Knowledge-Based Systems 309 (2025) 112890

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

Uncertainty-aware consistency learning for semi-supervised medical image segmentation

Min Dong [a,b], Ating Yang [a], Zhenhang Wang [a], Dezhen Li [c], Jing Yang [a], Rongchang Zhao [d,*]

[a] School of Electrical and Information Engineering, Zhengzhou University, Zhengzhou, 450001, China
[b] Industrial Technology Research Institute, Zhengzhou University, Zhengzhou, 450001, China
[c] Henan Institute of Advanced Technology, Zhengzhou University, Zhengzhou, 450001, China
[d] School computer science and engineering, Central South University, Changsha, 410012, China

ARTICLE INFO

Keywords:
Semi-supervised learning
Medical image segmentation
Consistency learning
Uncertainty estimation

ABSTRACT

Semi-supervised medical image segmentation faces two challenging issues: (1) insufficient exploration of latent structures leading to difficulty in comprehensively capturing complex features and structures in medical images; (2) sensitivity to noise, where unlabeled data lacks accurate label information, making the model more prone to noise interference during the learning process. In this paper, a method, uncertainty-aware consistency learning (UAC), is proposed to improve the poor generalization and suboptimal performance in semi-supervised medical image segmentation caused by insufficient information exploration and sensitivity to noise. Firstly, by employing multiple perturbation strategies at both the input and output levels, specifically through data-level and scale-level perturbations, the model is better equipped to capture structural information within organs and essential features that impact segmentation performance. Secondly, the perturbation uncertainty leverages perturbation prediction differences to measure uncertainty helps the model generate reliable predictions and avoid excessive focus on unreliable areas in the predictions. Experimental results on three public medical image segmentation datasets demonstrate that our UAC, utilizing multiple perturbation strategies and uncertainty estimation, exhibits generality across various organ segmentation tasks and achieves accurate segmentation, with the DICE of 91.15%(LA), 77.52%(Pancreas-CT) and 78.71%(PARSE) under a 10% label ratio setting. Comparative and ablation studies indicate that our method outperforms state-of-the-art semi-supervised medical image segmentation methods.

1. Introduction

Image segmentation has achieved outperformed results in many applications, such as disease detection [1,2], lesion segmentation [3,4], and pathological analysis [5,6]. Especially, the pre-trained fundamental models, MedSAM [7], MSA [8], are empowered with superpowers in zero-shot medical segmentation. However, due to the scarce labeled-data, semi-supervised learning (SSL) has attracted widespread attention in the field of medical image segmentation. In contrast to supervised learning algorithms, SSL methods harness a substantial amount of unlabeled data within the medical domain to boost learning efficacy.

Although great advantages have been made in existing methods, it is still challenging to achieve the medical image segmentation with small mount labeled data due to **(1) Insufficient mining of latent information.** In SLL, the lack of clear supervision signals in unlabeled data prevents the model from directly relying on labels to guide feature learning and model optimization, leading to underutilization of potential information. The valuable information embedded in unlabeled data is often overlooked, which may result in the model learning feature representations that are not comprehensive or accurate enough to handle segmentation tasks in complex areas (Fig. 1(a)). **(2) Sensitivity to noise.** Subjective annotations and noisy outliers lead to potential missegmentations and can significantly impact the performance of the segmentation model. As shown in Fig. 1(a), the model is prone to producing highly uncertain and erroneous segmentations in the organ adhesion region and some small branches because of the complex and intricate nature of these areas. Without constraining these low-confidence predictions, the model may overly focus on these noisy areas and unreliable predictions during training, leading to the learning of incorrect knowledge.

Uncertainty calibration is a potential approach to achieve semi-supervised image segmentation. By incorporating uncertainty calibration into the segmentation process, the model can learn more reliable

* Corresponding author.
E-mail addresses: iemdong@zzu.edu.cn (M. Dong), atyang@gs.zzu.edu.cn (A. Yang), wangzhl1999@163.com (Z. Wang), lidezhenw@163.com (D. Li), 13838638149@163.com (J. Yang), zhaorc@csu.edu.cn (R. Zhao).

https://doi.org/10.1016/j.knosys.2024.112890
Received 29 May 2024; Received in revised form 20 October 2024; Accepted 13 December 2024
Available online 20 December 2024

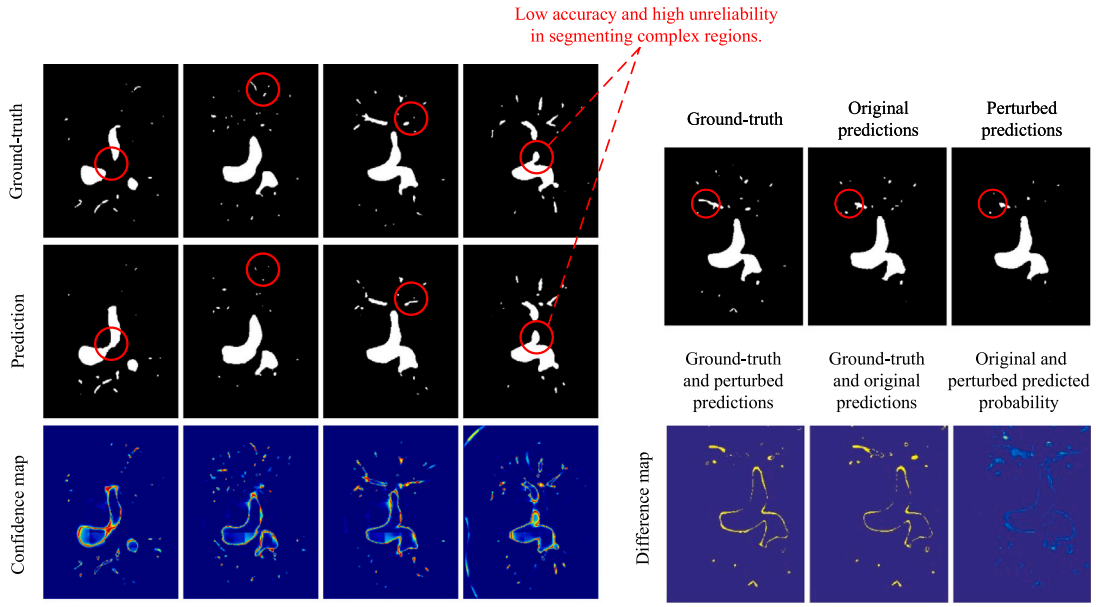

Fig. 1. (a) Implementing semi-supervised medical image segmentation faces two challenging issues: firstly, **insufficient mining of latent information** leads to the model struggling to cope with segmentation in challenging regions. Secondly, **sensitivity to noise**, that is, using unreliable or incorrect predictions directly for training, leads to subpar model performance. (b) The proposed UAC leverages perturbation uncertainty to learn reliable knowledge. It is easy to observe the similarity between the difference map of perturbed predictions and the difference map between perturbed predictions and labels. Regions with large differences in perturbed predictions have higher uncertainty, making it easier to generate incorrect predictions.

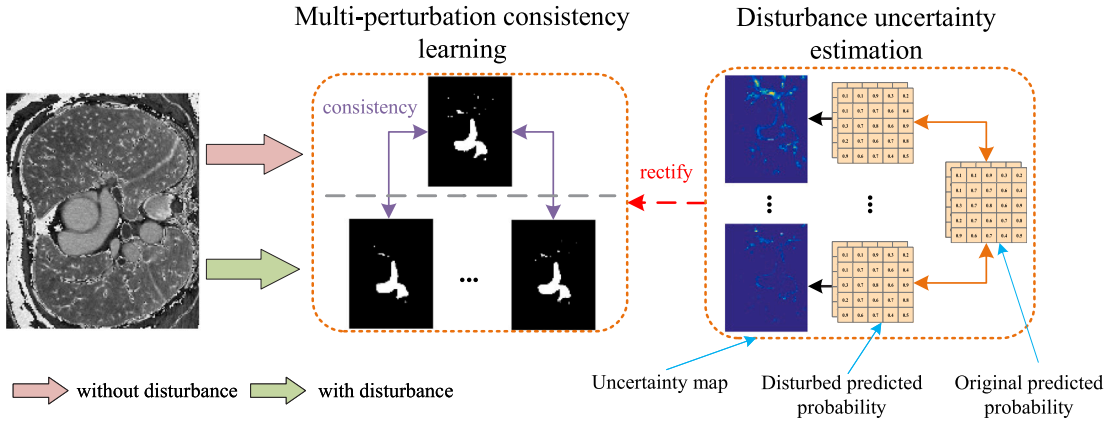

Fig. 2. The proposed UAC introduces a multi-level disturbance strategy for multi-scale consistency learning, enhancing the model's generalization ability. To prevent the model from focusing excessively on noisy regions and to learn more reliable knowledge, disturbance uncertainty estimation is introduced. By calculating the uncertainty map based on the degree of change between predictions before and after disturbance, the consistency learning is corrected.

information in unsupervised scenarios, thereby reducing the impact of noise interference. Generally, uncertainty estimation employs ensemble methods [9] or Monte Carlo sampling [10,11] to calculate model prediction confidence, and unreliable predictions are filtered out by setting thresholds to avoid the model focusing excessively on noisy regions and areas with high uncertainty. However, most of these methods are based on the assumption that higher entropy corresponds to higher uncertainty, making it difficult to capture the model's true uncertainty when the model is overly confident and the data distribution is uneven. Moreover, unreliable predictions often occur in regions with complex structures and blurry boundaries. Directly filtering out these unreliable predictions can result in the loss of information in these challenging areas.

In this paper, a novel consistency-based training strategy, Uncertainty-Aware Consistency learning (UAC, Fig. 2) is proposed to achieve semi-supervised medical image segmentation with few labeled images. As shown in Fig. 1(b), the predictions of model exhibit robustness in easily identifiable regions before and after perturbation, whereas in challenging regions, due to the complexity and ambiguity of the data, the model's predictions become unstable. There is a significant difference in prediction probabilities before and after perturbation in these difficult areas. By leveraging this characteristic, uncertainty can be estimated in a single forward pass, enabling dynamic adjustment of multi-level consistency. The proposed UAC is composed of three components: (1) Multi-perturbation strategy has been designed to enhance the model's generalization ability by perturbing data at both the data level and scale level simultaneously; (2) Disturbance uncertainty has

been introduced to guide output consistency by estimating uncertainty through calculating the difference in predictions before and after perturbation in a single forward pass. The corrected network focuses more on reliable prediction information, reducing the impact of noise; (3) Voxel contrastive learning has been introduced into the feature space to provide explicit supervision, further enhancing the separability of features, enabling the model to better distinguish features between different categories.

Overall, our proposed framework has three practical contributions:

- A novel consistency-based training strategy is proposed to achieve semi-supervised medical image segmentation with the uncertainty-calibrated multi-level consistency learning. The proposed method restricts consistency learning by computing model prediction uncertainty within a single forward pass to prevent the model from excessively focusing on high-risk areas during training and learning incorrect knowledge.
- A multi-level consistency learning method combining data-level and scale-level consistency is proposed to leverage unlabeled data for improving the generalization of SSL models.
- Extensive experiments are conducted on three publicly available benchmark datasets using different label ratios and compared with existing state of the art (SOTA). Experimental results show that our proposed method improves the segmentation performance and is superior to the SOTA.

2. Related work

2.1. Semi-supervised medical image segmentation

The emergence of SSL in medical image segmentation has increasingly captivated researchers, offering a solution to the data scarcity challenge prevalent in fully supervised methods. The SSL method can be roughly divided into two categories: (1) Pseudo-label learning [12], which uses unlabeled images to generate labels to guide model learning. However, the predictions of unlabeled data may contain noise, and directly using these pseudo-labels for training can lead to noise accumulation in the model. Selecting reliable predictions based on confidence [13,14] or generating soft pseudo-labels [15] to some extent alleviates this issue. (2) Consistency regularization [16–18] revolves around the core concept of obtaining outputs that remain invariant to perturbations. By bolstering the consistency of predictions across diverse viewpoints or multiple iterations, this approach diminishes the model's dependency on individual predictions and alleviates the pitfalls linked to overly confident pseudo-labels. Within the realm of consistency learning, the introduction of various perturbations, including data-level perturbations [19,20] and model-level perturbations [21,22], facilitates the acquisition of more resilient and broadly applicable feature representations by the model.

2.2. Uncertainty estimation

Uncertainty estimation is an important research direction in the fields of machine learning and deep learning. Its goal is to provide reliable assessments of model predictions and support model selection and ensemble methods. By estimating the uncertainty of model predictions, confidence information about the prediction results can be obtained, which helps us better understand the behavior and performance of the model. In the field of semi-supervised learning, Yu et al. [10] and Zhang et al. [11] estimated the predictive entropy of each target prediction as uncertainty using Monte Carlo sampling, filtering out unreliable predictions. While reducing the model's incorrect decisions, this may also lead to the loss of some important information. Methods using Bayesian neural networks [23] and Monte Carlo sampling [24] to estimate model uncertainty often require multiple forward passes and sampling, increasing computational complexity. Wu et al. [25] obtain the uncertainty of pseudo-labels by replicating the prediction head of the pre-trained model multiple times. Luo et al. [26] and Shi et al. [27] estimated uncertainty by calculating the variance between multi-views predictions. Building upon the above methods, we have improved the estimation method of model uncertainty. Cross-view and cross-scale uncertainty estimation have been introduced to guide consistency and minimize uncertainty as a regularization term to reduce prediction variance during training.

2.3. Contrastive learning

Contrastive Learning (CL) [28] is a crucial technique in unsupervised learning that has achieved state-of-the-art performance by leveraging abundant unlabeled data. The core idea of CL is to learn effective representations by comparing the similarity and dissimilarity between different parts of the data, essentially pulling similar pairs closer and pushing dissimilar pairs apart. The primary difference between CL-based frameworks lies in the strategies used to obtain positive and negative sample augmentations, such as utilizing momentum-updated memory banks to provide negative samples [29] and using other augmented samples within a batch as negative samples [30]. Dong et al. [31] believe that when selecting positive and negative samples, one should exclude samples that are erroneous or lack sufficient information. PC2Seg [32] and RCPS [33] transitions from image-level contrastive learning to dense voxel-level contrastive learning tasks and introduces a confidence negative sampling strategy. Drawing inspiration from the above, to further enhance model performance, a confidence negative sampling strategy is introduced in the feature space to improve the discriminative features of pixels.

3. Methodology

The proposed UAC framework (Fig. 3) conducts semi-supervised medical image segmentation with limited labeled-data by leveraging consistency learning. The UAC learns the consistent predictions to improve the segmentation generalization by introducing the multi-level perturbations. Specifically, the proposed UAC consists of three parts: (1) Multi-perturbation strategies explore the perturbation space by introducing various perturbations, uncovering latent information in unlabeled data to help the model learn more robust feature representations; (2) Perturbation uncertainty involves measuring perturbation prediction differences to estimate uncertainty in a single forward pass, which is then used to constrain consistency learning, preventing the model from overly focusing on noisy regions; (3) Voxel contrastive learning involves pulling similar class features closer together in feature space and pushing features from different classes further apart, further enhancing class separability and learning more discriminative features.

3.1. Multi-perturbation strategy for latent structure

The multi-perturbation strategies are designed to help the consistency learning model better understand the subtle structures within organs. Multiple perturbation strategies aim to introduce different types of perturbations, explore the feature space of data more comprehensively during training, uncover important information that influences segmentation performance in the medical image, help the model generalize better to new samples, and improve its performance and robustness.

When using perturbations, there is a balance issue with perturbation intensity. If the perturbation is too weak, the model may become overly reliant on initial predictions or local details, failing to fully leverage the latent information in unlabeled data. Conversely, if the perturbation is too strong, it may disrupt the structure and semantic information of the images, preventing the model from accurately learning effective features and patterns. To fully utilize the latent information in the data

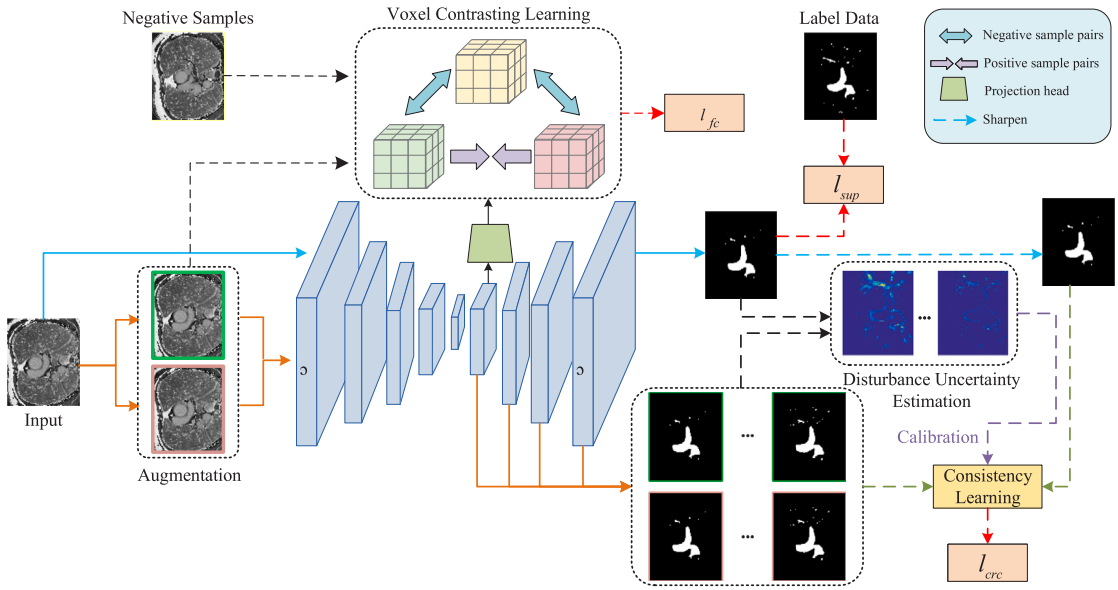

Fig. 3. The proposed UAC consists of three parts: (1) Multi-perturbation strategy is used to explore latent information contained in the data, aiding the model in learning more robust representations. (2) Disturbance uncertainty calibration helps the model prioritize learning reliable knowledge during the learning process by assigning different levels of attention to regions with varying degrees of reliability, thereby avoiding excessive focus on noisy areas. (3) Voxel contrastive loss provides explicit supervision for features, enhancing the separability between features of different classes.

while avoiding the instability and confusion caused by excessive perturbation, finding the right balance is crucial. Fixmatch [16] achieves image classification performance comparable to state-of-the-art methods by employing a simple strong–weak consistency framework. Inspired by this, the strong–weak consistency learning framework has been introduced into medical image segmentation tasks.

As shown in Fig. 3, in the input space, weak augmentations like random brightness transformations and random noise and strong augmentations such as random cropping and color jittering are applied to the input image x, to obtain augmented views x_S and x_W. Considering that common structures in medical images often exist at multiple scales, a pyramid structure is introduced in the decoder to generate multi-scale predictions, denoted as p_c. This helps the model capture structures and details at different levels in medical images. In this segmentation tasks, where $c = \{0, 1, 2, 3\}$, larger c values correspond to higher resolution output results. The multi-scale predictions are up-sampled to obtain outputs of the same size for alignment and comparison.

3.2. Perturbation uncertainty for consistency learning

The perturbation uncertainty is adopted to measure prediction confidence for preventing the model from excessively focusing on noisy areas during the training process. The difference between the multi-scale output predictions of perturbed images and the original predictions is calculated to represent the prediction uncertainty map. The prediction uncertainty map is used both to ensure the model generates reliable predictions and to dynamically adjust consistency learning to prevent the model from overly focusing on unreliable areas in the predictions. Over-focusing on unreliable areas in the predictions may lead to two potential issues: one is the risk of the model learning incorrect organ structure features, impacting segmentation performance, and the other is the potential for the model to overlook or confuse genuine organ structure features, resulting in the loss of crucial information in the predictions. The UAC framework includes the original data stream and the perturbed data stream. For the probability map p of the original input image, it is sharpened to obtain $\hat{p}$ [34]:

$$\hat{p} = \frac{p^{1/T}}{p^{1/T} + (1-p)^{1/T}} \tag{1}$$

Here, T is a hyperparameter that controls the degree of sharpening. It is worth noting that as T decreases, prediction results with low entropy constraint are obtained. However, if T is set too low, it may lead to overconfident predictions, ignoring the model's uncertainty. Choosing an appropriate T can help us achieve better segmentation results. Directly conducting multi-scale consistency learning may introduce unnecessary noise or errors. Therefore, the difference of predicted probability before and after perturbation is utilized to estimate uncertainty maps in a single forward pass:

$$\bar{p} = \frac{p_r + p}{2} \tag{2}$$

$$U_r\left(p_r, p\right) = \bar{p}(\log(\frac{\bar{p}}{p_r}) + log(\frac{\bar{p}}{p})) \tag{3}$$

Here, U_r denotes the difference map between the perturbed output and the original output. A significant difference indicates higher uncertainty in the model's predictions, suggesting a higher likelihood of learning erroneous knowledge. Inspired by the [35], to prevent the model from focusing excessively on noisy regions and incorrect information, the estimated uncertainty maps are use to guide the learning process:

$$l_{crc} = \sum_r \frac{l_{CE}\left(p_r, \hat{p}\right)}{exp\left(U_r\left(p_r, p\right)\right)} + \sum_r U_r\left(p_r, p\right) \tag{4}$$

In Eq. (4), the first term represents the corrected consistency loss obtained even after dynamic adjustment for uncertainty, while the second term calculates the overall uncertainty and uses it as a regularization term. By introducing additional penalty terms, the model is prevented from consistently generating predictions with high uncertainty.

3.3. Voxel contrastive learning for feature discrimination

To further learn a structured feature space, aiding the model in capturing finer-grained feature representations to handle complex structures and ambiguous boundaries. The voxel-level contrastive loss [32] is introduced to provide explicit supervision for features, by bringing similar samples closer together in the feature space and push samples from different classes farther apart, thereby improving the accuracy and robustness of segmentation. Specifically, a projection head is introduced at the second up-sampling block of the decoder to generate

feature representations. Two augmented views are considered as positive samples, while other samples in the dataset serve as negative samples. By maximizing the similarity between positive sample pairs and minimizing the similarity between negative sample pairs in the feature space, it encourages similar samples to have closer feature representations and dissimilar samples to have more dispersed feature representations. The feature pair contrastive loss function is defined as:

$$l_{fc} = -log\frac{exp\left(\frac{sim(r_S,\ r_W)}{\tau}\right)}{exp\left(\frac{sim(r_S,\ r_W)}{\tau}\right) + \sum_{\Lambda^-} exp\left(\frac{sim(r_S,\ r^-)}{\tau}\right)} \tag{5}$$

where r_S, r_W and r^- represent the feature representations of strong and weak augmented views, as well as negative samples. Additionally, $sim(\cdot)$ denotes the similarity between two feature maps and Λ^- is the set of negative sample images. In this paper, cosine distance is used as a metric to measure the similarity between two feature representations. Furthermore, unlabeled images are utilized as negative samples for labeled images, with a buffer used to store dynamically updated negative samples for computing the contrastive loss.

For pixel-level contrastive learning tasks, positive samples are the corresponding pixels in strong and weak augmented views. However, selecting negative samples following the methods used in image-level contrastive learning may lead to resource constraints. Additionally, semi-supervised segmentation tasks are sensitive to noise, where misclassifying a negative sample can lead to ineffective learning or even misguide the model's learning direction. Therefore, a confident negative sampling strategy is employed. Initially, to prevent the selection of pixels belonging to the same class as the positive samples in the negative sample images, a difference matrix is constructed using the segmentation results to select negative pixels. Subsequently, the top K pixels with the highest confidence are sampled from the negative pixels.

3.4. Overall training loss

The UAC is a universal semi-supervised framework where supervised loss and consistency loss are employed in a unified framework for learning from labeled and unlabeled data, and it combines uncertainty-guided multi-perturbation consistency learning with voxel contrastive learning. Specifically, in semi-supervised learning image segmentation, we assume that the dataset used contains only a small amount of labeled data and a lot of unlabeled data, where $D_L = \left\{x^l, y^l\right\}_{l=1}^{N_L}$ represents the labeled data set, $D_U = \{x^u\}_{u=1}^{N_U}$ represents the unlabeled data set, and $N_L \ll N_U$. For annotated data, the supervised loss function is directly computed:

$$l_{sup}(x^l, y^l) = l_{CE}(y, y_i) + l_{Dice}(y, y_i) \tag{6}$$

where l_{CE} represents the calculation of the cross-entropy loss function, and l_{Dice} represents the calculation of the dice loss function. Then, integrating multi-level uncertainty calibration and voxel-level contrastive learning into the proposed framework for unsupervised learning, the overall loss function is given by :

$$l_{total} = l_{sup}(x^l, y^l) + \alpha \cdot l_{crc}(x^l; x^u) + \beta \cdot l_{fc}(x^l; x^u) \tag{7}$$

where l_{sup}, l_{crc} and l_{fc} are the corrected consistency loss defined in Section 3.2 and the pixel-level contrastive loss defined in Section 3.3, respectively. α and β are hyperparameters that balance the loss, and their specific values depend on the specific task.

4. Experiments

To validate the superior performance of the UAC framework in medical image segmentation tasks, experiments were conducted, and the accuracy of segmentation predictions is evaluated on three different public medical datasets. The results are compared with the current state-of-the-art methods.

4.1. Dataset and pre-processing

The LA dataset. The LA benchmark dataset [36] is from the 2018 Left Atrial Segmentation Challenge, which contains 154 3D late gadolinium-enhanced magnetic resonance imaging scans (LGE-MRIs) of 60 patients, with spatial dimensions of either 576 × 576 × 88 or 640 × 640 × 88 pixels. Since the annotation of the test set is not publicly available, following the experimental setting in [10,17,21,33,37], the 100 scans of the training set are divided into 80 scans for training and 20 scans for testing. Before training, the intensity is first normalized to zero mean and unit variance, and then the region of interest (ROI) is cropped based on the label, with an enlarged boundary of 25 pixels. All the training volumes are randomly cropped to get a patch of 112 × 112 × 80 pixels, because of limited computing resources.

The Pancreas-CT Dataset. The Pancreas-CT dataset [38] is from the National Institutes of Health Clinical Center, which contains 82 3D abdominal contrast-enhanced CT scan of 80 patients, of which 53 are males and 27 are females. The CT scans, which have a resolution of 512 × 512 pixels, with different pixel sizes and slice thicknesses between 1.5 and 2.5 mm, were acquired by Philips and Siemens MDCT scanners. According to the experimental setting in [17,21,33], the dataset is randomly split into 20 testing cases and 62 training cases. Following the pre-processing in [17,33], the voxel values are clipped to the range of [−125, 275] Hounsfield Units (HU) and further re-sampled to an isotropic resolution of $1 \times 1 \times 1$ mm^3 firstly. Then the intensity is normalized to zero mean and unit variance, and crop out the region of interest (ROI) based on the label, with enlarged boundary of 25 pixels. All the training volumes are randomly cropped to get a patch of 96 × 96 × 96 pixels, because of limited computing resources.

The PARSE Dataset. The PARESE dataset [39] is from the 2022 Pulmonary Artery Segmentation challenge, which contains a computed tomography pulmonary angiography (CTPA) image set of 203 subjects diagnosed with pulmonary nodular disease, obtained from four different centers. The plane dimensions of the dataset are 512 × 512 pixels, and the number of slices ranges from 228 to 408. The data were divided into three parts, in which the number of training set, validation set, and test set is 100, 30, and 73 respectively, and the only annotated data we can publicly obtain is 100 samples of the training set, so in the training process, it is empirically divided into two parts, of which 80 samples are used as the training set and 20 samples are used as the validation set. For the pre-processing, the volumes are normalized, and the ROI is cropped. And the input of the model is set as a patch of 96 × 96 × 96 pixels.

4.2. Experimental settings

In this study, our experiment is conducted in PyTorch 1.11.0 on an NVIDIA 3090 GPU with fixed random seeds. For data augmentation, weak augmentation (i.e., random intensity augmentation and random noise) and strong augmentation (i.e., cutout) are used. Our model is trained via an SGD optimizer with a momentum of 0.9 and weight decay of 10–4, while the initial learning rate is set to 0.01 and decay with polynomial strategy slowly. The backbone is set as 3D U-Net. The batch size is set as 1, containing two labeled patches and two unlabeled patches. Our model is trained for 200 epochs on LA, 400 epochs on Pancreas-CT and 400 epochs on PARSE respectively. During training, we set $\beta = 0.1$, $\alpha = 0.1$ for LA, $\beta = 0.1$, $\alpha = 0.2$ for pancreas-CT, $\beta = 0.1$, $\alpha = 0.1$ for PARSE. Across three datasets with different label data ratios, the network was trained using the setting T=0.1. This setting was discussed in the ablation experiments in Section 5.3. In the testing stage, three metrics are adopted to evaluate the segmentation performance: Dice similarity coefficients (DSC), 95% Hausdorff Distance (95HD), and Average Symmetric Surface Distance (ASD). Additionally, to gain a better understanding of the trade-offs between performance improvements and computational costs, two metrics, Param and FLOPs, have been introduced to quantify the model's parameter quantity and computational load.

Table 1
The proposed UAC achieves highly accurate organ segmentation performance on three challenging datasets under different semi-supervised settings.

Dataset	Scans used		Metrics			Complexity	
	Labeled	Unlabeled	Dice(%)↑	HD95(voxel)↓	ASD(voxel)↓	Param(M)	FLOPs(G)
LA	8(10%)	72(90%)	91.15	5.27	1.68	6.005	71.530
Pancreas-CT	6(10%)	56(90%)	77.52	13.28	2.72	6.005	63.063
PARSE	8(10%)	72(90%)	78.71	11.12	2.10	6.005	63.063
LA	16(20%)	64(80%)	91.92	4.89	1.55	6.005	71.530
Pancreas-CT	12(20%)	50(80%)	80.92	6.22	1.82	6.005	63.063
PARSE	16(20%)	64(80%)	82.90	7.32	1.42	6.005	63.063

Table 2
Experimental results show that the UAC obtains a competitive performance compared with the SOTA method on the LA dataset. The *bold* indicate the best result.

Method	Scans used		Metrics			Complexity	
	Labeled	Unlabeled	Dice(%)↑	HD95(voxel)↓	ASD(voxel)↓	Param(M)	FLOPs(G)
UA-MT [10]	8(10%)	72(90%)	86.28	18.71	4.63	9.449	47.182
SASSNet [37]			85.22	11.18	2.89	9.443	**46.884**
DTC [17]			87.51	8.23	2.36	9.449	47.182
URPC [26]			85.01	15.37	3.96	**5.885**	69.360
MC-Net+ [21]			88.96	7.93	1.86	9.449	47.182
RCPS [33]			90.73	7.91	2.05	6.004	71.520
Proposed method			**91.15**	**5.27**	**1.68**	6.005	71.530
UA-MT [10]	16(20%)	64(80%)	88.74	8.39	2.32	9.449	47.182
SASSNet [37]			89.16	8.95	2.26	9.443	**46.884**
DTC [17]			89.52	7.07	1.96	9.449	47.182
URPC [26]			88.74	12.73	3.66	**5.885**	69.360
MC-Net+ [21]			91.07	5.84	1.67	9.449	47.182
RCPS [33]			91.21	6.54	1.81	6.004	71.520
Proposed method			**91.92**	**4.89**	**1.55**	6.005	71.530

5. Results and analysis

The proposed UAC framework has demonstrated excellent performance in the field of medical image segmentation, achieving high-precision segmentation results. The effectiveness of this framework in semi-supervised segmentation models has been validated in three aspects: (1) Performance of the UAC is examined on three different medical image datasets, LA, Pancreas-CT, and PARSE, confirming the universality of the UAC framework in semi-supervised medical image segmentation. (2) A comparison with existing methods has revealed the advantages of the proposed UAC framework in medical image segmentation. (3) Ablation study is conducted on the various components of the UAC framework to demonstrate the roles and contributions of different modules.

5.1. Overall segmentation performance

The experiments in Table 1 demonstrate that UAC can achieve accurate segmentation on three medical image datasets, proving the versatility of the UAC framework. Specifically, on the LA dataset with a 10% label data ratio, UAC achieved the best performance across all evaluation metrics, with 91.15% of Dice, 5.27 voxel of HD95, and 1.68 voxel of ASD. As the label data increased to 20%, there is a slight improvement in all evaluation metrics, with Dice only increasing by 0.77%. These results indicate that UAC effectively mines latent information from unlabeled data and enhances the segmentation performance of model through uncertainty-guided consistency learning and pixel-level contrastive learning. It is worth noting that the structures of the pancreas and pulmonary artery are typically more intricate than those of the heart, rendering segmentation tasks on the Pancreas-CT and PARSE datasets more demanding in comparison to the LA dataset. Nevertheless, UAC demonstrated outstanding segmentation performance on these two datasets. Under a 10% label data ratio on the Pancreas-CT and PARSE dataset, UAC achieved 77.52% of Dice, 13.28 voxel of HD95, 2.72 voxel of ASD and 78.71% of Dice, 11.12 voxel of HD95, 2.10 voxel of ASD, respectively. When the label ratio is increased to 20%, there is a respective increase of 3.40% and 4.19% in Dice scores. Compared to the LA dataset, the performance improvement was more significant, likely due to the more challenging nature of the segmentation tasks for the pancreas and pulmonary artery compared to the heart. Additionally, by calculating the Param and FLOPs of the model during testing, it was found that the proposed model has a parameter quantity of only 6.005M. The computational load is dependent on the input size, when we input a $112 \times 112 \times 112$ 3D cardiac image, the FLOPs amount to 71.530G.

5.2. Comparison with SOTA

Extensive experiments are conducted on three publicly available datasets demonstrate the segmentation accuracy of UAC including the LA dataset, Pancreas-CT dataset and PARSE dataset. The proposed method is compared with several existing methods, among which SASS [37], DTC [17] introduce task-level regularization for cross-task consistency learning, UA-MT [10], URPC [26], MC-Net+ [21] use output uncertainty to correct predictions, and RCPS [33] uses kl divergence to correct pseudo-labels. Experiments show that the proposed model has a certain improvement on the segmentation results compared with these state of the art.

Comparison on LA Dataset. The proposed UAC achieved the best results compared to state-of-the-art methods on the LA dataset with two different label data ratios, in terms of three evaluation metrics: Dice, HD95, and ASD. It is observed that varying degrees of improvement compared to the other methods in Table 2. Notably, the UAC incorporates a cross-sample consistency constraint and a cross-scale consistency constraint, which enables the network to capture boundary regions and voxel semantic changes more effectively. As a result, accurate organ contour is achieved without the need for shape consistency constraints during the training process, distinguishing us from methods such as SASS [37] and DTC [10]. Additionally, by leveraging uncertainty to constrain consistency, the issue of noise accumulation associated with unsupervised learning is mitigated. This allows us to extract valuable information from the unlabeled data, even when only

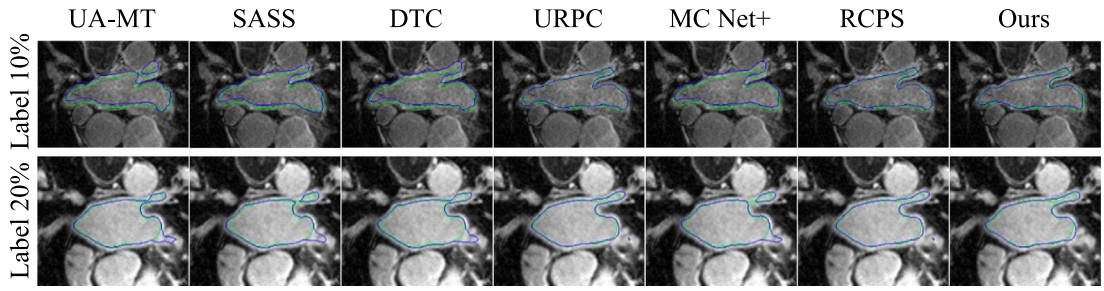

Fig. 4. 2D visualization of our method and comparison method at 10% labeled data and 20% labeled data of LA dataset. The blue line represents the prediction result and the green line represents the true label.

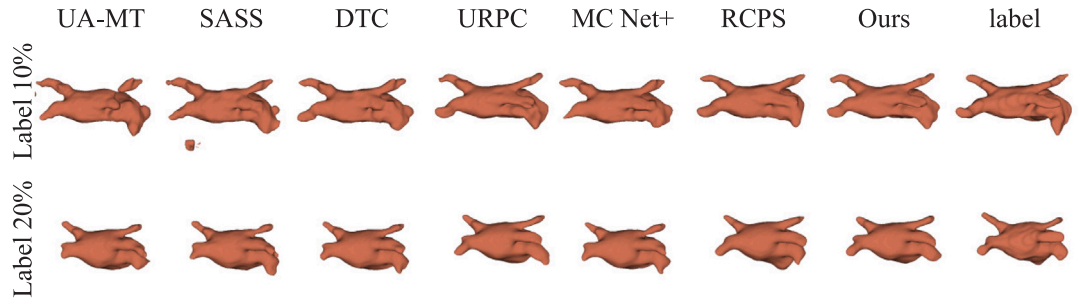

Fig. 5. 3D visualization of our method and comparison method at 10% labeled data and 20% labeled data of LA dataset, the last column is the real label.

a small number of labeled samples are available. In the scenarios with only 10% and 20% labeled data, the UAC achieves Dice values of 91.15% and 91.92%, respectively. Furthermore, compared to other methods, UAC does not introduce excessive inference costs.

It becomes evident that our method produces a more comprehensive left atrial segmentation compared to other methods, closely resembling the ground truth label. The segmentation results obtained by our method exhibit a higher level of accuracy and conformity to the anatomical structure of the left atrium. To visually showcase the segmentation progress of our method and recent semi-supervised medical image segmentation methods, two cases are chosen from the test set to illustrate the segmentation results. In Figs. 4 and 5, the 2D visualization and 3D visualization analyses of the segmentation outcomes on the LA dataset are presented using our method and the comparative method, respectively. Fig. 4 showcases the boundary delineation and internal details of the left atrium. Compared to other methods, the segmentation results of UAC are more accurate at the edges and junctions. Fig. 5 provides a three-dimensional representation of the segmented left atrium, allowing for a more comprehensive evaluation. The UAC achieves a more complete and precise segmentation, capturing the intricate shape and structure of the left atrium.

Comparison on Pancreas-CT Dataset. At a label data ratio of 10% on the Pancreas-CT dataset, UAC achieved the best performance in terms of Dice and HD95, while at a label data ratio of 20%, it achieved the best result in the HD95 metric. As shown in Table 3, compared to other existing state-of-the-art semi-supervised medical image segmentation algorithms, UAC shows varying degrees of improvement in different metrics. It is worth noting that UAC achieves good results even with a label ratio of 10%, indicating that UAC can make better use of the information extracted from unlabeled data. Additionally, in the case of a 20% label ratio, although the segmentation Dice and ASD values of UAC are slightly lower than the best-performing method, HD95 value is still improved, indicating that UAC can achieve more accurate organ contours while maintaining good segmentation accuracy.

To further illustrate the segmentation performance achieved by UAC, two cases are selected from the test set for visualization. Fig. 6 showcases the 2D visualization of the segmentation results. It demonstrates the boundary delineation and internal details of the pancreas, revealing that the segmentation results of UAC have smoother boundaries and shapes that are closer to the ground truth. While RCPS [33] achieves high-precision segmentation, its boundaries and shapes are not as accurate. On the other hand, SASS [37] can produce accurate boundaries, but there are some outliers and significant deviations in its segmentation results, which affect the segmentation performance. Fig. 7 showcases the 3D visualization analysis of the segmentation results. It provides a three-dimensional representation of the segmented pancreas, revealing that UAC captures the complex shape and structure of the pancreas, closely resembling the ground truth label.

Comparison on PARSE Dataset. UAC achieved the best results compared to state-of-the-art methods on the PARSE dataset with two different label data ratios, except for a slightly lower Dice compared to the best result at a 10% label data ratio setting. Table 4 presents the segmentation results of our proposed method and six other state-of-the-art methods on the PARSE dataset. Specifically, at a label data ratio of 20%, UAC achieved the best performance with a Dice of 82.90%, HD95 of 7.32 voxel, and ASD of 1.42 voxel. To further illustrate the segmentation performance achieved by our method, two cases are selected from the test set for visualization. In Fig. 8, the 3D visualization analysis results of the segmentation using our method and the comparative methods on the PARSE dataset are presented. It can be observed that our method generates fewer false positive results compared to the other methods.

5.3. Ablation study

The ablation experiments in Table 5 have demonstrated the effectiveness of each component in UAC. Disturbance uncertainty calibration and voxel contrastive learning were experimented with using 10% labeled data on the LA, Pancreas-CT, and PARSE datasets, respectively. The experimental results indicate that each module improves the model's performance compared to the baseline. Moreover, the perturbation uncertainty has a more significant impact on enhancing model performance compared to voxel contrastive learning. When both modules are used simultaneously, the model achieves the best performance.

Table 3
Experimental results show that the UAC obtains a competitive performance compared with the SOTA method on the Pancreas-CT dataset. The *bold* indicate the best result.

Method	Scans used		Metrics			Complexity	
	Labeled	Unlabeled	Dice(%)↑	HD95(voxel)↓	ASD(voxel)↓	Param(M)	FLOPs(G)
UA-MT [10]	6(10%)	56(90%)	66.44	17.04	3.03	9.449	41.597
SASSNet [37]			68.97	18.83	**1.96**	9.443	**41.334**
DTC [17]			66.58	15.46	4.16	9.449	41.597
URPC [26]			73.53	22.57	7.85	**5.885**	61.150
MC-Net+ [21]			70.00	16.03	3.87	9.449	41.597
RCPS [33]			76.62	16.32	3.01	6.004	63.054
Proposed method			**77.52**	**13.28**	2.72	6.005	63.063
UA-MT [10]	12(20%)	50(80%)	76.10	10.84	2.43	9.449	41.597
SASSNet [37]			76.39	11.06	**1.42**	9.443	**41.334**
DTC [17]			76.27	8.70	2.20	9.449	41.597
URPC [26]			80.02	8.51	1.98	**5.885**	61.150
MC-Net+ [21]			79.37	8.52	1.72	9.449	41.597
RCPS [33]			**81.59**	7.50	2.03	6.004	63.054
Proposed method			80.92	**6.22**	1.82	6.005	63.063

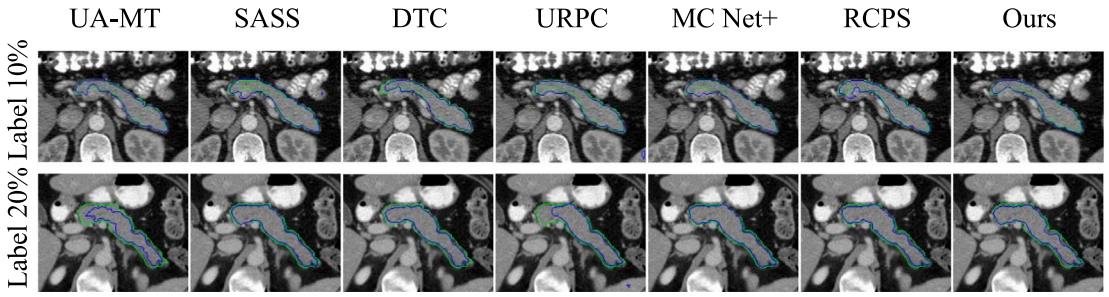

Fig. 6. 2D visualization of our method and comparison method at 10% labeled data and 20% labeled data of Pancreas-CT dataset. The blue line represents the prediction result and the green line represents the true label.

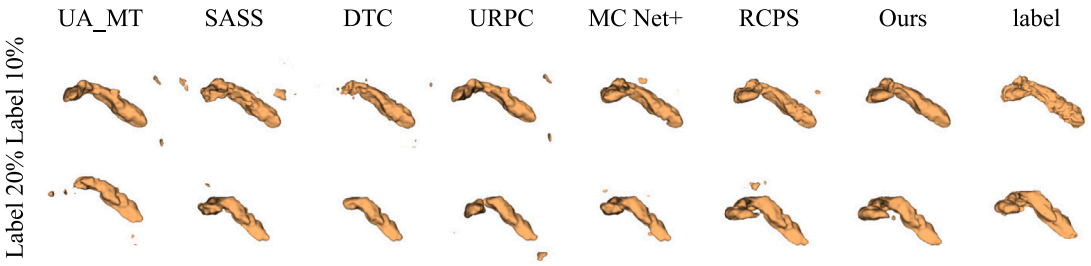

Fig. 7. 3D visualization of our method and comparison method at 10% labeled data and 20% labeled data of Pancreas-CT dataset, the last column is the real label.

Table 4
Experimental results show that the UAC obtains a competitive performance compared with the SOTA method on the PARSE dataset. The *bold* indicate the best result.

Method	Scans used		Metrics			Complexity	
	Labeled	Unlabeled	Dice(%)↑	HD95(voxel)↓	ASD(voxel)↓	Param(M)	FLOPs(G)
UA-MT [10]	8(10%)	72(90%)	61.84	16.04	6.24	9.449	41.597
SASSNet [37]			69.91	13.77	5.18	9.443	**41.334**
DTC [17]			59.20	47.04	2.59	9.449	41.597
URPC [26]			77.19	**9.94**	2.47	**5.885**	61.150
MC-Net+ [21]			77.57	11.38	2.38	9.449	41.597
RCPS [33]			**78.90**	11.48	2.27	6.004	63.054
Proposed method			78.71	11.12	**2.10**	6.005	63.063
UA-MT [10]	16(20%)	64(80%)	62.97	15.21	6.16	9.449	41.597
SASSNet [37]			70.57	12.98	4.38	9.443	**41.334**
DTC [17]			66.34	21.88	3.83	9.449	41.597
URPC [26]			80.37	9.26	2.19	**5.885**	61.150
MC-Net+ [21]			80.31	9.78	2.55	9.449	41.597
RCPS [33]			82.53	8.10	1.50	6.004	63.054
Proposed method			**82.90**	**7.32**	**1.42**	6.005	63.063

To illustrate the necessity of using multi-level perturbations, experiments are conducted separately for single-level perturbation uncertainty and multi-level perturbation uncertainty. As shown in Table 6, using a multi-level perturbation strategy on different datasets led to

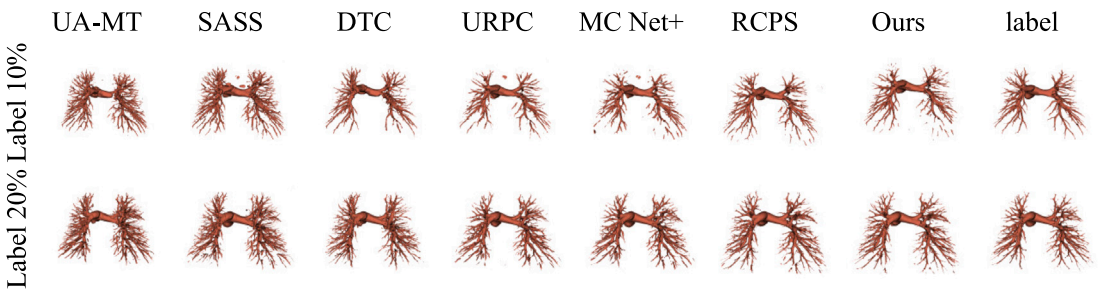

Fig. 8. 3D visualization of our method and comparison method at 10% labeled data and 20% labeled data of PARSE dataset, the last column is the real label.

Table 5
The UAC segmentation performance on different datasets with different methods. Here, PU and VCL respectively represent perturbation uncertainty, voxel contrastive learning.

Dataset	Method	Scans used	Metrics		
		Labeled Unlabeled	Dice(%)↑	HD95(voxel)↓	ASD(voxel)↓
LA	Baseline	8(10%) 72(90%)	74.01	24.42	7.24
	Baseline + VCL		83.49	10.45	2.52
	Baseline + PU		91.01	5.38	1.66
	Baseline + PU + VCL		**91.15**	**5.27**	**1.68**
Pancreas-CT	Baseline	6(10%) 56(90%)	55.45	23.87	7.21
	Baseline + VCL		68.82	26.01	4.74
	Baseline + PU		72.82	20.01	4.05
	Baseline + PU + VCL		**77.52**	**13.28**	**2.72**
PARSE	Baseline	8(10%) 72(90%)	48.74	19.51	5.08
	Baseline + VCL		71.92	17.61	4.09
	Baseline + PU		72.42	12.89	2.94
	Baseline + PU + VCL		**78.71**	**11.12**	**2.10**

Black and bold indicate the best result.

Table 6
The UAC segmentation performance on different datasets with different perturbation strategies. Here, MPU and SPU respectively represent multi-level perturbation uncertainty, single-level perturbation uncertainty.

Dataset	Method	Scans used	Metrics		
		Labeled Unlabeled	Dice(%)↑	HD95(voxel)↓	ASD(voxel)↓
LA	Baseline + SPU	8(10%) 72(90%)	89.75	10.45	4.05
	Baseline + MPU		**91.01**	**5.38**	**1.66**
Pancreas-CT	Baseline + SPU	6(10%) 56(90%)	69.53	32.48	5.34
	Baseline + MPU		**72.82**	**20.01**	**4.05**
PARSE	Baseline + SPU	8(10%) 72(90%)	72.18	13.12	2.98
	Baseline + MPU		**72.42**	**12.89**	**2.94**

Black and bold indicate the best result.

varying degrees of performance improvement for the model. Introducing diverse perturbations can encourage the model to learn more robust and general feature representations, thereby reducing the risk of overfitting.

In order to illustrate the effect of the sharpening degree of the original prediction on the performance in the model, different sizes of T are selected for ablation experiments. The results shown in Fig. 9 indicate that choosing an appropriate temperature parameter when generating soft pseudo-labels for training can improve the performance of the model. A higher temperature parameter makes the probability distribution of the model smoother, with probabilities of different classes being closer to each other. However, this may result in blurred boundaries and loss of details. On the other hand, a lower temperature parameter sharpens the probability distribution, emphasizing classes with higher confidence. This helps highlight the edges and details. If the temperature parameter is set too low, it may lead to overconfident predictions, disregarding the model's uncertainty.

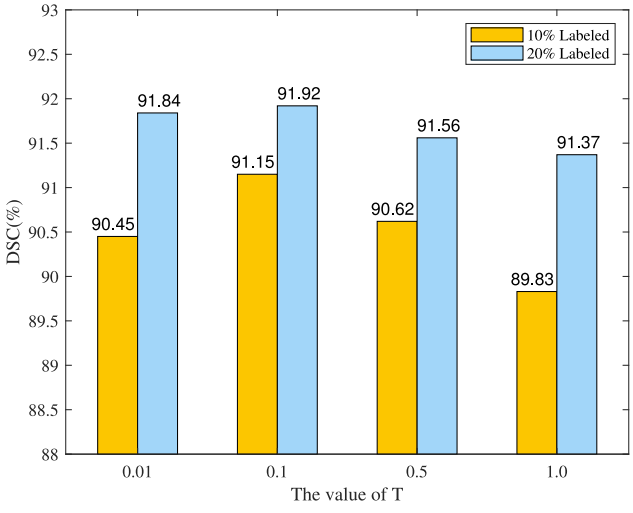

Fig. 9. Segmentation results on the LA dataset with different T settings. The model achieves the best segmentation performance when T=0.1.

5.4. Limitations and future work

Although our semi-supervised model has demonstrated excellent performance in segmenting images of multiple organs, our current

discussions have been focused on single-domain datasets. In the context of semi-supervised settings in the text, both labeled and unlabeled data come from the same domain. However, in real life, the datasets we acquire may originate from multiple medical centers, leading to domain discrepancies and inconsistencies in annotations between different centers. In our future work, we aim to further investigate how semi-supervised methods perform when there are domain differences between labeled and unlabeled data, and integrate cross-domain learning and domain adaptation techniques to enhance the model's generalization capabilities.

6. Conclusion

In this paper, UAC is propose to address the issues of noise sensitivity and insufficient information in semi-supervised medical image segmentation. The proposed UAC aims to better utilize the information contained in unlabeled data through consistency learning by exploring a larger perturbation space. A multi-level perturbation strategy is introduce to enhance model generalization and address the issue of low prediction confidence in challenging areas by proposing disturbance uncertainty estimation. This dynamically adjusts the constraints on consistency to prevent the model from overly focusing on noisy regions. Experimental results on three public datasets demonstrate that our method exhibits excellent segmentation performance compared to state-of-the-art semi-supervised medical image segmentation methods.

CRediT authorship contribution statement

Min Dong: Writing – review & editing, Supervision, Resources, Project administration, Methodology, Conceptualization. **Ating Yang:** Writing – original draft, Software, Methodology. **Zhenhang Wang:** Formal analysis, Data curation. **Dezhen Li:** Investigation, Formal analysis. **Jing Yang:** Visualization, Validation. **Rongchang Zhao:** Writing – review & editing, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No. U21B2037), the National Natural Science Foundation of China (Grant No. 62376253), the Education Ministry's Collaborative Education Program with Industry, China (Grand No. 231003226205059), the Henan Provincial Science and Technology Research Project, China (Grant No. 242102210002), the Zhengzhou's Science and Technology Innovation Guidance Program in Healthcare, China (Grant No. 2024YLZDJH026), the National Natural Science Foundation of China (Grant No. 62372474) and the 111 Project, China (Grant No. B18059).

Data availability

The data used in this study are publicly available.

References

[1] R. Zhao, Y. Wang, Dual gradient alignment for unsupervised domain adaptation on optic disc and cup segmentation, in: 2022 IEEE International Conference on Bioinformatics and Biomedicine, BIBM, IEEE, 2022, pp. 1405–1410.
[2] Q. Zhou, J. Qin, X. Xiang, Y. Tan, Y. Ren, MOLS-net: Multi-organ and lesion segmentation network based on sequence feature pyramid and attention mechanism for aortic dissection diagnosis, Knowl.-Based Syst. 239 (2022) 107853.
[3] S. Li, M. Dong, G. Du, X. Mu, Attention dense-u-net for automatic breast mass segmentation in digital mammogram, IEEE Access 7 (2019) 59037–59047.
[4] S. Ahmad, Z. Ullah, J. Gwak, Multi-teacher cross-modal distillation with cooperative deep supervision fusion learning for unimodal segmentation, Knowl.-Based Syst. (2024) 111854.
[5] R. Su, X. Liu, Q. Jin, X. Liu, L. Wei, Identification of glioblastoma molecular subtype and prognosis based on deep MRI features, Knowl.-Based Syst. 232 (2021) 107490.
[6] Z. Chen, Y. Hou, H. Liu, Z. Ye, R. Zhao, H. Shen, FDCT: Fusion-guided dual-view consistency training for semi-supervised tissue segmentation on MRI, Comput. Biol. Med. 160 (2023) 106908.
[7] J. Ma, Y. He, F. Li, L. Han, C. You, B. Wang, Segment anything in medical images, Nature Commun. 15 (1) (2024) 654.
[8] J. Wu, R. Fu, H. Fang, Y. Liu, Z. Wang, Y. Xu, Y. Jin, T. Arbel, Medical sam adapter: Adapting segment anything model for medical image segmentation, 2023, arXiv preprint arXiv:2304.12620.
[9] B. Lakshminarayanan, A. Pritzel, C. Blundell, Simple and scalable predictive uncertainty estimation using deep ensembles, Adv. Neural Inf. Process. Syst. 30 (2017).
[10] L. Yu, S. Wang, X. Li, C.-W. Fu, P.-A. Heng, Uncertainty-aware self-ensembling model for semi-supervised 3D left atrium segmentation, in: Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part II 22, Springer, 2019, pp. 605–613.
[11] Y. Zhang, R. Jiao, Q. Liao, D. Li, J. Zhang, Uncertainty-guided mutual consistency learning for semi-supervised medical image segmentation, Artif. Intell. Med. 138 (2023) 102476.
[12] D.-H. Lee, et al., Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks, in: Workshop on Challenges in Representation Learning, ICML, Vol. 3, Atlanta, 2013, p. 896.
[13] P. Cascante-Bonilla, F. Tan, Y. Qi, V. Ordonez, Curriculum labeling: Revisiting pseudo-labeling for semi-supervised learning, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 35, 2021, pp. 6912–6920.
[14] Y. Wang, H. Chen, Q. Heng, W. Hou, Y. Fan, Z. Wu, J. Wang, M. Savvides, T. Shinozaki, B. Raj, et al., Freematch: Self-adaptive thresholding for semi-supervised learning, 2022, arXiv preprint arXiv:2205.07246.
[15] E. Arazo, D. Ortego, P. Albert, N.E. O'Connor, K. McGuinness, Pseudo-labeling and confirmation bias in deep semi-supervised learning, in: 2020 International Joint Conference on Neural Networks, IJCNN, IEEE, 2020, pp. 1–8.
[16] K. Sohn, D. Berthelot, N. Carlini, Z. Zhang, H. Zhang, C.A. Raffel, E.D. Cubuk, A. Kurakin, C.-L. Li, Fixmatch: Simplifying semi-supervised learning with consistency and confidence, Adv. Neural Inf. Process. Syst. 33 (2020) 596–608.
[17] X. Luo, J. Chen, T. Song, G. Wang, Semi-supervised medical image segmentation through dual-task consistency, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 35, 2021, pp. 8801–8809.
[18] Z. Chen, Y. Xiong, H. Wei, R. Zhao, X. Duan, H. Shen, Dual-consistency semi-supervision combined with self-supervision for vessel segmentation in retinal OCTA images, Biomed. Opt. Express 13 (5) (2022) 2824–2834.
[19] G. Bortsova, F. Dubost, L. Hogeweg, I. Katramados, M. De Bruijne, Semi-supervised medical image segmentation via learning consistency under transformations, in: Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part VI 22, Springer, 2019, pp. 810–818.
[20] S. Jiang, H. Wu, J. Chen, Q. Zhang, J. Qin, PH-net: Semi-supervised breast lesion segmentation via patch-wise hardness, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024, pp. 11418–11427.
[21] Y. Wu, Z. Ge, D. Zhang, M. Xu, L. Zhang, Y. Xia, J. Cai, Mutual consistency learning for semi-supervised medical image segmentation, Med. Image Anal. 81 (2022) 102530.
[22] M. Chen, C. Wang, Multi-head co-training: An uncertainty-aware and robust semi-supervised learning framework, Knowl.-Based Syst. 302 (2024) 112325.
[23] Y. Wang, Y. Zhang, J. Tian, C. Zhong, Z. Shi, Y. Zhang, Z. He, Double-uncertainty weighted method for semi-supervised learning, in: Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part I 23, Springer, 2020, pp. 542–551.
[24] K. Wang, B. Zhan, C. Zu, X. Wu, J. Zhou, L. Zhou, Y. Wang, Semi-supervised medical image segmentation via a tripled-uncertainty guided mean teacher model with contrastive learning, Med. Image Anal. 79 (2022) 102447.

[25] J. Wu, G. Wang, R. Gu, T. Lu, Y. Chen, W. Zhu, T. Vercauteren, S. Ourselin, S. Zhang, Upl-sfda: Uncertainty-aware pseudo label guided source-free domain adaptation for medical image segmentation, IEEE Trans. Med. Imaging (2023).
[26] X. Luo, W. Liao, J. Chen, T. Song, Y. Chen, S. Zhang, N. Chen, G. Wang, S. Zhang, Efficient semi-supervised gross target volume of nasopharyngeal carcinoma segmentation via uncertainty rectified pyramid consistency, in: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part II 24, Springer, 2021, pp. 318–329.
[27] Y. Shi, C. Zu, P. Yang, S. Tan, H. Ren, X. Wu, J. Zhou, Y. Wang, Uncertainty-weighted and relation-driven consistency training for semi-supervised head-and-neck tumor segmentation, Knowl.-Based Syst. 272 (2023) 110598.
[28] R. Zhao, X. Chen, Z. Chen, S. Li, Diagnosing glaucoma on imbalanced data with self-ensemble dual-curriculum learning, Med. Image Anal. 75 (2022) 102295.
[29] K. He, H. Fan, Y. Wu, S. Xie, R. Girshick, Momentum contrast for unsupervised visual representation learning, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 9729–9738.
[30] T. Chen, S. Kornblith, M. Norouzi, G. Hinton, A simple framework for contrastive learning of visual representations, in: International Conference on Machine Learning, PMLR, 2020, pp. 1597–1607.
[31] H. Dong, X. Long, Y. Li, Rethinking samples selection for contrastive learning: Mining of potential samples, Knowl.-Based Syst. (2024) 111979.
[32] Y. Zhong, B. Yuan, H. Wu, Z. Yuan, J. Peng, Y.-X. Wang, Pixel contrastive-consistent semi-supervised semantic segmentation, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 7273–7282.
[33] X. Zhao, Z. Qi, S. Wang, Q. Wang, X. Wu, Y. Mao, L. Zhang, Rcps: Rectified contrastive pseudo supervision for semi-supervised medical image segmentation, IEEE J. Biomed. Health Inf. (2023).
[34] Q. Xie, Z. Dai, E. Hovy, T. Luong, Q. Le, Unsupervised data augmentation for consistency training, Adv. Neural Inf. Process. Syst. 33 (2020) 6256–6268.
[35] Z. Zheng, Y. Yang, Rectifying pseudo label learning via uncertainty estimation for domain adaptive semantic segmentation, Int. J. Comput. Vis. 129 (4) (2021) 1106–1120.
[36] Z. Xiong, Q. Xia, Z. Hu, N. Huang, C. Bian, Y. Zheng, S. Vesal, N. Ravikumar, A. Maier, X. Yang, et al., A global benchmark of algorithms for segmenting the left atrium from late gadolinium-enhanced cardiac magnetic resonance imaging, Med. Image Anal. 67 (2021) 101832.
[37] S. Li, C. Zhang, X. He, Shape-aware semi-supervised 3D semantic segmentation for medical images, in: Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part I 23, Springer, 2020, pp. 552–561.
[38] H.R. Roth, A. Farag, E. Turkbey, L. Lu, J. Liu, R.M. Summers, Data from pancreas-ct. the cancer imaging archive, IEEE Trans. Image Process. 10 (2016) K9.
[39] G. Luo, K. Wang, J. Liu, S. Li, X. Liang, X. Li, S. Gan, W. Wang, S. Dong, W. Wang, et al., Efficient automatic segmentation for multi-level pulmonary arteries: The PARSE challenge, 2023, arXiv preprint arXiv:2304.03708.