., Zhang, X., Li, J., Qiu, R., Xu, H., Tian, Q., 2023. Semi-supervised contrastive learning with similarity co-calibration. IEEE Trans. Multimed. 25, 1749–1759.

Zhao, X., Fang, C., Fan, D.-J., Lin, X., Gao, F., Li, G., 2022. Cross-level contrastive learning and consistency constraint for semi-supervised medical image segmentation. In: IEEE International Symposium on Biomedical Imaging. pp. 1–5.

Zhao, X., Vemulapalli, R., Mansfield, P.A., Gong, B., Green, B., Shapira, L., Wu, Y., 2021. Contrastive learning for label efficient semantic segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 10623–10633.

Zhao, Z., Yang, L., Long, S., Pi, J., Zhou, L., Wang, J., 2023. Augmentation matters: A simple-yet-effective approach to semi-supervised semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11350–11359.

Zhou, Y., Chen, H., Lin, H., Heng, P.-A., 2020. Deep semi-supervised knowledge distillation for overlapping cervical cell instance segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 521–531.

Zhou, Y., Xu, H., Zhang, W., Gao, B., Heng, P.-A., 2021. C3-semiseg: Contrastive semi-supervised segmentation via cross-set learning and dynamic class-balancing. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 7036–7045.

Zhuang, J.-X., Cai, J., Zhang, J., Zheng, W.-s., Wang, R., 2023. Class attention to regions of lesion for imbalanced medical image recognition. Neurocomputing 555, 126577.