Source-Free Domain Adaptation for Medical Image Segmentation via Prototype-Anchored Feature Alignment and Contrastive Learning

Qinji Yu[1], Nan Xi[2], Junsong Yuan[2], Ziyu Zhou[1], Kang Dang[3], and Xiaowei Ding[1,3(✉)]

[1] Shanghai Jiao Tong University, Shanghai, China
dingxiaowei@sjtu.edu.cn
[2] State University of New York at Buffalo, New York, USA
[3] VoxelCloud, Inc., Los Angeles, USA

Abstract. Unsupervised domain adaptation (UDA) has increasingly gained interests for its capacity to transfer the knowledge learned from a labeled source domain to an unlabeled target domain. However, typical UDA methods require concurrent access to both the source and target domain data, which largely limits its application in medical scenarios where source data is often unavailable due to privacy concern. To tackle the source data-absent problem, we present a novel two-stage source-free domain adaptation (SFDA) framework for medical image segmentation, where only a well-trained source segmentation model and unlabeled target data are available during domain adaptation. Specifically, in the prototype-anchored feature alignment stage, we first utilize the weights of the pre-trained pixel-wise classifier as source prototypes, which preserve the information of source features. Then, we introduce the bi-directional transport to align the target features with class prototypes by minimizing its expected cost. On top of that, a contrastive learning stage is further devised to utilize those pixels with unreliable predictions for a more compact target feature distribution. Extensive experiments on a cross-modality medical segmentation task demonstrate the superiority of our method in large domain discrepancy settings compared with the state-of-the-art SFDA approaches and even some UDA methods. Code is available at: https://github.com/CSCYQJ/MICCAI23-ProtoContra-SFDA.

Keywords: Domain adaptation · Source-free · Contrastive learning

---

K. Dang—Co-Corresponding Author.

---

1   Introduction

The inception of deep neural networks has revolutionized the landscape of medical image segmentation [14,24]. This tremendous success, however, is conditioned on the assumption that the training and testing data are drawn from the same distribution. Unfortunately, in real-world clinical scenarios, due to different acquisition protocols or various imaging modalities, domain shift is widespread between training (*i.e.*, source domain) and testing (*i.e.*, target domain) datasets [15]. This distribution gap usually degenerates the model performance on the target domain. To achieve reliable performance across different domains, a straightforward way is manually labeling some target data and fine-tuning the pretrained model on them [13]. However, obtaining expert-level annotation data in the medical imaging domain incurs significant time and expense [22]. Recently, unsupervised domain adaptation (UDA) has been widely investigated to reduce domain gap through transferring the knowledge learned from a rich-labeled source domain to an unlabeled target domain [4,7,17,19]. Existing UDA methods typically require sharing source data during adaptation, and enforce distribution alignment to diminish the domain discrepancy between source and target domains. This requirement limits the application of UDA methods when source domain data are not accessible. Hence, some very recent works have started to explore a more practical setting, source-free domain adaptation (SFDA), that adapts a pre-trained source model to unlabeled target domains without accessing any source data [1,5,6,12,20,21].
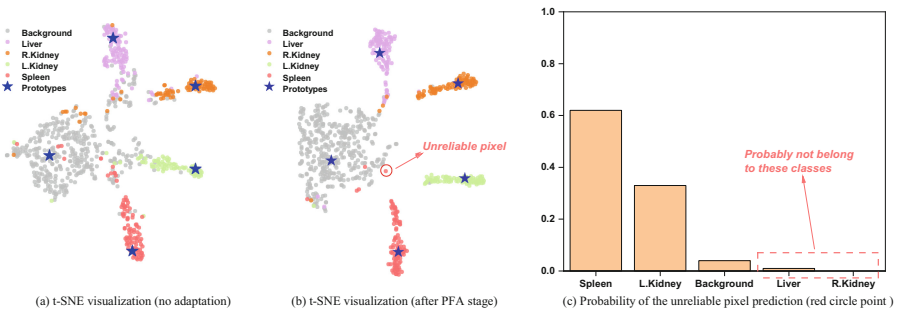

(a) t-SNE visualization (no adaptation)   (b) t-SNE visualization (after PFA stage)   (c) Probability of the unreliable pixel prediction (red circle point)

Fig. 1. (a→b) t-SNE visualization of target feature distributions in embedding space before and after Prototype-anchored Feature Alignment (PFA). (c) Category-wise probability of the unreliable pixel in (b).

Among these methods, [5] and [20] focus on generating reliable pseudo labels for target domain data by developing various denoising strategies. Unavoidably, these self-training methods depends heavily on initial probability maps produced by the source model, which are considerably unreliable when the domain discrepancy is large (*e.g.*, CT and MRI). To relieve the issues caused by noisy pseudo labels, Bateson et al. [1] proposed a prior-aware entropy minimization method to

minimize the label-free entropy loss for target predictions. Furthermore, unlike the above self-adaption methods, Yang et al. [21] utilized the statistic information stored in the batch normalization layer of the source model and mutual Fourier Transform to synthesize the source-like image. However, the quality of the generated image is still influenced by the domain discrepancy.

In this work, we propose a novel SFDA framework for cross-modality medical image segmentation. Our framework contains two sequentially conducted stages, *i.e.*, Prototype-anchored Feature Alignment (PFA) stage and Contrastive Learning (CL) stage. As previous works [12] noted, the weights of the pre-trained classifier (*i.e.*, projection head) can be employed as the source prototypes during domain adaptation. That means we can characterize the features of each class with a source prototype and align the target features with them instead of the inaccessible source features. To that end, during the PFA stage, we first provide a target-to-prototype transport to ensure the target features get close to the corresponding prototypes. Then, considering the trivial solution that all target features are assigned to the dominant class prototype (*e.g.*, background), we add a reverse prototype-to-target transport to encourage diversity. However, although most target features have been assigned to the correct class prototype after PFA, some hard samples with high prediction uncertainty still exist in the decision boundary (see Fig. 1(a→b)). Moreover, we observe that those unreliable predictions usually get confused among only a few classes instead of all classes [18]. Taking the unreliable pixel in Fig. 1(b, c) for example, though it achieves similar high probabilities on the spleen and left kidney, the model is pretty sure about this pixel not belonging to the liver and right kidney. Inspired by this, we use confusing pixels as the negative samples for those unlikely classes, and then introduce the CL stage to pursue a more compact target feature distribution. Finally, we conduct experiments on a cross-modality abdominal multi-organ segmentation task. With only a source model and unlabeled target data, our method outperforms the state-of-the-art SFDA and even achieves comparable results with some classical UDA approaches.

2   Methods

We are first provided a segmentation model $\mathcal{M}^s$ trained on N_s labeled samples $\{(x_n^s, y_n^s)\}_{n=1}^{N_s}$ from the source domain $\mathcal{D}^s$, and an unlabeled dataset with N_t samples $\{x_m^t\}_{m=1}^{N_t}$ from the target domain $\mathcal{D}^t$, where $x^s, x^t \in \mathbb{R}^{H \times W \times D}$, $y_n^s \in \mathbb{R}^{H \times W}$, H and W are the height and width of the samples. The goal of SFDA is to adapt the source model $\mathcal{M}^s$ with only unlabeled x^t to predict pixel-wise label y^t for the target domain data. In general, the segmentation model consists of two parts: 1) a feature extractor $F_\theta : x_i \to \boldsymbol{f}_i \in \mathbb{R}^{D_f}$, parameterized by θ, mapping each pixel $i \in \{1, \cdots, H \times W\}$ in image x to the feature $\boldsymbol{f}_i$ in the embedding space; 2) a one-layer pixel-wise classifier $\phi : \boldsymbol{f}_i \to \boldsymbol{p}_i \in \mathbb{R}^C$, that projects pixel feature into the semantic label space with C classes.

In the SFDA task, the source classifier ϕ^s encounters a domain shift problem when classifying the target domain feature. To tackle this challenge, we propose a novel SFDA framework mainly including two stages, shown in Fig. 2. We will elaborate on the details in the following.
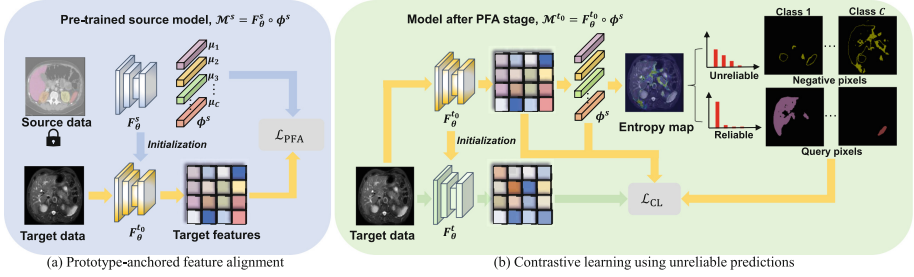
(a) Prototype-anchored feature alignment   (b) Contrastive learning using unreliable predictions

Fig. 2. An overview of the proposed two-stage SFDA framework. (a) is the first PFA stage. We freeze the classifier ϕ^s and use its weights for prototype-anchored feature alignment. (b) is the following CL stage. Given a target image, we first use $\mathcal{M}^{t_0}$ to make a prediction, and separate the pixels into query one and negative ones for each class based on their reliability (entropy). Then, features of query pixels come from F_θ^t (query samples), while features of negative pixels are from $F_\theta^{t_0}$ (negative samples), when minimizing $\mathcal{L}_{\mathrm{CL}}$.

2.1 Prototype-Anchored Feature Alignment

Since source data is not available, explicit feature alignment that directly minimizes the domain gap between the source and target data like many UDA methods [4,8] is inoperative. As shown by previous methods [12], the weights $[\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \cdots, \boldsymbol{\mu}_C] \in \mathbb{R}^{D_f \times C}$ of the source domain classifier ϕ^s can be interpreted as the source prototypes, which characterize the features of each class. Thus, we introduce a bi-directional transport cost to align the target features with these prototypes instead of the unaccessible source features.

Following [23], given a mini-batch $\{x_m^t\}_{m=1}^M$ with M images, we first adopt the cosine distance $d(\boldsymbol{\mu}_c, \boldsymbol{f}_{m,i}^t) = 1 - \langle \boldsymbol{\mu}_c, \boldsymbol{f}_{m,i}^t \rangle$ to define a point-to-point transport cost between $\boldsymbol{f}_{m,i}^t$ and $\boldsymbol{\mu}_c$, where $\langle \cdot, \cdot \rangle$ is the cosine similarity. Then, a conditional distribution $\pi_\theta\left(\boldsymbol{\mu}_c \mid \boldsymbol{f}_{m,i}^t\right)$ specifying the probability of transporting from $\boldsymbol{f}_{m,i}^t$ to $\boldsymbol{\mu}_c$ can be constructed as,

$$\pi_\theta\left(\boldsymbol{\mu}_c \mid \boldsymbol{f}_{m,i}^t\right) = \frac{\hat{p}\left(\boldsymbol{\mu}_c\right) \exp\left(\boldsymbol{\mu}_c^T \boldsymbol{f}_{m,i}^t / \tau\right)}{\sum_{c'=1}^C \hat{p}\left(\boldsymbol{\mu}_{c'}\right) \exp\left(\boldsymbol{\mu}_{c'}^T \boldsymbol{f}_{m,i}^t / \tau\right)} \tag{1}$$

where τ is the temperature parameter, and $\hat{p}\left(\boldsymbol{\mu}_c\right)$ is the prior distribution (*i.e.*, class proportion) over the C classes for the target domain. As the true class distribution is unavailable in the target domain, we use the EM algorithm to infer $\hat{p}\left(\boldsymbol{\mu}_c\right)$ instead of using a uniform prior distribution (see more details in [16]). Note that in Eq. 1, a target point is more likely to be transported to the class prototypes closer to it or those with higher class propotion.

With the conditional distribution and point-to-point transport cost, we can derive the target-to-prototype (T2P) expected cost of moving the target features

in this mini-batch to source prototypes,

$$\mathcal{L}_{\text{T2P}} = \frac{1}{M \times H \times W} \sum_{m=1}^{M} \sum_{i=1}^{H \times W} \sum_{c=1}^{C} d(\boldsymbol{\mu}_c, \boldsymbol{f}_{m,i}^t) \pi_\theta \left(\boldsymbol{\mu}_c \mid \boldsymbol{f}_{m,i}^t \right) \quad (2)$$

In this target-to-prototype direction, we assign each target pixel to the prototypes according to their similarities and the class distribution. However, like many entropy minimization methods [1,2], optimizing target-to-prototype cost alone may result in degenerate trivial solutions, biasing the prediction towards a single dominant class [16]. To avoid mapping most of the target features to only a few prototypes, we add a prototype-to-target (P2T) transport cost in the opposite direction, which ensures that each prototype can be assigned to some target features. Similarly, we have:

$$\mathcal{L}_{\text{P2T}} = \sum_{c=1}^{C} \hat{p}(\boldsymbol{\mu}_c) \sum_{m=1}^{M} \sum_{i=1}^{H \times W} d(\boldsymbol{\mu}_c, \boldsymbol{f}_{m,i}^t) \frac{\exp\left(\boldsymbol{\mu}_c^T \boldsymbol{f}_{m,i}^t/\tau\right)}{\sum_{m'=1}^{M} \sum_{i'=1}^{H \times W} \exp\left(\boldsymbol{\mu}_c^T \boldsymbol{f}_{m',i'}^t/\tau\right)} \quad (3)$$

Then, combining the conditional transport cost in these two directions, we define the total prototype-anchored feature alignment (PFA) loss:

$$\mathcal{L}_{\text{PFA}} = \mathcal{L}_{\text{T2P}} + \mathcal{L}_{\text{P2T}} \quad (4)$$

Similar to [6], we initialize the adaptation model $\mathcal{M}^{t_0}$ with the pre-trained source model $\mathcal{M}^s$ and fix the weights of the classifier during adaptation.

2.2 Contrastive Learning Using Unreliable Predictions

After the PFA stage, the clusters of target features are shifted towards their corresponding source prototypes, which brings remarkable improvements for the initial noisy prediction (see Fig. 3(b)). To further improve the compactness of the target feature distribution, previous self-training methods mainly focus on strengthening the reliability of pseudo labels by developing denoising strategies [5,20], but discard those low-confidence predictions. However, such contempt for unreliable predictions may result in information loss. For example, in Fig. 1(c), the probability of the unreliable pixel hovers between spleen and left kidney, yet is confident enough to indicate the categories it does not belong to.

With this intuition, we denote $\boldsymbol{p}_{m,i}^t$ as the softmax probabilities generated by model $\mathcal{M}^{t_0}$ for the target data $x_{m,i}^t$. Then, for each class c, we construct three components, named query samples, positive prototypes, and negative samples, to explore those unreliable predictions as [18].

Query Samples. During training, we employ the per-pixel entropy as uncertainty metric [18], and sample the pixels with low entropy (reliable pixel) in the current mini-batch as query candidates. We denote the set of features of all query pixels for class c as $\mathcal{P}_c$,

$$\mathcal{P}_c = \{ \boldsymbol{f}_{m,i}^t \mid \mathcal{H}(\boldsymbol{p}_{m,i}^t) \leq \gamma_c, \ \arg\max_{c'} \boldsymbol{p}_{m,i}^t = c \} \quad (5)$$

where $\mathcal{H}(\cdot)$ is the entropy of the input probabilities and γ_c is the entropy threshold for class c. Here we set γ_c as the α_c-th percentile of all the entropy values of pixels assigned a pseudo label c.

Positive Prototypes. The positive prototype is the same for all query pixels from the same class. Instead of using the center of query samples like [18], we set them the same as the previous source prototype, which is denoted as $\boldsymbol{z}_c^+ = \boldsymbol{\mu}_c$.

Negative Samples. For a query sample from class c, its qualified negative samples should satisfy: 1) unreliable; 2) highly probable not belong to class c. Therefore, we introduce the pixel-level category order $\mathcal{O}_{m,i}^t = \mathrm{argsort}(\boldsymbol{p}_{m,i}^t)$. For example, we have $\mathcal{O}_{m,i}^t(\arg\max \boldsymbol{p}_{m,i}^t) = 1$ and $\mathcal{O}_{m,i}^t(\arg\min \boldsymbol{p}_{m,i}^t) = C$. Thus, we can use $\mathcal{O}_{m,i}^t(c)$ to define the set of all negative samples:

$$\mathcal{N}_c = \{\boldsymbol{f}_{m,i}^t \mid \mathcal{H}(\boldsymbol{p}_{m,i}^t) > \gamma_c, \ \mathcal{O}_{m,i}^t(c) \geq r_l\} \tag{6}$$

where r_l is the low rank threshold and is set to 3 in our task.

With the above definition, we have the pixel-level contrastive loss as:

$$\mathcal{L}_{\mathrm{CL}} = -\frac{1}{C \times K} \sum_{c=1}^{C} \sum_{k=1}^{K} \log \left[\frac{e^{\langle z_{c,k}, z_c^+ \rangle / \tau}}{e^{\langle z_{c,k}, z_c^+ \rangle / \tau} + \sum_{j=1}^{N} e^{\langle z_{c,k}, z_{c,k,j}^- \rangle / \tau}} \right] \tag{7}$$

where K is the number of query samples, and $\boldsymbol{z}_{c,k} \in \mathcal{P}_c$ denotes the k-th query sample from class c. Each query sample is paired with a positive prototype $\boldsymbol{z}_c^+$ and N negative samples $\boldsymbol{z}_{c,k,j}^- \in \mathcal{N}_c$.

3   Experiments and Results

3.1   Experimental Setup

Datasets and Evaluation Metrics. We evaluate our SFDA approach on a cross-modality abdominal multi-organ segmentation task. For the abdominal datasets, we obtain 20 MRI volumes from the 2019 CHAOS Challenge [10] and 30 CT volumes from MICCAI 2015 [11], respectively. Both datasets are under the Creative Commons Attribution 4.0 International license and involve segmentation masks for the following abdominal organs: liver, right kidney, left kidney and spleen. We complete adaptation experiments both in the "MRI to CT" direction and in the "CT to MRI" direction. For the "MRI to CT" direction, we take the MRI modality to train the source model and vice verse. Both modalities are randomly divided into 80% for domain adaptation training and 20% for evaluation. For both datasets, we discard the axial slices that do not contain foreground and crop out the non-body region [3]. The value range in CT volumes is first clipped to $[-125, 275]$. Then min-max normalization has been performed on both datasets to normalize the intensity value to $[0, 1]$. After that, all the MRI and CT volumes are uniformly resized to 256×256 in axial plane. Due to the large variance in the slice thickness of CT and MRI modality, we split the volume into slices for the model training.

For the evaluation, two main metrics, dice similarity coefficient (Dice) and average symmetric surface distance (ASSD) are used to quantitatively evaluate the segmentation results [4,15].

Implementation Details. We adopt classic U-Net structure for the segmentation model as the previous work [1]. The source segmentation model is trained in a fully-supervised manner for 10k iterations. During adaptation, we use Adam optimizer with the learning rate 1×10^{-4} and a weight decay of 5×10^{-4}. The temperature τ and batch size is set as 0.1 and 16, respectively. In PFA stage, we freeze the classifier and optimize $F_\theta^{t_0}$ for 200 iterations. In CL stage, we empirically set hyper-parameters $\alpha_c = 80$, $K = 64$, and $N = 256$ for all classes. All experiments are conducted with PyTorch on a single NVIDIA RTX 3090 GPU of 24 GB memory. Data augmentation such as random cropping, rotation, and brightness are adopted for source domain training and target domain adaptation.

3.2 Results of Source-Free Domain Adaptation

Comparision with Other Methods. In our experiments, "no adaptation" lower bound denotes learning a model on the source domain and directly test on the target domain without adaptation. And "supervised" upper bound means training and testing in the same target domain. We compared our methods with recent SFDA methods all designed for medical image segmentation scenarios, including a denoised pseudo-labeling approach (DPL) [5], a prior-aware entropy minimization approach (AdaMI) [1], a fourier style mining approach (FSM) [21], and a feature map statistics-guided approach [9]. We also considered top-performing UDA methods (*i.e.*, SIFA [4], DAG-Net [19]). For a fair comparison, we utilized the same backbone for these methods [1,4,5,21] and reimplemented them according to their official codes. Note that we reported the results of methods [9,19] from papers, since their official codes were not released.

The quantitative evaluation results are presented in Table 1. Compared to the upper and lower bounds in both directions, a huge performance gap can be observed due to the severe domain shifts between MRI and CT modalities. In "MRI to CT" direction, our method remarkably outperforms all other SFDA approaches on the right kidney and spleen, achieving the highest average Dice
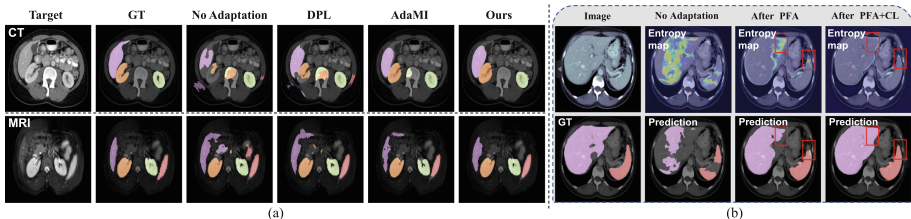

Fig. 3. (a) Qualitative segmentation results of different methods for abdominal images. (b) Visualized evolution of the model uncertainty and predictions in different stages.

Table 1. Comparision with other methods on abdominal multi-organ datasets.

| Method	Source-free	Dice (%, mean ± std) ↑					ASSD (voxel, mean ± std) ↓				
		Liver	R. Kidney	L. Kidney	Spleen	Avg	Liver	R. Kidney	L. Kidney	Spleen	Avg
Abdominal MRI → Abdominal CT											
No Adaptation	–	49.2±9.6	43.4±24.9	65.4±18.1	62.4±20.9	55.1±14.3	8.7±2.0	10.8±4.6	7.4±3.1	4.5±2.9	7.9±2.5
Supervised	–	93.5±1.2	91.3±1.2	92.1±1.2	91.1±5.6	92.0±2.7	0.7±0.1	0.6±0.1	0.6±0.2	0.5±0.2	0.6±0.2
SIFA [4]	✗	89.0±3.2	83.8±4.0	82.7±5.8	84.6±8.0	85.0±5.9	1.2±0.5	1.3±0.6	1.5±0.7	1.6±0.9	1.4±0.8
DAG-Net [19]	✗	84.8±4.6	85.9±3.9	86.7±3.6	88.1±7.4	86.4±4.9	1.6±0.6	1.1±0.5	1.2±0.8	0.9±0.7	1.2±0.7
DPL [5]	✔	70.1±6.9	52.9±14.2	65.7±12.5	70.9±13.2	64.9±8.8	4.6±2.0	7.9±3.1	7.5±2.8	3.2±2.6	5.8±2.0
FSM [21]	✔	83.2±3.8	74.5±4.5	75.1±4.6	76.2±9.8	77.3±7.0	2.8±0.8	4.2±1.6	5.0±1.9	2.6±1.4	3.7±1.3
AdaMI [1]	✔	**90.2 ± 1.0**	81.4±3.3	82.6±4.7	80.2±7.1	83.6±6.6	**1.0 ± 0.5**	1.8±0.9	1.6±1.0	2.4±1.2	1.7±0.9
Hong et al [9]	✔	88.1	80.8	**88.1**	79.2	84.1	–	–	–	–	–
Ours	✔	89.9±2.7	**84.5 ± 6.8**	84.9±4.0	**85.2 ± 7.8**	**86.1 ± 6.3**	1.3±0.5	**1.4 ± 0.4**	**1.3 ± 0.6**	**1.2 ± 0.8**	**1.4 ± 0.7**
Abdominal CT → Abdominal MRI											
No Adaptation	–	66.5±6.8	81.6±14.6	78.8±19.5	70.1±11.6	74.3±11.0	3.5±1.6	1.9±1.0	3.0±1.6	5.1±2.3	3.4±2.1
Supervised	–	93.6±3.5	94.2±1.8	90.4±3.7	92.2±2.6	92.6±2.1	0.5±0.2	0.3±0.1	0.4±0.1	0.4±0.2	0.4±0.2
SIFA [4]	✗	87.1±4.6	89.1±2.8	84.2±3.9	88.3±2.4	87.2±3.8	1.6±0.6	0.8±0.2	1.5±0.8	1.7±0.9	1.5±0.8
DAG-Net [19]	✗	86.3±3.3	89.0±2.0	89.9±2.0	90.6±2.9	89.0±2.6	1.8±0.6	0.9±0.2	0.9±0.4	1.3±1.0	1.2±0.6
DPL [5]	✔	77.2±2.0	83.5±12.5	80.3±13.6	83.3±7.2	81.1±8.3	3.0±0.6	2.3±1.6	2.5±1.4	3.1±1.7	2.7±1.4
FSM [21]	✔	84.3±3.1	83.5±6.6	82.1±9.0	84.2±5.6	83.5±4.3	2.1±0.5	1.5±0.6	1.6±0.7	2.2±0.6	1.9±0.6
AdaMI [1]	✔	85.5±2.4	86.4±5.1	82.1±7.6	89.9±3.2	86.0±3.1	**1.9 ± 0.4**	1.2±0.5	2.2±0.6	**1.3 ± 0.7**	1.7±0.5
Hong et al [9]	✔	**88.4**	89.1	86.4	**91.1**	88.8	–	–	–	–	–
Ours	✔	86.1±0.5	**91.7 ± 5.1**	**88.6 ± 8.0**	90.4±2.2	**89.2 ± 3.3**	2.0±0.3	**0.7 ± 0.2**	**1.0 ± 0.4**	1.5±0.5	**1.3 ± 0.5**
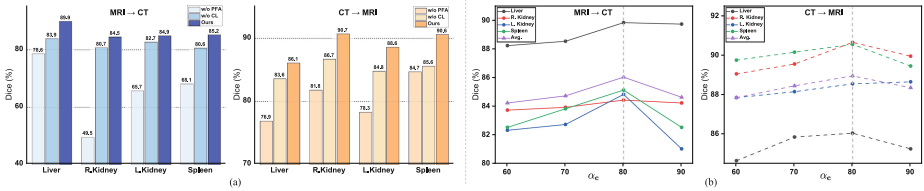
Fig. 4. (a) Ablation analysis of proposed two SFDA stages. "w/o CL" denotes only the PFA is performed; "w/o PFA" denotes directly optimizing the contrastive loss according to the source model prediction. (b) Effect of different uncertainty percentile α_c on the adaptation performance.

of 86.1% and the lowest average ASSD of 1.4. Moreover, compared with recent UDA methods, our method obtains competitive results on average Dice and ASSD, which may be due to the use of unreliable predictions. As for "CT to MRI" direction, our method similarly shows great superiority on most organs as well, achieving the best performance in terms of both the average Dice (89.2%) and ASSD (1.3) among all SFDA methods. Figure 3(a) shows the segmentation results obtained by existing and our methods in both modalities. As observed, DPL is prone to amplify the initial noisy regions since it directly discards the unreliable pixels in self-training. For comparison, our method substantially rectificate the uncertain regions from the initial prediction, and details are shown in Fig. 3(b).

Ablation Study. In Fig. 4(a), we verify the effectiveness of the proposed two SFDA stages by removing each stage while keeping the other. The consecutive two stage adaptation leads to the best performance, while the drop in Dice is more significant if we remove the PFA stage. This result is not surprising

because, without PFA, the source model prediction is too noisy to sample the qualified query and negative pixels for contrastative learning. We also study the impact of different uncertainty percentile α_c in Fig. 4(b). This parameter has a certain impact on performance, and we find $\alpha_c = 80\%$ achieves the best performance for most organs. Large α_c may introduce low-confidence query samples for supervision, and small α_c will drop some informative negative samples.

4   Conclusion

In this paper, we propose a novel two-stage framework to address the source-free domain adaptation problem in medical image segmentation. We first introduce a bi-directional transport cost to encourage the alignment between target features and source class prototypes in the prototype-anchored feature alignment stage. Also, a contrastive learning stage using unreliable predictions is further devised to learn a more compact target feature distribution. Sufficient experiments on the cross-modality abdominal multi-organ segmentation task validate the effectiveness and superiority of our method against other strong SFDA baselines, even some classical UDA approaches.

References

1. Bateson, M., Kervadec, H., Dolz, J., Lombaert, H., Ayed, I.B.: Source-free domain adaptation for image segmentation. Med. Image Anal. **82**, 102617 (2022)
2. Bateson, M., Lombaert, H., Ben Ayed, I.: Test-time adaptation with shape moments for image segmentation. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) MICCAI 2022. LNCS, vol. 13434, pp. 736–745. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16440-8_70
3. Bian, C., Yuan, C., Ma, K., Yu, S., Wei, D., Zheng, Y.: Domain adaptation meets zero-shot learning: an annotation-efficient approach to multi-modality medical image segmentation. IEEE Trans. Med. Imaging **41**(5), 1043–1056 (2021)
4. Chen, C., Dou, Q., Chen, H., Qin, J., Heng, P.A.: Unsupervised bidirectional cross-modality adaptation via deeply synergistic image and feature alignment for medical image segmentation. IEEE Trans. Med. Imaging **39**(7), 2494–2505 (2020)
5. Chen, C., Liu, Q., Jin, Y., Dou, Q., Heng, P.-A.: Source-free domain adaptive fundus image segmentation with denoised pseudo-labeling. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12905, pp. 225–235. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87240-3_22
6. Ding, N., Xu, Y., Tang, Y., Xu, C., Wang, Y., Tao, D.: Source-free domain adaptation via distribution estimation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 7212–7222 (2022)
7. Dou, Q., Ouyang, C., Chen, C., Chen, H., Heng, P.A.: Unsupervised cross-modality domain adaptation of convnets for biomedical image segmentations with adversarial loss. In: Proceedings of the 27th International Joint Conference on Artificial Intelligence, pp. 691–697 (2018)
8. Han, X., et al.: Deep symmetric adaptation network for cross-modality medical image segmentation. IEEE Trans. Med. Imaging **41**(1), 121–132 (2021)

9. Hong, J., Zhang, Y.D., Chen, W.: Source-free unsupervised domain adaptation for cross-modality abdominal multi-organ segmentation. Knowl.-Based Syst. 109155 (2022)

10. Kavur, A.E., et al.: Chaos challenge-combined (CT-MR) healthy abdominal organ segmentation. Med. Image Anal. **69**, 101950 (2021)

11. Landman, B., Xu, Z., Igelsias, J., Styner, M., Langerak, T., Klein, A.: MICCAI multi-atlas labeling beyond the cranial vault-workshop and challenge. In: Proceedings of MICCAI Multi-Atlas Labeling Beyond Cranial Vault-Workshop Challenge, vol. 5, p. 12 (2015)

12. Liu, Y., Chen, Y., Dai, W., Gou, M., Huang, C.T., Xiong, H.: Source-free domain adaptation with contrastive domain alignment and self-supervised exploration for face anti-spoofing. In: Avidan, S., Brostow, G., Cissé, M., Farinella, G.M., Hassner, T. (eds.) ECCV 2022. LNCS, vol. 13672, pp. 511–528. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-19775-8_30

13. Motiian, S., Jones, Q., Iranmanesh, S., Doretto, G.: Few-shot adversarial domain adaptation. In: Advances in Neural Information Processing Systems, vol. 30 (2017)

14. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28

15. Stan, S., Rostami, M.: Privacy preserving domain adaptation for semantic segmentation of medical images. arXiv preprint arXiv:2101.00522 (2021)

16. Tanwisuth, K., et al.: A prototype-oriented framework for unsupervised domain adaptation. Adv. Neural. Inf. Process. Syst. **34**, 17194–17208 (2021)

17. Tzeng, E., Hoffman, J., Saenko, K., Darrell, T.: Adversarial discriminative domain adaptation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7167–7176 (2017)

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

19. Xian, J., et al.: Unsupervised cross-modality adaptation via dual structural-oriented guidance for 3D medical image segmentation. IEEE Trans. Med. Imaging (2023)

20. Xu, Z., et al.: Denoising for relaxing: unsupervised domain adaptive fundus image segmentation without source data. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) MICCAI 2022. LNCS, vol. 13435, pp. 214–224. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16443-9_21

21. Yang, C., Guo, X., Chen, Z., Yuan, Y.: Source free domain adaptation for medical image segmentation with fourier style mining. Med. Image Anal. **79**, 102457 (2022)

22. Yu, Q., Dang, K., Tajbakhsh, N., Terzopoulos, D., Ding, X.: A location-sensitive local prototype network for few-shot medical image segmentation. In: 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI), pp. 262–266. IEEE (2021)

23. Zheng, H., Zhou, M.: Exploiting chain rule and Bayes' theorem to compare probability distributions. Adv. Neural. Inf. Process. Syst. **34**, 14993–15006 (2021)

24. Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N., Liang, J.: UNet++: a nested U-net architecture for medical image segmentation. In: Stoyanov, D., et al. (eds.) DLMIA/ML-CDS -2018. LNCS, vol. 11045, pp. 3–11. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00889-5_1