SCIENCE CHINA
Information Sciences

November 2023, Vol. 66 210104:1–210104:15
https://doi.org/10.1007/s11432-022-3871-0

• RESEARCH PAPER •
Special Topic: Deep Learning for Computer Vision

Confidence-weighted mutual supervision on dual networks for unsupervised cross-modality image segmentation

Yajie CHEN[1*], Xin YANG[1*] & Xiang BAI[2*]

[1]*School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan* 430074*, China;*
[2]*School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan* 430074*, China*

Received 27 November 2022/Revised 5 June 2023/Accepted 2 August 2023/Published online 26 October 2023

Abstract The unsupervised cross-modality image segmentation has gained much attention. Many methods attempt to align different modalities via adversarial learning. Recently, self-training with pseudo labels for the unsupervised target modality has also been widely used and achieved very promising results. The pseudo labels are usually obtained by selecting reliable predictions whose highest predicted probability is larger than an empirically set value. Such pseudo label generation inevitably has noise and training a segmentation model using incorrect pseudo labels could yield nontrivial errors for the target modality. In this paper, we propose a confidence-weighted mutual supervision on dual networks for unsupervised cross-modality image segmentation. Specifically, we independently initialize two networks with the same architecture, and propose a novel confidence-weighted Dice loss to mutually supervise the two networks using their predicted results for unlabeled data. In this way, we make full use of all predictions of unlabeled images and leverage the prediction confidence to alleviate the negative impact of noisy pseudo labels. Extensive experiments on three widely-used unsupervised cross-modality image segmentation datasets (i.e., MM-WHS 2017, Brats 2018, and Multi-organ segmentation) demonstrate that the proposed method achieves superior performance to some state-of-the-art methods.

Keywords domain adaptation, pseudo label, mutual supervision, cross-modality, image segmentation

Citation Chen Y J, Yang X, Bai X. Confidence-weighted mutual supervision on dual networks for unsupervised cross-modality image segmentation. Sci China Inf Sci, 2023, 66(11): 210104, https://doi.org/10.1007/s11432-022-3871-0

1 Introduction

Semantic image segmentation has achieved great progress [1–10] thanks to the development of deep learning. Most of the existing semantic segmentation methods heavily rely on a large amount of precisely annotated data, which is expensive to obtain. Besides, the presence of different imagining modalities (e.g., visible and infrared light images, different medical imaging modalities) categorized by the method in which images are generated further exacerbates the annotation expense. Therefore, there has been a growing interest in the field of unsupervised cross-modality image segmentation, which aims to achieve segmentation without relying on manual annotation specifically for the target modality, instead using only manual annotations from the source modality.

Most methods frame the unsupervised cross-modality image segmentation as an unsupervised domain adaptation (UDA) segmentation task. These UDA methods aim at performing domain adaptation by transferring the knowledge learned from labeled source data to unlabeled target data with very different image appearances. Image translation methods [6, 11, 12] that attempt to align the image distribution between the source and the target domain via adversarial training are widely adopted in UDA segmentation tasks. Since the source and target data share some common features (such as semantic content

* Corresponding author (email: yajiechen@hust.edu.cn, xinyang2014@hust.edu.cn, xbai@hust.edu.cn)

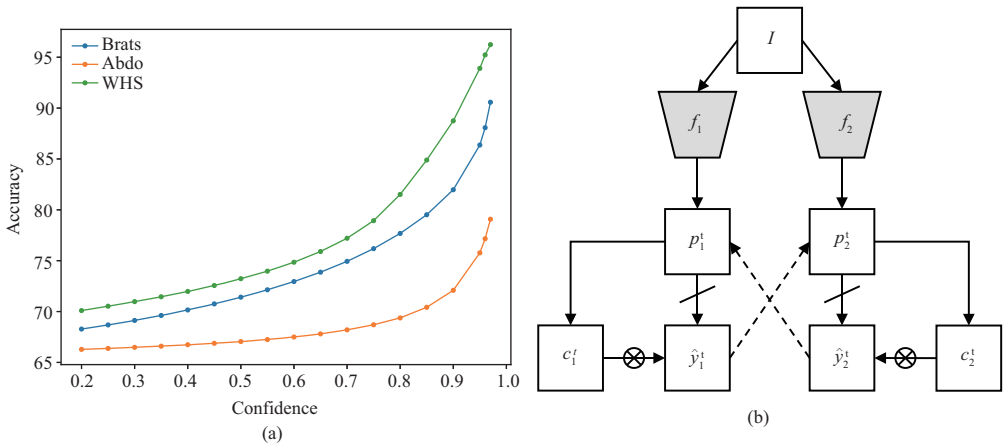

Figure 1 (a) Evolution of accuracy on whole heart segmentation (WHS), brain tumor segmentation (Brats), and abdominal multi-organ segmentation (Abdo), with respect to the confidence threshold used for extracting reliable pseudo labels. (b) Schematic view of the proposed confidence-weighted mutual supervision on dual networks. $\rightarrow$, $\dashrightarrow$, $\nrightarrow$, and $\otimes$ denote forward, backward propagation, gradient stop, and pixel-wise multiplication, respectively. We leverage the prediction confidence c to alleviate the impact of inevitable noisy pseudo labels, while making full use of all pseudo labels.

and geometric shape), some other methods [13–15] alternate to align intermediate features to learn domain invariant representation for UDA segmentation. In addition to these image/feature-level alignment methods, some other methods [15–18] directly align the structured segmentation output, which preserves more sufficient spatial complexity and rich semantic information for the segmentation tasks [19].

The above alignment-based UDA semantic segmentation methods mainly focus on reducing the domain gap, rather than learning precise segmentation networks for unlabeled target domains. To fully utilize the unlabeled target data, self-training methods [20–22] that rely on pseudo labels as supervision for the unlabeled target data have been commonly used. Self-ensembling methods [20, 21] use the pseudo labels generated from the teacher model to train the student model. The bias toward the dominant semantic classes in pseudo labels is still inevitable. Some methods [23–25] aim at tackling such class-imbalance issues in the pseudo labels.

Thanks to the pseudo labels of extended data from the target domain [20, 21, 23–25], self-training methods have achieved very encouraging results in UDA semantic segmentation. However, without the exact supervision information from precise mask annotation, self-training-based UDA segmentation methods are still prone to generate noisy labels. This can cause the error accumulation in the network [26], degrading segmentation results.

To tackle the issue of noisy pseudo labels, CCT [27] learns discriminative features from different views to improve the accuracy of pseudo labels. FixMatch [28] alternates to learn from strong augmented unlabeled data using the selected pseudo labels of the weakly augmented one. Though these methods [27, 28] do alleviate the harm from the noisy pseudo label, they are still prone to mistake the noisy pseudo labels as the clean one due to the confirmation bias of a single network [26], and neglect the risk that the network memorizes noisy label during training processing [29, 30].

In this paper, we propose the confidence-weighted mutual supervision on dual networks to address the problem of noisy labels for unsupervised cross-modality image segmentation. Specifically, pioneered by the work of deep mutual learning in [31], we make use of dual networks initialized independently based on [32] to get rid of some noisy pseudo labels caused by the confirmation bias [26] in a single network. We further introduce a confidence measure based on the entropy of the predicted score, to assess the quality of pseudo labels. Figure 1 shows that for most cases as the confidence increases, the segmentation accuracy gets better, implying that predictions with higher confidence are more likely to provide correct pseudo labels. However, entropy-based confidence could still contain noise, in particular in the setting of UDA segmentation in which the domain shift increases the difficulties in pseudo label generation and in turn yield unreliable entropy-based measure. To address this problem, we propose a novel confidence-weighted Dice loss on the unlabeled target data to mutually supervise the dual networks. In this way, we make full use of the pseudo labels, and effectively alleviate the negative impact of noisy pseudo labels.

The main contributions of this paper are threefolds.

(1) We propose a novel approach for unsupervised cross-modality image segmentation by introducing

dual networks with confidence-weighted mutual supervision. To the best of our knowledge, this is the first application of the mutual supervision on dual networks to address the challenges of unsupervised cross-modality image segmentation.

(2) We introduce a unique confidence-weighted Dice loss, which enables confidence-weighted mutual supervision. This innovative loss function allows us to effectively leverage pseudo labels while considering their quality.

(3) We validate the proposed method in segmenting cross-modality medical images, which often exhibit shared anatomical structures and larger gaps compared with the domain gap present in cross-domain natural images. Through extensive experiments, our approach achieves state-of-the-art results on three popular unsupervised cross-modality image segmentation tasks.

2 Related work

The unsupervised cross-modality image segmentation is usually framed as UDA semantic segmentation. Therefore, we mainly focus on reviewing related UDA semantic segmentation methods, which can be roughly categorized into two classes: alignment-based and self-learning-based methods.

2.1 Alignment-based UDA semantic segmentation

The source and target domains usually exhibit different appearance distributions. Many methods attempt to achieve UDA semantic segmentation by reducing such domain gaps caused by distribution shifts. For that, most methods mainly align source and target distribution in three-levels: image/feature/output level. Specifically, some methods [25, 33] aim to transfer the labeled source images to unlabeled target distribution. The image distribution alignment is usually achieved by image translation networks such as CycleGAN [11] and DSFN [12]. Yang and Soatto [33] transferred the source image to the target style by using the high-frequency component of the Fourier frequency spectrum of target images. In addition to the image level alignment, the feature level alignment [13–15, 34] is also widely used to learn invariant features for both source and target domains. These methods [13–15, 34] usually perform feature level alignment using adversarial training. Recently, output level alignment is widely used in many methods [15–18, 35, 36]. Directly aligning the structured output achieves impressive results. Some other methods [37–41] perform both image level and output level alignment. Such a combination learns the unified input and output distribution relationship between the source domain and the target domain [38].

The proposed method also adopts CycleGAN [11] to reduce the domain gap between the source and target domain by aligning the image appearance distribution. Different from these alignment-based methods, we propose a confidence-weighted mutual supervision on dual networks to make full use of the pseudo labels, yielding many improved results.

2.2 Self-training methods

Recently, self-training [20, 21, 42] that relies on pseudo labels is widely adopted for unsupervised domain adaptive semantic segmentation. Most methods [20,21,43] generate the pseudo labels on unlabeled target images by thresholding the predicted probability or predicted confidence with an empirically set value. The pseudo labels are then used to re-train the segmentation model on the unlabeled target images. For instance, self-ensembling [20] utilizes the pseudo labels generated from an online updated teacher model to supervise the student model trained on the labeled source data. Li et al. [21] first selected reliable pseudo labels based on the consistency between the predictions on different image views. The selected pseudo labels are then used for retraining on the unlabeled target data. In addition to the thresholding-based approaches, some methods [23–25] aim at tackling the pseudo labels bias among the semantic classes.

Self-training is also popularized in semi-supervised learning. For example, FixMatch [28] and CCT [27] deal with noisy pseudo labels to avoid the error flow in the iteration. FixMatch [28] combats the noisy supervision information by learning from both strongly and weakly perturbed unlabeled data under the supervision of thresholded pseudo labels. CCT [27] de-noises the pseudo labels by learning consistent features from two perturbed networks. Following the pioneer work of deep mutual learning in [8], CPS [32] improves CCT [27] by learning multiple features using pseudo labels produced from two perturbed networks with the same architecture. The method in [44] extends CPS [32] by adopting one shared encoder

and two segmentation heads, and using strongly and weakly augmented unlabeled data. Fan et al. [44] selected the pseudo labels based on the confidence of the prediction to supervise the prediction on strongly augmented data. Specifically, in [44], for each pixel, only the prediction with higher confidence is selected as the pseudo labels to supervise the other network with lower confidence on that pixel using the cross-entropy loss.

Compared with those UDA semantic segmentation methods based on self-training, the proposed method leverages dual networks to alleviate the negative impact of noisy pseudo labels predicted by the single network. Besides, we do not threshold the predicted probability to get somehow reliable pseudo labels. Instead, we make full use of all pseudo labels by considering the confidence of prediction. The most related studies are [32,44] dedicated for semi-supervised semantic segmentation. The proposed method adopts mutual supervision on dual networks for unsupervised cross-modality image segmentation. Besides, we introduce a novel confidence-weighted Dice loss to consider the quality of all pseudo labels for image segmentation. In this way, we get rid of some noisy pseudo labels in mutual supervision on dual networks, yielding improved segmentation results.

3 Method

Problem setting. Let $X^{\mathrm{S}} = \{x_j^{\mathrm{s}}\}_{j=1}^{N_{\mathrm{S}}}$ denote the set of source modality images with the corresponding ground-truth segmentation label $Y^{\mathrm{S}} = \{y_j^{\mathrm{s}}\}_{j=1}^{N_{\mathrm{S}}}$, and $X^{\mathrm{T}} = \{x_j^{\mathrm{t}}\}_{j=1}^{N_{\mathrm{T}}}$ denote the unlabeled target modality images. The goal is to accurately segment the target modality images. To put it simply, we train the segmentation network f with the labeled source data $(X^{\mathrm{S}}, Y^{\mathrm{S}})$ and unlabeled target data X_{T}, and aim to make such trained segmentation network f generalize well to the unseen target modality images.

3.1 Overview

Unsupervised cross-modality image segmentation is a very challenging task. Most methods draw inspiration from unsupervised domain adaptive semantic segmentation, where self-training that leverages pseudo labels for network training is widely used. Based on the observation that the prediction with a higher probability is more likely to be correct, many self-training methods [20, 28] only select reliable pseudo labels whose predicted probability is higher than an empirically set threshold value. Despite the encouraging results, recent self-training methods still suffer from noisy pseudo labels existing in prediction with high probability, due to the confirmation bias in a single network [26]. This may lead the network to overfit some incorrect pseudo labels, resulting in degraded segmentation accuracy.

To cope with the confirmation bias, some semi-supervised methods [27,32] train two independent networks with mutual supervision for unlabeled data to alleviate the bias of noisy pseudo labels given by a single network. Such a mutual supervision strategy reduces the negative effect caused by noisy pseudo labels. Yet, it is still inevitable for the network to memorize some noisy pseudo labels [29,30] during the training process. Besides, the unreliable prediction with low probability may also contain some useful information [45, 46]. Intuitively, a more sophisticated way is to take into account the quality of pseudo labels in mutual supervision on dual networks. Therefore, we adopt a dual network (described in Subsection 3.3), and design a confidence measure based on the entropy of predicted probability distribution to assess the quality of pseudo labels. We then propose a confidence-weighted Dice loss (detailed in Subsection 3.4) to mutually supervise the dual networks. This helps to combat against the noisy pseudo labels and further explore the useful information in the noisy pseudo labels. To further improve the quality of the pseudo labels in cross-modality image segmentation, we also apply image translation based on CycleGAN (see Subsection 3.2) before the self-training process. The overall pipeline is depicted in Figure 2.

3.2 Image translation

Images of different modalities often have very different appearances. To reduce the domain gap between the source modality images and the target modality images, we transfer the source modality images to the style of the target modality images using CycleGAN [11]. Concretely, we first translate the source images to the target style by adversarial training based on the generator $G_{\mathrm{S}\to\mathrm{T}}$ and discriminator D_{T}. Similarly, we train the target to source translation networks based on the generator $G_{\mathrm{T}\to\mathrm{S}}$ and discriminator D_{S}. We also enforce the preservation of source content.

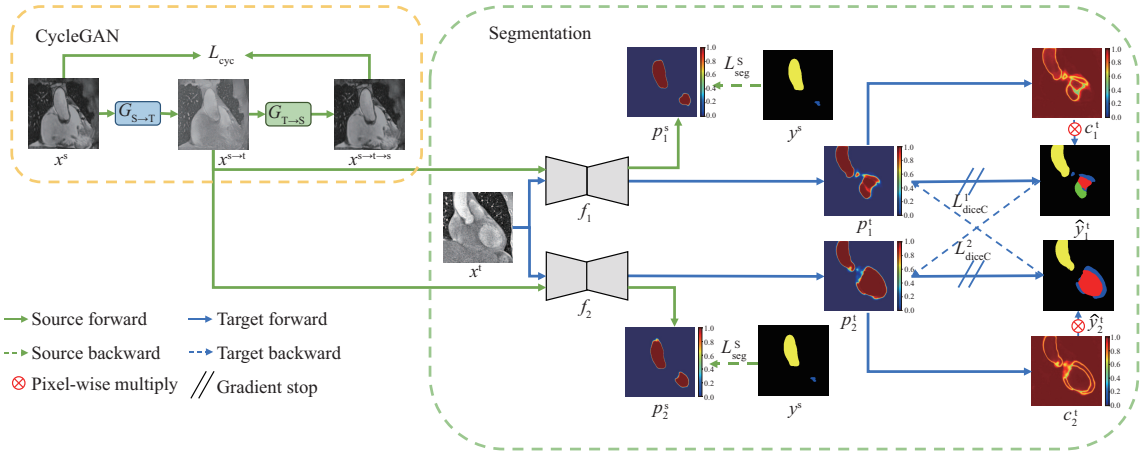

Figure 2 The pipeline of the proposed method, consisting of a CycleGAN module and dual segmentation networks f_1 and f_2 with different initialization. For the source modality images of the target style, we adopt the classical segmentation loss $L_{\text{seg}}^{\text{S}}$ to supervise the prediction p^{s} of both networks. For the target modality images, we introduce a confidence measure c^{t} and design a confidence-weighted Dice loss L_{diceC} based on the pseudo labels $\hat{y}^{\text{t}}$ to mutually supervise the dual networks.

3.3 Segmentation network architecture

The family of variants of U-Net [2] achieves good results in medical image segmentation. Though some variants [47,48] achieve better results than the original U-Net on some specific tasks, it has been shown that the original U-Net still achieves good results in most cases [49]. Therefore, we simply adopt the de facto U-Net for the dual segmentation networks f_1 and f_2. A stronger segmentation network may lead to better results. Yet, since the main contribution of the proposed approach does not lie in the segmentation network, we do not seek for stronger segmentation network. It is noteworthy that the dual networks have the same U-Net architecture, but are initialized differently.

3.4 Confidence-weighted mutual supervision

We online feed the source modality images with the target style and target modality images to the dual networks during training. Let $p^{\text{s}} = f(x^{\text{s}})$ and $p^{\text{t}} = f(x^{\text{t}})$ denote the predicted probability map for the source and target images, respectively. $\hat{y}^{\text{t}} = \text{argmax}(p^{\text{t}})$ stands for the pseudo labels of the target image. The training objective for the source and target modality images is detailed in the following.

Training objective for the target style-like source modality images. Since the source images have accurate annotation, we adopt the classical cross-entropy loss L_{ce} and Dice loss L_{dice} to train both networks f_1 and f_2. Specifically, for a given source image, the corresponding losses are given bellow:

$$L_{\text{ce}}(p^{\text{s}}, y^{\text{s}}) = -\frac{1}{H \times W} \sum_{i=1}^{H \times W} \sum_{k=0}^{K} y_k^{\text{s}}(i) \times \log p_k^{\text{s}}(i), \tag{1}$$

$$L_{\text{dice}}(p^{\text{s}}, y^{\text{s}}) = \frac{1}{K} \sum_{k=0}^{K} \left(1 - 2 \times \frac{\sum_{i=1}^{H \times W} y_k^{\text{s}}(i) \times p_k^{\text{s}}(i) + \gamma}{\sum_{i=1}^{H \times W} (y_k^{\text{s}}(i) + p_k^{\text{s}}(i)) + \gamma} \right), \tag{2}$$

where H, W, and K denote the height, width of the image, and the number of foreground semantic classes, respectively. γ is used to prevent the zero division and is set to 0.0001. For both networks, the overall segmentation loss $L_{\text{seg}}^{\text{S}}$ for the source images is given by

$$L_{\text{seg}}^{\text{S}} = \frac{1}{N_{\text{S}}} \sum_{x^{\text{s}} \sim X^{\text{S}}} \left(L_{\text{ce}}(p^{\text{s}}, y^{\text{s}}) + L_{\text{dice}}(p^{\text{s}}, y^{\text{s}}) \right). \tag{3}$$

Training objective for the target modality images. Since there is no annotation for the target modality images, we leverage the pseudo labels $\hat{y}_1^{\text{t}}$ and $\hat{y}_2^{\text{t}}$ of the dual networks to mutually supervise each other. Specifically, in the early stage of training, both networks are not well optimized. The pseudo labels are not very reliable. Therefore, we simply apply mutual supervision using the Dice loss based

on the pseudo labels. More precisely, in the early stage, the dual network f_1 and f_2 are optimized with $L_1^{\mathrm{t}} = \frac{1}{N_{\mathrm{T}}} \sum_{x^{\mathrm{t}} \sim X^{\mathrm{T}}} L_{\mathrm{dice}}(p_1^{\mathrm{t}}, \hat{y}_2^{\mathrm{t}})$ and $L_2^{\mathrm{t}} = \frac{1}{N_{\mathrm{T}}} \sum_{x^{\mathrm{t}} \sim X^{\mathrm{T}}} L_{\mathrm{dice}}(p_2^{\mathrm{t}}, \hat{y}_1^{\mathrm{t}})$, respectively.

With the process of training, the quality of the pseudo labels for target modality images becomes more and more reliable. Yet, both pseudo labels of the dual networks contain some noises. The straight forward mutual supervision may make the network to memorize the noisy label, leading to degraded segmentation performance on the unlabeled target data. To alleviate such issues of noisy labels in the mutual supervision, we start to take into account the quality of pseudo labels in the late training stage. For that, we define a confidence map c^{t} based on the entropy E of the predicted probability map p^{t}. For the i-th pixel, the entropy $E^{\mathrm{t}}(i)$ is given by

$$E^{\mathrm{t}}(i) = -\sum_{k=0}^{K} p_k^{\mathrm{t}}(i) \times \log p_k^{\mathrm{t}}(i). \tag{4}$$

The confidence map $c^{\mathrm{t}}(i)$ on the i-th pixel is defined as follows:

$$c^{\mathrm{t}}(i) = \frac{2}{1 + \exp\left(E^{\mathrm{t}}(i)\right)}, \tag{5}$$

which is in the range (0, 1). In the late training stage, e.g., starting from the n_{e}-th epoch, as shown in Figure 2, we pay more attention to confident pseudo labels and trust less on the noisy pseudo labels of low confidence. This is achieved by multiplying the estimated confidence map c^{t} to the pseudo labels in the Dice loss for mutual supervision. Such confidence-weighted Dice loss is given by

$$L_{\mathrm{diceC}}(c^{\mathrm{t}}, p^{\mathrm{t}}, \hat{y}^{\mathrm{t}}) = \frac{1}{K} \sum_{k=0}^{K} \left(1 - 2 \times \frac{\sum_{i=1}^{H \times W} c^{\mathrm{t}}(i) \times y_k^{\mathrm{t}}(i) \times p_k^{\mathrm{t}}(i) + \gamma}{\sum_{i=1}^{H \times W} (c^{\mathrm{t}}(i) \times y_k^{\mathrm{t}}(i) + c^{\mathrm{t}}(i) \times p_k^{\mathrm{t}}(i)) + \gamma}\right). \tag{6}$$

More specifically, in the late stage, for the target images, the dual network f_1 and f_2 are optimized with $L_1^{\mathrm{T}} = \frac{1}{N_{\mathrm{T}}} \sum_{x^{\mathrm{t}} \sim X^{\mathrm{T}}} L_{\mathrm{diceC}}(c_2^{\mathrm{t}}, p_1^{\mathrm{t}}, y_2^{\mathrm{t}})$ and $L_2^{\mathrm{T}} = \frac{1}{N_{\mathrm{T}}} \sum_{x^{\mathrm{t}} \sim X^{\mathrm{T}}} L_{\mathrm{diceC}}(c_1^{\mathrm{t}}, p_2^{\mathrm{t}}, y_1^{\mathrm{t}})$, respectively. Since the pseudo label for the target images is generally less confidence than the ground-truth annotation for the source images, we also weigh the confidence-weighted Dice loss on the target domain with a hyper-parameter λ_{t}.

4 Experiments

4.1 Dataset and evaluation protocol

(a) Datasets. We conduct experiments on three types of unsupervised cross-modality image segmentation tasks, including whole heart segmentation on the MM-WHS 2017 dataset [50], brain tumor segmentation on the Brats 2018 dataset [51], and abdominal Multi-organ segmentation dataset consisting of the CHAOS challenge dataset and the BTCV dataset [52, 53]. The detail of these involved datasets is given in the following.

MM-WHS 2017. The Multi-Modality Whole Heart Segmentation (MM-WHS) Challenge 2017 dataset [50] consists of unpaired 20 MRI and 20 CT volumes from different clinical sites. The goal is to segment four cardiac structures: the ascending aorta (AA), the left atrium blood cavity (LAC), the left ventricle blood cavity (LVC), and the myocardium of the left ventricle (MYO). The original size of each slice ranges from 256×256 to 512×512. We conduct experiments under both "MRI to CT" (MRI2CT) and "CT to MRI" (CT2MRI) cross-modality settings.

Brats 2018. The Multi-Modality Brain Tumor Segmentation Challenge 2018 dataset [51] contains MRI images with four different modalities (FLAIR, T1, T1CE, and T2) from 75 low-graded glioma (LGG) cases. We conduct cross-modality brain tumor segmentation experiments by regarding the T2 modality as the source domain, and the rest modalities as the target domains. The original size of the slices in the Brats 2018 dataset [51] is 240×240. There are three types of tumors in the Brats 2018 dataset, i.e., the enhancing tumor, the peritumoral edema, and the necrotic and non-enhancing tumor core. We follow [12] to combine the three types of tumors as a single tumor class for a fair comparison with existing studies [25, 37].

Multi-organ segmentation. The last type of dataset is from abdominal multi-organ segmentation task [52,53]. Similarly to the whole heart segmentation task, we also conduct experiments under MRI2CT and CT2MRI settings. The MRI modality images are from the Combined Healthy Abdominal Organ Segmentation (CHAOS) challenge dataset [52], containing 20 T2w MRI volumes. The CT modality images are from the Multi-Atlas Labeling Beyond the Cranial Vault Challenge (BTCV) [53]. This dataset consists of 30 contrast-enhanced portal venous phase CT scans. We aim to segment the liver, left and right kidneys (LKid and RKid), and spleen. The original size of each slice in BTCV [53] is 512×512.

For all datasets, all the source modality images are used as the training set. Following SIFA [39] and DSFN [12], for the target modality images, we randomly select 80% cases as the training set and 20% as the test set. Each slice is re-sampled into the size of 256×256 for a fair comparison with the other methods. The image intensity is first normalized by subtracting the mean intensity and dividing by the standard deviation, and then normalized into range $[-1, 1]$. We also perform some classical dataset augmentations such as random crop and rotation during training.

(b) Evaluation protocol. We adopt the Dice score (%) to benchmark all the unsupervised cross-modality image segmentation tasks. The Dice score evaluates the similarity between the predicted segmentation and the ground-truth 3D mask annotation. Besides, we also adopt the distance-based metrics. Specifically, we adopt the average symmetric surface distance (ASSD) in terms of voxel for the MM-WHS 2017 [50] and Multi-organ segmentation datasets [52, 53], and Hausdorff distance (HD) for the Brats 2018 [51] dataset. ASSD is obtained by computing the average distance between the surface of the prediction and ground-truth 3D segmentation and vice versa. HD is the maximum distance between two sets of voxels from the predicted mask and ground-truth segmentation.

4.2 Implementation details

We implement the proposed method with the PyTorch framework on a workstation with two Nvidia Titan X (Pascal) 12 GB memory GPU. For the CycleGAN [11] module described in Subsection 3.2, we adopt 9 layer ResNet [54] architecture as the generator and basic 70×70 PatchGAN [55] as the discriminator. We use the Adam optimizer [56] to train the CycleGAN. The learning rate is set to 1×10^{-4} for both the generator and the discriminator. For the segmentation network detailed in Subsection 3.4, we train the dual networks for 100 epochs on the MM-WHS 2017 [50] and Brats 2018 dataset [51], 200 epochs on the Multi-organ segmentation dataset [52, 53], using also the Adam optimizer [56]. We set the batchsize to 16 for all datasets. The learning rate is set to 1×10^{-3}. The hyperparameter λ_{t} involved in the loss on target modality images is set to 0.5, while γ in the Dice loss is set to 10^{-4}. The starting epoch n_{e} for using the confidence-weighted Dice loss is set to 50, 50, and 120 for the MM-WHS 2017, Brats 2018, and Multi-organ segmentation dataset, respectively.

4.3 Experimental results

We conduct experiments on the MM-WHS 2017 dataset [50], Brats 2018 dataset [51], and Multi-organ segmentation dataset [52, 53]. We mainly compare the proposed method with some state-of-the-art approaches and the baseline model of mutual supervision on the dual networks denoted by MS-Dual. Since the dual networks preform similarly, for a fair comparison, we report the performance given by the averaged predicted scores of the dual networks for both the baseline model MS-Dual and the proposed method denoted in the following by CWMS-Dual, which leverages the confidence-weighted Dice loss for mutual supervision on dual networks.

Experimental results on the MM-WHS 2017 dataset. We first evaluate the proposed method on the cross-modality whole heart segmentation on the MM-WHS 2017 dataset. Some qualitative illustrations are given in Figure 3. Since both networks f_1 and f_2 perform similarly, we simply illustrate the segmentation result given by the first network f_1. As shown in Figure 3, the proposed method achieves accurate segmentation, and outperforms the baseline model MS-Dual.

The quantitative comparison of the proposed method with some state-of-the-art methods for the MRI2CT setting is depicted in Table 1. The proposed method improves the baseline model MS-Dual by 1.08% Dice score, demonstrating the usefulness of the proposed confidence-weighted Dice loss and the confidence-guided fusion. Compared with SIFA [39] which simultaneously learns common feature across domains and translate images from the source to the target using the shared encoder, the proposed method achieves 12.68% improvement in terms of Dice score. Compared with DSFN [12] that bridges the gap between two domains by translating images in both source-to-target and target-to-source directions,

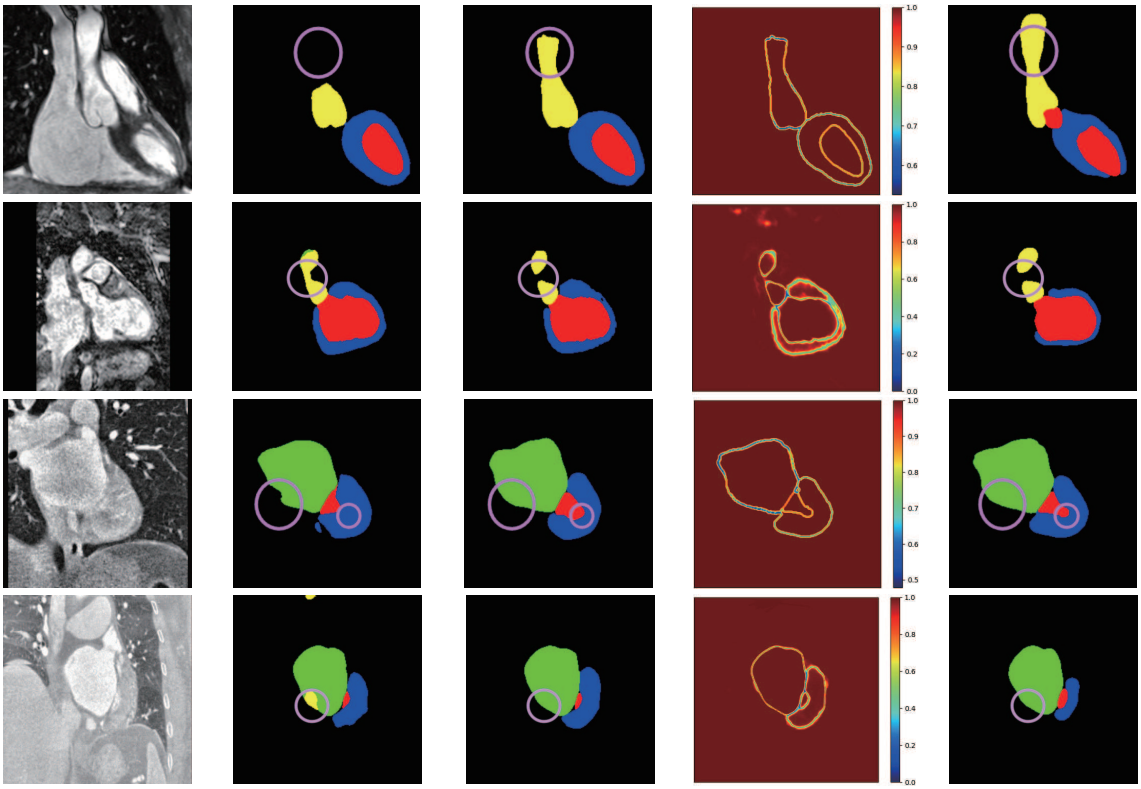

Figure 3 Some segmentation results on the MM-WHS 2017 dataset. Top two rows: CT2MRI segmentation; bottom two rows: MRI2CT segmentation. Different colors in the segmentation denote different regions. Yellow: ascending aorta (AA); green: left atrium blood cavity (LAC); red: left ventricle blood (LVC); blue: myocardium of the left ventricle (MYO). (a) Image; (b) baseline; (c) ours; (d) confidence map; (e) ground truth.

Table 1 Quantitative benchmark of MRI2CT segmentation results on MM-WHS 2017 [50] [a)]

Method	Dice score (%) ↑					ASSD (voxel) ↓				
	MYO	LAC	LVC	AA	Average	MYO	LAC	LVC	AA	Average
w/o adaptation	17.71	73.60	7.68	66.24	41.31	23.05	11.58	60.68	11.99	26.82
CycleGAN [11]	28.11	68.88	28.89	79.82	51.42	13.83	6.00	14.25	10.05	11.03
SIFA [39]	61.60	79.50	73.80	81.30	74.10	8.50	6.20	5.50	7.90	7.00
IB-GAN [13]	61.50	79.67	76.27	80.24	74.42	6.09	6.34	7.15	11.88	7.86
ARL-GAN [14]	**81.60**	80.60	69.00	71.30	75.70	6.50	5.90	6.70	6.30	6.40
DSFN [12]	62.40	76.90	79.10	84.70	75.80	15.70	11.90	10.60	7.40	11.40
DADASeg-Net [34]	61.20	80.70	77.90	87.00	76.70	5.50	5.60	4.70	4.50	5.10
DSAN [37]	66.52	84.76	82.77	79.92	78.50	5.59	6.65	3.77	7.68	5.92
UMDA-SNA-SFCNN [15]	66.20	82.70	82.60	89.20	80.20	4.50	3.60	3.00	6.70	4.40
ICMSC [41]	72.40	86.40	84.30	85.60	82.20	**3.20**	3.30	3.40	**2.40**	3.10
DaLST [17]	67.59	90.09	86.13	89.92	83.44	–	–	–	–	–
MPSCL [43]	72.51	87.08	**86.45**	90.26	84.08	3.41	3.16	**2.85**	3.47	3.47
MS-Dual (baseline)	75.40	89.08	86.25	92.10	85.70	3.76	3.56	3.00	3.14	3.37
CWMS-Dual (ours)	77.48	**90.20**	86.41	**93.02**	**86.78**	3.62	**2.79**	2.88	2.54	**2.96**

a) The best results are in bold.

the proposed method improves it by 10.98% Dice score. The proposed method also outperforms the most recent methods DaLST [17] and MPSCL [43] by 3.34% and 2.70%, respectively. Besides, the proposed method also achieves the best ASSD.

Table 2 lists the quantitative results of CT2MRI segmentation. The proposed method outperforms the baseline model by 2.14% Dice score. Compared with the state-of-the-art approaches, the proposed method improves SIFA [39] by a large margin. Compared with DaLST [17], we make an improvement of 1.31% in terms of the Dice score. The proposed method also scores the best according to ASSD.

Table 2 Quantitative evaluation of CT2MRI segmentation results on MM-WHS 2017 [50] [a)]

Method	Dice score (%) ↑					ASSD (voxel) ↓				
	MYO	LAC	LVC	AA	Average	MYO	LAC	LVC	AA	Average
w/o adaptation	0.64	17.80	14.61	7.16	10.05	30.53	26.88	22.32	24.50	26.06
CycleGAN [11]	39.33	12.73	70.91	31.29	38.56	7.09	17.87	7.76	12.13	11.21
SIFA [39]	47.30	62.30	78.90	65.30	63.40	4.40	7.40	3.80	7.30	5.70
DSAN [37]	52.07	66.23	76.30	71.29	66.45	4.25	7.30	5.46	**4.44**	5.36
MPSCL [43]	55.90	77.34	81.61	64.66	69.87	3.50	2.64	3.44	5.59	3.80
DaLST [17]	**73.85**	78.58	**92.97**	69.36	78.69	–	–	–	–	–
MS-Dual (baseline)	62.93	83.21	89.34	75.95	77.86	3.89	2.30	3.33	6.37	3.97
CWMS-Dual (ours)	67.19	**83.82**	91.92	**77.08**	**80.00**	**3.03**	**2.45**	**1.86**	6.10	**3.36**

a) The best results are in bold.

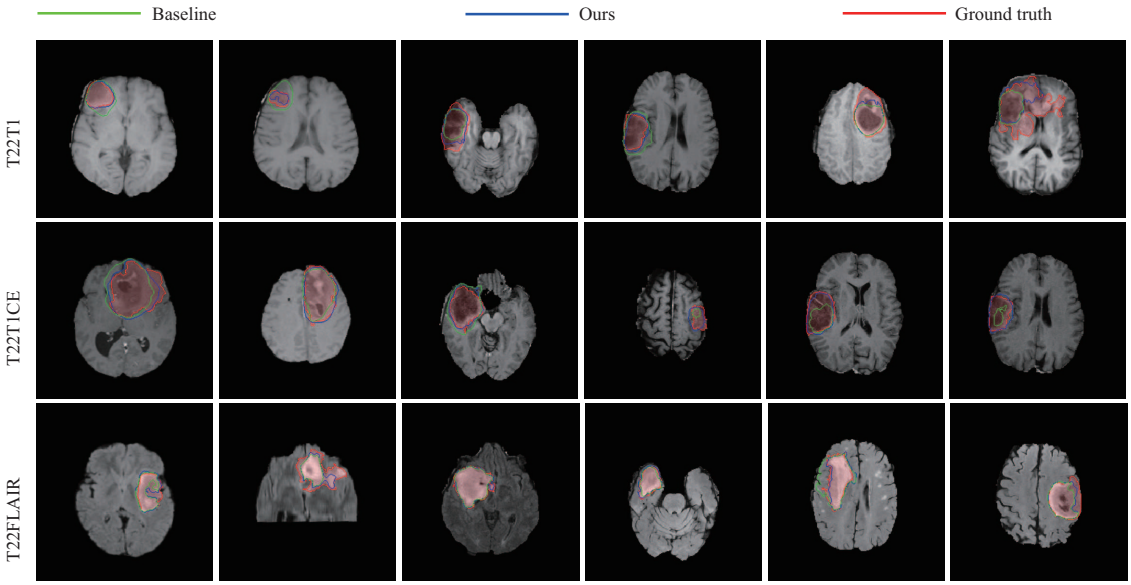

Figure 4 Some qualitative segmentation results on the Brats 2018 dataset.

Experimental results on the Brats 2018 dataset. We then conduct experiments on the Brats 2018 dataset. As shown in Figure 4, the proposed method achieves accurate brain tumor cross-modality segmentation. The quantitative evaluation is depicted in Table 3. The proposed method improves the baseline model by 1.55% Dice score. Compared with SIFA [39], the proposed method increases the Dice score from 59.30% to 75.37%. We outperform DSFN [12] by 9.27% Dice score. Compared with the more recent methods, DSAN [37] and Self Semantic Contour [40], the proposed method achieves 8.19% and 6.80% Dice score improvement, respectively. Based on the Hausdorff distance measure, the proposed method also performs the best on the unsupervised cross-modality brain tumor segmentation task.

Experimental results on the Multi-organ segmentation dataset. Figure 5 illustrates some qualitative cross-modality multi-organ segmentation results. As shown in Figure 5, the proposed method achieves a rather accurate cross-modality segmentation result. The quantitative comparison with some state-of-the-art methods on CT2MRI and MRI2CT settings is depicted in Tables 4 and 5, respectively. The proposed method performs the best for CT2MRI segmentation in terms of both Dice score and ASSD. The proposed method improves the baseline model by 0.13%. Compared with DSAN [37] and DaLST [17], we achieve 0.92% and 1.11% Dice score improvement, respectively. The proposed method outperforms PSIGAN [57] by 0.81%. For the MRI2CT segmentation, the proposed method performs the best in terms of ASSD. Based on the Dice score, we outperform the baseline model by 2.04%. It is noteworthy that though the proposed method performs slightly worse than DaLST [17] and AttENT [38] in terms of Dice score, DaLST [17] and AttENT [38] use PSPNet [3] and Deeplab-V2 [4] with ResNet101 [54] as the backbone. Besides, they are both pre-trained on the ImageNet [58].

Table 3 Quantitative benchmark of different methods on Brats 2018 [51] a)

Method	Dice score (%) ↑				Hausdorff distance (mm) ↓			
	T1	T1CE	FLAIR	Average	T1	T1CE	FLAIR	Average
w/o adaptation	8.99	5.55	64.60	26.38	76.31	80.01	49.17	68.50
CycleGAN [11]	38.10	42.10	63.30	47.80	25.40	23.20	17.20	21.90
SIFA [39]	51.70	58.20	68.00	59.30	19.60	15.10	16.90	17.10
DSFN [12]	57.30	62.20	78.90	66.10	17.50	15.50	13.80	15.60
DaLST [17]	–	–	81.26	–	–	–	–	–
DSAN [37]	57.70	62.04	81.79	67.18	14.24	13.70	8.62	12.19
Self Semantic Contour [40]	59.30	63.50	82.90	68.57	12.50	11.20	7.90	10.53
MS-Dual (baseline)	71.97	66.63	82.85	73.82	10.86	11.78	4.64	9.10
CWMS-Dual (ours)	**73.29**	**68.95**	**83.87**	**75.37**	**9.87**	**10.69**	**4.33**	**8.30**

a) The best results are in bold.

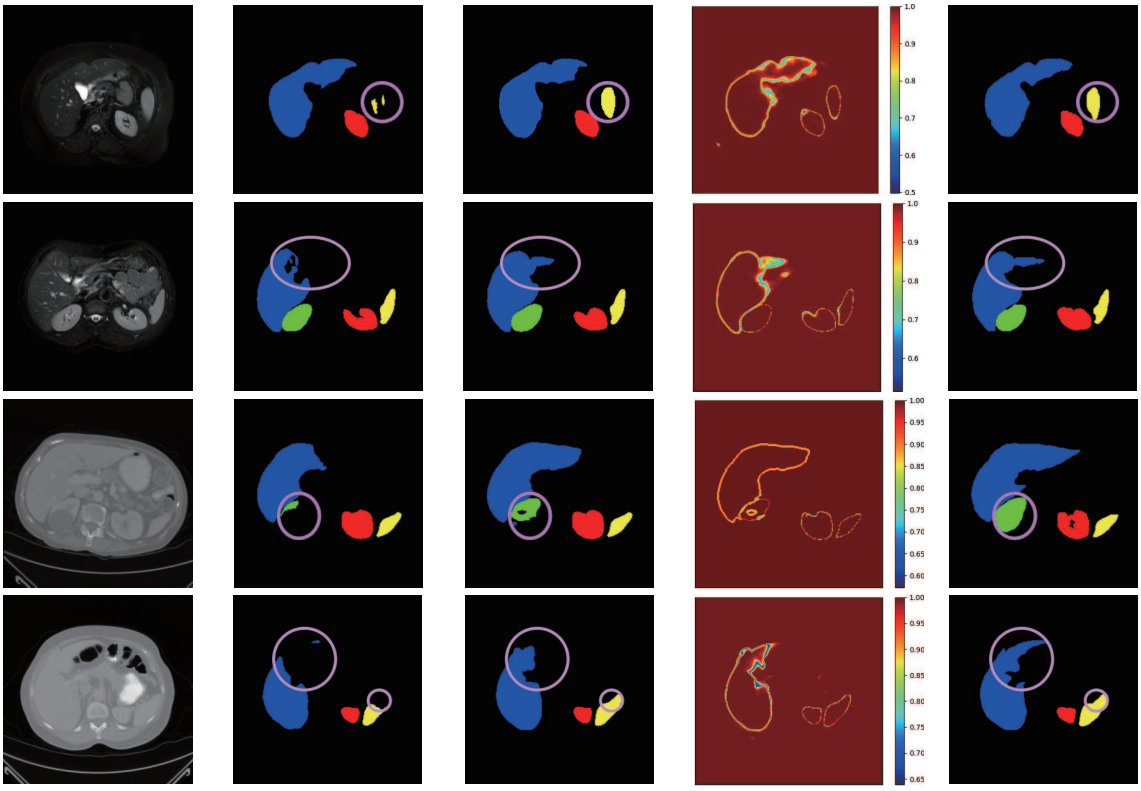

Figure 5 CT2MRI (top two rows) and MRI2CT (bottom two rows) segmentation results on the Multi-organ segmentation dataset. Different colors denote different regions. Yellow: spleen; green: left kidney; red: right kidney; blue: liver. (a) Image; (b) baseline; (c) ours; (d) confidence map; (e) ground truth.

5 Discussion

The core idea of the proposed method revolves around confidence-weighted mutual supervision. We first analyze the confidence map defined in (5). As illustrated in Figure 1(a), the confidence map effectively evaluates the quality of pseudo labels. Pseudo labels with higher confidence are more likely to be correct. Therefore, integrating such confidence measures in mutual supervision enables better utilization of pseudo labels from the dual networks, thereby enhancing the segmentation results. As illustrated in Figures 3 and 5, pixels with low confidence mainly locate around region boundaries, which is reasonable. By mitigating the impact of false supervision in these areas, the segmentation accuracy is significantly improved.

We then conduct three types of ablation studies to discuss the effect of the proposed method. Firstly, we evaluate the effect of different modules involved in the proposed pipeline: CycleGAN, mutual supervision on dual networks (MS), and confidence-weighted Dice loss (CWDL). For this study, in addition to the result of averaged predicted scores of dual networks, we also show the result for each of the dual networks

Table 4 Evaluation of different CT2MRI segmentation results on the Multi-organ segmentation dataset [52, 53]

Method	Dice score (%) ↑					ASSD (voxel) ↓				
	Liver	RKid	LKid	Spleen	Average	Liver	RKid	LKid	Spleen	Average
w/o adaptation	14.12	3.87	46.47	20.09	21.14	8.22	8.92	5.61	7.63	7.59
CycleGAN [11]	86.00	75.00	88.00	87.00	84.00	2.00	3.20	1.90	2.60	2.40
SIFA [39]	90.00	89.10	80.20	82.30	85.40	1.50	0.60	1.50	2.40	1.50
AttENT [38]	**91.05**	81.38	80.51	89.75	85.67	0.99	1.03	1.26	1.12	1.10
DaLST [17]	90.67	87.36	87.52	**93.03**	89.65	–	–	–	–	–
DSAN [37]	89.30	90.16	90.09	89.83	89.84	2.18	1.25	1.10	1.10	1.41
PSIGAN [57]	91.00	87.00	91.00	90.00	90.00	–	–	–	–	–
MS-Dual (baseline)	87.50	**93.21**	**92.05**	89.74	90.63	0.68	**0.12**	**0.15**	**0.35**	0.32
CWMS-Dual (ours)	88.02	92.54	91.81	90.59	**90.76**	**0.51**	0.13	0.24	0.36	**0.31**

Table 5 Evaluation of different MRI2CT segmentation results on the Multi-organ segmentation dataset [52, 53]

Method	Dice score (%) ↑					ASSD (voxel) ↓				
	Liver	RKid	LKid	Spleen	Average	Liver	RKid	LKid	Spleen	Average
w/o adaptation	40.03	15.14	35.16	36.24	31.64	8.94	Infinite	11.17	11.83	Infinite
CycleGAN [11]	84.50	78.60	80.30	76.90	80.10	1.80	1.30	1.20	1.90	1.60
SIFA [39]	88.00	83.30	80.90	82.60	83.70	1.20	1.00	1.50	1.60	1.30
DSAN [37]	87.50	83.40	82.90	83.63	84.36	–	–	–	–	–
AttENT [38]	88.56	80.66	85.59	86.34	85.29	0.68	1.31	1.43	1.21	1.16
DaLST [17]	90.52	**85.08**	**87.36**	**86.64**	**87.40**	–	–	–	–	–
MS-Dual (baseline)	89.29	82.55	83.22	75.38	82.52	0.32	0.55	0.46	1.26	0.65
CWMS-Dual (ours)	**90.59**	83.41	85.50	79.55	84.94	**0.26**	**0.53**	**0.35**	**0.74**	**0.47**

Table 6 Ablation study on each component for MRI2CT segmentation on MM-WHS 2017 dataset[a)]

Method	Dice score (%) ↑					ASSD (voxel) ↓				
	MYO	LAC	LVC	AA	Average	MYO	LAC	LVC	AA	Average
CycleGAN only	28.11	68.88	28.89	79.82	51.42	13.83	6.00	14.25	10.05	11.03
MS-Dual (baseline)	75.40	89.08	86.25	92.10	85.70	3.76	3.56	3.00	3.14	3.37
CycleGAN+MS+CWDL (U-Net1)	76.47	89.51	85.30	92.49	85.94	3.81	3.22	3.27	4.43	3.68
CycleGAN+MS+CWDL (U-Net2)	77.34	90.08	86.26	92.73	86.61	3.83	3.05	**2.85**	2.61	3.08
CycleGAN+MS+CWDL (averaged score)	**77.48**	**90.20**	**86.41**	**93.02**	**86.78**	**3.62**	**2.79**	2.88	**2.54**	**2.96**

a) MS and CWDL denote mutual supervision and confidence-weighted Dice loss, respectively. U-Net1, U-Net2, and averaged score are the results of the prediction from each of the dual networks and the averaged score of dual networks, respectively. The best results are in bold.

Table 7 Ablation study on each component for CT2MRI segmentation on MM-WHS 2017 dataset[a)]

Method	Dice score (%) ↑					ASSD (voxel) ↓				
	MYO	LAC	LVC	AA	Average	MYO	LAC	LVC	AA	Average
CycleGAN only	39.33	12.73	70.91	31.29	38.56	7.09	17.87	7.76	12.13	11.21
MS-Dual (baseline)	62.96	83.21	89.34	75.95	77.86	3.89	2.30	3.33	6.37	3.97
CycleGAN+MS+CWDL (U-Net1)	66.34	83.16	91.42	76.98	79.47	3.54	4.76	2.12	6.07	4.12
CycleGAN+MS+CWDL (U-Net2)	**67.48**	83.22	**92.11**	76.99	79.95	3.20	**2.37**	**1.77**	**5.87**	**3.30**
CycleGAN+MS+CWDL (averaged score)	67.19	**83.82**	91.92	**77.08**	**80.00**	**3.03**	2.45	1.86	6.10	3.36

a) MS and CWDL denote mutual supervision and confidence-weighted Dice loss, respectively. U-Net1, U-Net2, and averaged score are the results of the prediction from each of the dual networks and the averaged score of dual networks, respectively. The best results are in bold.

denoted by U-Net1 and U-Net2. Then we conduct an ablation study on when to apply the confidence-weighted Dice loss. Finally, we compare our confidence weighting strategy with the hard thresholding strategy to validate the advantage of our confidence-weighted Dice loss. We conduct all ablation studies on the MM-WHS 2017 dataset [50], which is widely used in unsupervised cross-modality image segmentation.

Ablation study on different modules. As depicted in Tables 6 and 7, using only CycleGAN achieves relatively reasonable results, but is far from satisfied. The mutual supervision on dual networks is quite effective. The proposed confidence-weighted Dice loss further improves the result of CycleGAN by 35.36% and 41.44% Dice score, under MRI2CT and CT2MRI settings. Compared with the baseline model

Table 8 Ablation study on the starting epoch n_e of applying the confidence-weighted Dice loss to the MRI2CT and CT2MRI segmentation on the MM-WHS 2017 dataset[a)]

Starting epoch n_e	Dice score (%) ↑					ASSD (voxel) ↓				
	MYO	LAC	LVC	AA	Average	MYO	LAC	LVC	AA	Average
MRI2CT										
40th Epoch	75.50	89.21	85.21	92.32	85.56	3.66	3.62	3.10	6.09	4.12
45th Epoch	76.76	89.25	85.80	92.05	85.96	3.51	3.53	3.10	4.67	3.70
50th Epoch (ours)	**77.48**	**90.20**	86.41	**93.02**	**86.78**	3.62	**2.79**	2.88	**2.54**	**2.96**
55th Epoch	74.26	89.20	**87.01**	92.72	85.80	3.81	3.83	**2.78**	5.33	3.94
60th Epoch	76.25	89.62	86.02	92.77	86.17	**3.44**	3.42	3.17	3.07	3.30
CT2MRI										
40th Epoch	67.68	**84.43**	**93.67**	76.17	**80.49**	**2.98**	**2.22**	**1.23**	6.26	**3.17**
45th Epoch	66.38	84.16	91.67	76.05	79.56	3.25	2.80	1.80	6.61	3.62
50th Epoch (ours)	67.19	83.82	91.92	77.08	80.00	3.03	2.45	1.86	6.10	3.36
55th Epoch	**67.98**	82.94	91.79	**77.66**	80.09	3.35	2.53	1.68	**5.90**	3.37
60th Epoch	67.86	83.31	92.72	76.79	80.17	2.98	2.26	1.71	6.44	3.35

a) The best results are in bold.

Table 9 Ablation study on hard thresholding and confidence weighting in MRI2CT and CT2MRI tasks[a)]

Method	Dice score (%) ↑					ASSD (voxel) ↓				
	MYO	LAC	LVC	AA	Average	MYO	LAC	LVC	AA	Average
MRI2CT										
thresh = 0.80	75.06	89.46	86.42	92.85	85.95	3.64	3.74	2.86	3.68	3.48
thresh = 0.85	74.26	89.22	85.82	92.20	85.37	3.50	3.32	3.09	3.97	3.47
thresh = 0.90	75.26	89.84	87.18	91.65	85.98	**3.50**	3.09	3.02	2.59	3.05
thresh = 0.95	76.64	89.44	86.44	92.51	86.25	3.63	3.49	2.90	4.08	3.53
thresh = 0.96	75.32	89.05	**87.33**	91.34	85.76	3.57	3.61	**2.80**	2.95	3.23
CWMS-Dual (ours)	**77.48**	**90.20**	86.41	**93.02**	**86.78**	3.62	**2.79**	2.88	**2.54**	**2.96**
CT2MRI										
thresh = 0.80	64.98	83.49	90.18	75.68	78.58	3.19	2.58	2.59	6.10	3.61
thresh = 0.85	**68.17**	82.29	**91.98**	75.50	79.48	**3.00**	2.78	1.88	6.90	3.64
thresh = 0.90	67.15	82.08	91.89	75.44	79.13	3.23	2.55	1.95	6.35	3.53
thresh = 0.95	67.64	82.46	91.12	74.53	78.93	3.06	3.62	1.89	7.04	3.90
thresh = 0.96	66.42	82.55	90.55	75.12	78.66	3.28	2.80	2.68	6.77	3.87
CWMS-Dual (ours)	67.19	**83.82**	91.92	**77.08**	**80.00**	3.03	**2.45**	**1.86**	**6.10**	**3.36**

a) The best results are in bold.

MS-Dual, the proposed method achieves an improvement 1.08% and 2.14% Dice score under MRI2CT and CT2MRI setting, respectively. Besides, each of the dual networks also outperforms the result of averaged scores from the dual networks in the baseline model.

Ablation study on the starting epoch n_e. As shown in Table 8, the starting epoch of applying the confidence-weighted Dice loss somehow influences the segmentation accuracy. Yet, the performance is rather stable for a wide range of starting epochs from 40 to 60 under both MRI2CT and CT2MRI settings. We choose $n_e = 50$ as the optimal starting epoch for the cross-modality whole heart segmentation and brain tumor segmentation on Brats 2018 dataset. For the multi-organ segmentation task, we set the starting epoch n_e to 120.

Ablation study on hard thresholding vs. confidence-weighting. As shown in Table 9, hard thresholding on the pseudo labels may also achieve interesting results. Yet, it is somehow difficult to set a fixed threshold value (e.g., 0.95 for MRI2CT task and 0.85 for CT2MRI task) to get good results. Besides, compared with all hard thresholding settings, our soft confidence weighting strategy achieves superior performance thanks to the full utilization of the information from the pseudo labels.

6 Conclusion

In this paper, we aim to tackle the problem of unsupervised cross-modality image segmentation. For that, we propose a pipeline of confidence-weighted mutual supervision on dual networks. Specifically, we first

apply image translation based on CycleGAN to reduce the domain shift between different modalities. We then adopt self-learning and rely on mutual supervision to get rid of some noisy pseudo labels caused by the confirmation bias of a single network. To further alleviate the negative impact of noisy pseudo labels, we propose a confidence-weighted Dice loss to take into account the quality of pseudo labels for the dual networks. Extensive experimental results on three widely used datasets demonstrate that the proposed method consistently improves the baseline model, and outperforms some state-of-the-art methods.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant Nos. 62061160490, 62122029, U20B2064).

References

1 Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015. 3431–3440
2 Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention, 2015. 234–241
3 Zhao H, Shi J, Qi X, et al. Pyramid scene parsing network. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017. 2881–2890
4 Chen L C, Papandreou G, Kokkinos I, et al. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Trans Pattern Anal Mach Intell, 2017, 40: 834–848
5 Wang J H, Liu B, Xu K. Semantic segmentation of high-resolution images. Sci China Inf Sci, 2017, 60: 123101
6 Geng Q C, Zhou Z, Cao X C. Survey of recent progress in semantic image segmentation with CNNs. Sci China Inf Sci, 2018, 61: 051101
7 Ma S, Pang Y W, Pan J, et al. Preserving details in semantics-aware context for scene parsing. Sci China Inf Sci, 2020, 63: 120106
8 Zhang Z J, Pang Y W. CGNet: cross-guidance network for semantic segmentation. Sci China Inf Sci, 2020, 63: 120104
9 Isensee F, Jaeger P F, Kohl S A A, et al. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat Methods, 2021, 18: 203–211
10 Xu Q, Xi X M, Meng X J, et al. Difficulty-aware bi-network with spatial attention constrained graph for axillary lymph node segmentation. Sci China Inf Sci, 2022, 65: 192102
11 Zhu J Y, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of IEEE International Conference on Computer Vision (ICCV), 2017. 2223–2232
12 Zou D, Zhu Q, Yan P. Unsupervised domain adaptation with dual-scheme fusion network for medical image segmentation. In: Proceedings of the 29th International Joint Conference on Artificial Intelligence, 2020. 3291–3298
13 Chen J, Zhang Z, Xie X, et al. Beyond mutual information: generative adversarial network for domain adaptation using information bottleneck constraint. IEEE Trans Med Imag, 2021, 41: 595–607
14 Chen X, Lian C, Wang L, et al. Anatomy-regularized representation learning for cross-modality medical image segmentation. IEEE Trans Med Imag, 2020, 40: 274–285
15 Liu J, Liu H, Gong S, et al. Automated cardiac segmentation of cross-modal medical images using unsupervised multi-domain adaptation and spatial neural attention structure. Med Image Anal, 2021, 72: 102135
16 Jafari M, Francis S, Garibaldi J M, et al. LMISA: a lightweight multi-modality image segmentation network via domain adaptation using gradient magnitude and shape constraint. Med Image Anal, 2022, 81: 102536
17 Xie Q, Li Y, He N, et al. Unsupervised domain adaptation for medical image segmentation by disentanglement learning and self-training. IEEE Trans Med Imag, 2022. doi: 10.1109/TMI.2022.3192303
18 Zhou W, Wang Y, Chu J, et al. Affinity space adaptation for semantic segmentation across domains. IEEE Trans Image Process, 2020, 30: 2549–2561
19 Toldo M, Maracani A, Michieli U, et al. Unsupervised domain adaptation in semantic segmentation: a review. Technologies, 2020, 8: 35
20 French G, Mackiewicz M, Fisher M. Self-ensembling for visual domain adaptation. In: Proceedings of International Conference On Learning Representations, 2018
21 Li J, Zhou K, Qian S H, et al. Feature re-representation and reliable pseudo label retraining for cross-domain semantic segmentation. IEEE Trans Pattern Anal Mach Intell, 2022. doi: 10.1109/TPAMI.2022.3154933
22 Tranheden W, Olsson V, Pinto J, et al. DACS: domain adaptation via cross-domain mixed sampling. In: Proceedings of IEEE Winter Conference on Applications of Computer Vision (WACV), 2021. 1379–1389
23 Chen M, Xue H, Cai D. Domain adaptation for semantic segmentation with maximum squares loss. In: Proceedings of International Conference on Computer Vision, 2019. 2090–2099
24 Spadotto T, Toldo M, Michieli U, et al. Unsupervised domain adaptation with multiple domain discriminators and adaptive self-training. In: Proceedings of International Conference on Pattern Recognition, 2021. 2845–2852

25 Zou Y, Yu Z, Kumar B, et al. Unsupervised domain adaptation for semantic segmentation via class-balanced self-training. In: Proceedings of European Conference on Computer Vision, 2018. 289–305

26 Han B, Yao Q, Yu X, et al. Co-teaching: robust training of deep neural networks with extremely noisy labels. In: Proceedings of the 32nd International Conference on Neural Information Processing Systems, 2018. 8536–8546

27 Ouali Y, Hudelot C, Tami M. Semi-supervised semantic segmentation with cross-consistency training. In: Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020. 12674–12684

28 Sohn K, Berthelot D, Carlini N, et al. FixMatch: simplifying semi-supervised learning with consistency and confidence. In: Proceedings of Advances in Neural Information Processing Systems, 2020. 33: 596–608

29 Liu S, Niles-Weed J, Razavian N, et al. Early-learning regularization prevents memorization of noisy labels. In: Proceedings of Advances in Neural Information Processing Systems, 2020. 33: 20331–20342

30 Liu S, Liu K, Zhu W, et al. Adaptive early-learning correction for segmentation from noisy annotations. In: Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022. 2606–2616

31 Zhang Y, Xiang T, Hospedales T M, et al. Deep mutual learning. In: Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2018. 4320–4328

32 Chen X, Yuan Y, Zeng G, et al. Semi-supervised semantic segmentation with cross pseudo supervision. In: Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021. 2613–2622

33 Yang Y, Soatto S. FDA: Fourier domain adaptation for semantic segmentation. In: Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020. 4085–4095

34 Chen X, Kuang T, Deng H, et al. Dual adversarial attention mechanism for unsupervised domain adaptive medical image segmentation. IEEE Trans Med Imag, 2022, 41: 3445–3453

35 Vu T H, Jain H, Bucher M, et al. Advent: adversarial entropy minimization for domain adaptation in semantic segmentation. In: Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019. 2517–2526

36 Truong T D, Duong C N, Le N, et al. BiMaL: Bijective maximum likelihood approach to domain adaptation in semantic scene segmentation. In: Proceedings of IEEE/CVF International Conference on Computer Vision (ICCV), 2021. 8548–8557

37 Han X, Qi L, Yu Q, et al. Deep symmetric adaptation network for cross-modality medical image segmentation. IEEE Trans Med Imag, 2021, 41: 121–132

38 Li C, Luo X, Chen W, et al. AttENT: domain-adaptive medical image segmentation via attention-aware translation and adversarial entropy minimization. In: Proceedings of IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2021. 952–959

39 Chen C, Dou Q, Chen H, et al. Unsupervised bidirectional cross-modality adaptation via deeply synergistic image and feature alignment for medical image segmentation. IEEE Trans Med Imag, 2020, 39: 2494–2505

40 Liu X, Xing F, Fakhri G E, et al. Self-semantic contour adaptation for cross modality brain tumor segmentation. In: Proceedings of IEEE 19th International Symposium on Biomedical Imaging (ISBI), 2022. 1–5

41 Zeng G, Lerch T D, Schmaranzer F, et al. Semantic consistent unsupervised domain adaptation for cross-modality medical image segmentation. In: Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention, 2021. 201–210

42 Choi J, Kim T, Kim C. Self-ensembling with GAN-based data augmentation for domain adaptation in semantic segmentation. In: Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019. 6830–6840

43 Liu Z, Zhu Z, Zheng S, et al. Margin preserving self-paced contrastive learning towards domain adaptation for medical image segmentation. IEEE J Biomed Health Inform, 2022, 26: 638–647

44 Fan J, Gao B, Jin H, et al. UCC: uncertainty guided cross-head co-training for semi-supervised semantic segmentation. In: Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022. 9947–9956

45 Wang Y, Wang H, Shen Y, et al. Semi-supervised semantic segmentation using unreliable pseudo-labels. In: Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022. 4248–4257

46 Gu X, Sun J, Xu Z. Unsupervised and semi-supervised robust spherical space domain adaptation. IEEE Trans Pattern Anal Mach Intell, 2022. doi: 10.1109/TPAMI.2022.3158637

47 He N J, Fang L Y, Plaza A. Hybrid first and second order attention Unet for building segmentation in remote sensing images. Sci China Inf Sci, 2020, 63: 140305

48 Zhou Z, Siddiquee M M R, Tajbakhsh N, et al. UNet++: redesigning skip connections to exploit multiscale features in image segmentation. IEEE Trans Med Imag, 2020, 39: 1856–1867

49 Gut D, Tabor Z, Szymkowski M, et al. Benchmarking of deep architectures for segmentation of medical images. IEEE Trans Med Imag, 2022, 41: 3231–3241

50 Zhuang X, Shen J. Multi-scale patch and multi-modality atlases for whole heart segmentation of MRI. Med Image Anal, 2016, 31: 77–87

51 Menze B H, Jakab A, Bauer S, et al. The multimodal brain tumor image segmentation benchmark (BRATS). IEEE Trans Med Imag, 2014, 34: 1993–2024

52 Kavur A E, Gezer N S, Barış M, et al. CHAOS Challenge-combined (CT-MR) healthy abdominal organ segmentation. Med Image Anal, 2021, 69: 101950

53 Landman B, Xu Z, Igelsias J, et al. MICCAI multi-atlas labeling beyond the cranial vault–workshop and challenge. In: Proceedings of MICCAI Workshop, 2015. 12

54 He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. In: Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2016. 770–778

55 Isola P, Zhu J Y, Zhou T, et al. Image-to-image translation with conditional adversarial networks. In: Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2017. 1125–1134

56 Kingma D P, Ba J. Adam: a method for stochastic optimization. In: Proceedings of International Conference on Learning Representations, 2015

57 Jiang J, Hu Y C, Tyagi N, et al. PSIGAN: joint probabilistic segmentation and image distribution matching for unpaired cross-modality adaptation-based MRI segmentation. IEEE Trans Med Imag, 2020, 39: 4071–4084

58 Deng J, Dong W, Socher R, et al. ImageNet: a large-scale hierarchical image database. In: Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2009. 248–255