ACT: Semi-supervised Domain-Adaptive Medical Image Segmentation with Asymmetric Co-training

Xiaofeng Liu[1](✉), Fangxu Xing[1], Nadya Shusharina[2], Ruth Lim[1], C.-C. Jay Kuo[3], Georges El Fakhri[1], and Jonghye Woo[1]

[1] Gordon Center for Medical Imaging, Department of Radiology, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114, USA
xliu61@mgh.harvard.edu

[2] Division of Radiation Biophysics, Department of radiation Oncology, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114, USA

[3] Department of Electrical and Computer Engineering, University of Southern California, Los Angeles, CA 90007, USA

Abstract. Unsupervised domain adaptation (UDA) has been vastly explored to alleviate domain shifts between source and target domains, by applying a well-performed model in an unlabeled target domain via supervision of a labeled source domain. Recent literature, however, has indicated that the performance is still far from satisfactory in the presence of significant domain shifts. Nonetheless, delineating a few target samples is usually manageable and particularly worthwhile, due to the substantial performance gain. Inspired by this, we aim to develop semi-supervised domain adaptation (SSDA) for medical image segmentation, which is largely underexplored. We, thus, propose to exploit both labeled source and target domain data, in addition to unlabeled target data in a unified manner. Specifically, we present a novel asymmetric co-training (ACT) framework to integrate these subsets and avoid the domination of the source domain data. Following a divide-and-conquer strategy, we explicitly decouple the label supervisions in SSDA into two asymmetric sub-tasks, including semi-supervised learning (SSL) and UDA, and leverage different knowledge from two segmentors to take into account the distinction between the source and target label supervisions. The knowledge learned in the two modules is then adaptively integrated with ACT, by iteratively teaching each other, based on the confidence-aware pseudo-label. In addition, pseudo label noise is well-controlled with an exponential MixUp decay scheme for smooth propagation. Experiments on cross-modality brain tumor MRI segmentation tasks using the BraTS18 database showed, even with limited labeled target samples, ACT yielded marked improvements over UDA and state-of-the-art SSDA methods and approached an "upper bound" of supervised joint training.

1 Introduction

Accurate delineation of lesions or anatomical structures is a vital step for clinical diagnosis, intervention, and treatment planning [24]. While recently flourished

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022
L. Wang et al. (Eds.): MICCAI 2022, LNCS 13435, pp. 66–76, 2022.
https://doi.org/10.1007/978-3-031-16443-9_7

deep learning methods excel at segmenting those structures, deep learning-based segmentors cannot generalize well in a heterogeneous domain, e.g., different clinical centers, scanner vendors, or imaging modalities [4,14,16,20]. To alleviate this issue, unsupervised domain adaptation (UDA) has been actively developed, by applying a well-performed model in an unlabeled target domain via supervision of a labeled source domain [5,15,18,19]. Due to diverse target domains, however, the performance of UDA is far from satisfactory [9,17,31]. Instead, labeling a small set of target domain data is usually more feasible [25]. As such, semi-supervised domain adaptation (SSDA) has shown great potential as a solution to domain shifts, as it can utilize both labeled source and target data, in addition to unlabeled target data. To date, while several SSDA classification methods have been proposed [8,13,23,29], based on discriminative class boundaries, they cannot be directly applied to segmentation, since segmentation involves complex and dense pixel-wise predictions.

Recently, while a few works [6,10,26] have been proposed to extend SSDA for segmentation on natural images, to our knowledge, no SSDA for medical image segmentation has yet been explored. For example, a depth estimation for natural images is used as an auxiliary task as in [10], but that approach cannot be applied to medical imaging data, e.g., MRI, as they do not have perspective depth maps. Wang et al. [26] simply added supervision from labeled target samples to conventional adversarial UDA. Chen et al. [6] averaged labeled source and target domain images at both region and sample levels to mitigate the domain gap. However, source domain supervision can easily dominate the training, when we directly combine the labeled source data with the target data [23]. In other words, the extra small amount of labeled target data has not been effectively utilized, because the volume of labeled source data is much larger than labeled target data, and there is significant divergence across domains [23].

To mitigate the aforementioned limitations, we propose a practical asymmetric co-training (ACT) framework to take each subset of data in SSDA in a unified and balanced manner. In order to prevent a segmentor, jointly trained by both domains, from being dominated by the source data only, we adopt a divide-and-conquer strategy to decouple the label supervisions for the two asymmetric segmentors, which share the same objective of carrying out a decent segmentation performance for the unlabeled data. By "asymmetric," we mean that the two segmentors are assigned different roles to utilize the labeled data in either source or target domain, thereby providing a complementary view for the unlabeled data. That is, the first segmentor learns on the labeled source domain data and unlabeled target domain data as a conventional UDA task, while the other segmentor learns on the labeled and unlabeled target domain data as a semi-supervised learning (SSL) task. To integrate these two asymmetric branches, we extend the idea of co-training [1,3,22], which is one of the most established multi-view learning methods. Instead of modeling two views on the same set of data with different feature extractors or adversarial sample generation in conventional co-training [1,3,22], our two cross-domain views are explicitly provided by the segmentors with the correlated and complementary UDA and SSL tasks.

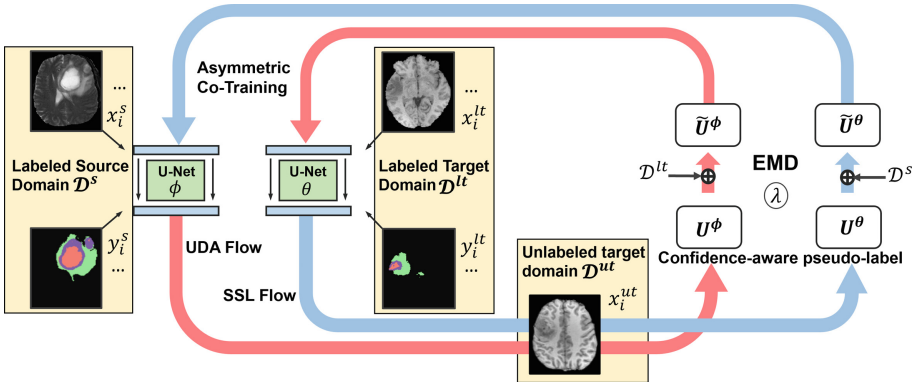

Fig. 1. Illustration of our proposed ACT framework for SSDA cross-modality (e.g., T2-weighted to T1-weighted MRI) image segmentation. Note that only target domain specific segmentor θ will be used in testing.

Specifically, we construct the pseudo label of the unlabeled target sample based on the pixel-wise confident predictions of the other segmentor. Then, the segmentors are trained on the pseudo labeled data iteratively with an exponential MixUp decay (EMD) scheme for smooth propagation. Finally, the target segmentor carries out the target domain segmentation.

The contributions of this work can be summarized as follows:

- We present a novel SSDA segmentation framework to exploit the different supervisions with the correlated and complementary asymmetric UDA and SSL sub-tasks, following a divide-and-conquer strategy. The knowledge is then integrated with confidence-aware pseudo-label based co-training.
- An EMD scheme is further proposed to mitigate the noisy pseudo label in early epochs of training for smooth propagation.
- To our knowledge, this is the first attempt at investigating SSDA for medical image segmentation. Comprehensive evaluations on cross-modality brain tumor (i.e., T2-weighted MRI to T1-weighted/T1ce/FLAIR MRI) segmentation tasks using the BraTS18 database demonstrate superiority performance over conventional source-relaxed/source-based UDA methods.

2 Methodology

In our SSDA setting for segmentation, we are given a labeled source set $\mathcal{D}^s = \{(x_i^s, y_i^s)\}_{i=1}^{N^s}$, a labeled target set $\mathcal{D}^{lt} = \{(x_i^{lt}, y_i^{lt})\}_{i=1}^{N^{lt}}$, and an unlabeled target set $\mathcal{D}^{ut} = \{(x_i^{ut})\}_{i=1}^{N^{ut}}$, where N^s, N^{lt}, and N^{ut} are the number of samples for each set, respectively. Note that the slice x_i^s, x_i^{lt}, and x_i^{ut}, and the segmentation mask labels y_i^s, and y_i^{lt} have the same spatial size of $H \times W$. In addition, for each pixel $y_{i:n}^s$ or $y_{i:n}^{lt}$ indexed by $n \in \mathbb{R}^{H \times W}$, the label has C classes, i.e.,

$y^s_{i:n}, y^{lt}_{i:n} \in \{1, \cdots, C\}$. There is a distribution divergence between source domain samples, $\mathcal{D}^s$, and target domain samples, $\mathcal{D}^{lt}$ and $\mathcal{D}^{ut}$. Usually, N^{lt} is much smaller than N^s. The learning objective is to perform well in the target domain.

2.1 Asymmetric Co-training for SSDA Segmentation

To decouple SSDA via a divide-and-conquer strategy, we integrate $\mathcal{D}^{ut}$ with either $\mathcal{D}^s$ or $\mathcal{D}^{lt}$ to form the correlated and complementary sub-tasks of UDA and SSL. We configure a cross-domain UDA segmentor ϕ and a target domain SSL segmentor θ, which share the same objective of achieving a decent segmentation performance in $\mathcal{D}^{ut}$. The knowledge learned from the two segmentors is then integrated with ACT. The overall framework of this work is shown in Fig. 1.

Conventional co-training has focused on two independent views of the source and target data or generated artificial multi-views with adversarial examples, which learns two classifiers for each of the views and teaches each other on the unlabeled data [3,22]. By contrast, in SSDA, without multiple views of the data, we propose to leverage the distinct yet correlated supervision, based on the inherent discrepancy of the labeled source and target data. We note that the sub-tasks and datasets adopted are different for the UDA and SSL branches. Therefore, all of the data subsets can be exploited, following well-established UDA and SSL solutions without interfering with each other.

To achieve co-training, we adopt a simple deep pseudo labeling method [27], which assigns the pixel-wise pseudo label $\hat{y}_{i:n}$ for $x^{ut}_{i:n}$. Though UDA and SSL can be achieved by different advanced algorithms, deep pseudo labeling can be applied to either UDA [32] or SSL [27]. Therefore, we can apply the same algorithm to the two sub-tasks, thereby greatly simplifying our overall framework. We note that while a few methods [28] can be applied to either SSL or UDA like pseudo labeling, they have not been jointly adopted in the context of SSDA.

Specifically, we assign the pseudo label for each pixel $x^{ut}_{i:n}$ in $\mathcal{D}^{ut}$ with the prediction of either ϕ or θ, therefore constructing the pseudo labeled sets U^ϕ and U^θ for the training of another segmentor θ and ϕ, respectively:

$$U^\phi = \{(x^{ut}_{i:n}, \hat{y}^\phi_{i:n} = \arg\max_c p(c|x^{ut}_{i:n};\phi));\ \text{if}\ \max_c p(c|x^{ut}_{i:n};\phi) > \epsilon\}, \tag{1}$$

$$U^\theta = \{(x^{ut}_{i:n}, \hat{y}^\theta_{i:n} = \arg\max_c p(c|x^{ut}_{i:n};\theta));\ \text{if}\ \max_c p(c|x^{ut}_{i:n};\theta) > \epsilon\}, \tag{2}$$

where $p(c|x^{ut}_{i:n};\theta)$ and $p(c|x^{ut}_{i:n};\phi)$ are the predicted probability of class $c \in \{1, \cdots, C\}$ w.r.t. $x^{ut}_{i:n}$ using θ and ϕ, respectively. ϵ is a confidence threshold. Note that the low softmax prediction probability indicates the low confidence for training [18,32]. Then, the pixels in the selected pseudo label sets are merged with the labeled data to construct $\{\mathcal{D}^s, U^\theta\}$ and $\{\mathcal{D}^{lt}, U^\phi\}$ for the training of ϕ and θ with a conventional supervised segmentation loss, respectively. Therefore, the two segmentors with asymmetrical tasks act as teacher and student of each other to distillate the knowledge with highly confident predictions.

Algorithm 1: An iteration of the ACT algorithm.

Input: batch size N, λ, η, ϵ, $\mathcal{D}^s$, $\mathcal{D}^{lt}$, $\mathcal{D}^{ut}$, current network parameters ω_ϕ, ω_θ;
Sample $\{(x_i^s, y_i^s)\}_{i=1}^N$, $\{(x_i^{lt}, y_i^{lt})\}_{i=1}^N$, and$\{(x_i^{ut})\}_{i=1}^N$from $\mathcal{D}^s$, $\mathcal{D}^{lt}$, and$\mathcal{D}^{ut}$, respectively;
Initialize $U^\phi = \emptyset$, $U^\theta = \emptyset$;
for $i \leftarrow 1$ **to** N **do**
  $\hat{y}_{i:n}^\phi = \arg\max_c p(c|x_{i:n}^{ut}; \phi)$; and $\hat{y}_{i:n}^\theta = \arg\max_c p(c|x_{i:n}^{ut}; \theta)$
  if $\max_c p(c|x_{i:n}^{ut}; \phi) > \epsilon$: **update** $U^\phi \leftarrow U^\phi \cup \{(x_{i:n}^{ut}, \hat{y}_{i:n}^\theta)$ with Eq. (1);
  if $\max_c p(c|x_{i:n}^{ut}; \theta) > \epsilon$: **update** $U^\theta \leftarrow U^\theta \cup \{(x_{i:n}^{ut}, \hat{y}_{i:n}^\theta)$ with Eq. (2);
end
Obtain $\tilde{U}^\phi = \{\text{EMD}(U_i^\phi, \{(x_i^{lt}, y_i^{lt})\}_{i=1}^N; \lambda)\}_{i=1}^{|U^\phi|\times N}$ with Eq. (3);
Obtain $\tilde{U}^\theta = \{\text{EMD}(U_i^\theta, \{(x_i^s, y_i^s)\}_{i=1}^N; \lambda)\}_{i=1}^{|U^\theta|\times N}$ with Eq. (4);
Update $\omega_\phi \leftarrow \omega_\phi - \eta\nabla(\mathcal{L}(\omega_\phi, \mathcal{D}^s) + \mathcal{L}(\omega_\phi, \tilde{U}^\theta))$; $\omega_\theta \leftarrow \omega_\theta - \eta\nabla(\mathcal{L}(\omega_\theta, \mathcal{D}^{lt}) + \mathcal{L}(\omega_\theta, \tilde{U}^\phi))$;
Output: Updated network parameters ω_ϕ and ω_θ.

2.2 Pseudo-label with Exponential MixUp Decay

Initially generated pseudo labels with the two segmentors are typically noisy, which is significantly acute in the initial epochs, thus leading to a deviated solution with propagated errors. Numerous conventional co-training methods relied on simple assumptions that there is no domain shift, and the predictions of the teacher model can be reliable and be simply used as ground truth. Due to the domain shift, however, the prediction of ϕ in the target domain could be noisy and lead to an aleatoric uncertainty [7,11,12]. In addition, insufficient labeled target domain data can lead to an epistemic uncertainty related to the model parameters [7,11,12].

To smoothly exploit the pseudo labels, we propose to adjust the contribution of the supervision signals from both labels and pseudo labels as the training progresses. Previously, vanilla MixUp [30] was developed for efficient data augmentation, by combining both samples and their labels to generate new data for training. We note that the MixUp used in SSL [2,6] adopted a constant sampling, and did not take the decay scheme for gradual co-training. Thus, we propose to gradually exploit the pseudo label by mixing up $\mathcal{D}^s$ or $\mathcal{D}^{lt}$ with pseudo labeled $\mathcal{D}^{ut}$, and adjust their ratio with the EMD scheme. For the selected U^ϕ and U^θ with the number of slices $|U^\phi|$ and $|U^\theta|$, we mix up each pseudo labeled image with all images from $\mathcal{D}^s$ or $\mathcal{D}^{lt}$ to form the mixed pseudo labeled sets $\tilde{U}^\theta$ and $\tilde{U}^\phi$. Specifically, our EMD can be formulated as:

$$\tilde{U}^\phi = \{(\tilde{x}_{i:n}^{lt} = \lambda x_{i:n}^{lt} + (1-\lambda)x_{i:n}^{ut}, \lambda\tilde{y}_{i:n}^{lt} = \lambda y_{i:n}^{lt} + (1-\lambda)\hat{y}_{i:n}^\theta)\}_i^{|U^\theta|\times N}, \quad (3)$$

$$\tilde{U}^\theta = \{(\tilde{x}_{i:n}^{s} = \lambda x_{i:n}^{s} + (1-\lambda)x_{i:n}^{ut}, \lambda\tilde{y}_{i:n}^{s} = \lambda y_{i:n}^{s} + (1-\lambda)\hat{y}_{i:n}^\phi)\}_i^{|U^\phi|\times N}, \quad (4)$$

where $\lambda = \lambda^0\exp(-I)$ is the MixUp parameter with the exponential decay w.r.t. iteration I. λ^0 is the initial weight of ground truth samples and labels, which is empirically set to 1. Therefore, along with the increase over iteration I, we have smaller λ, which adjusts the contribution of the ground truth label to be large at the start of the training, while utilizing the pseudo labels at the later training epochs. Therefore, $\tilde{U}^\phi$ and $\tilde{U}^\theta$ gradually represent the pseudo label sets

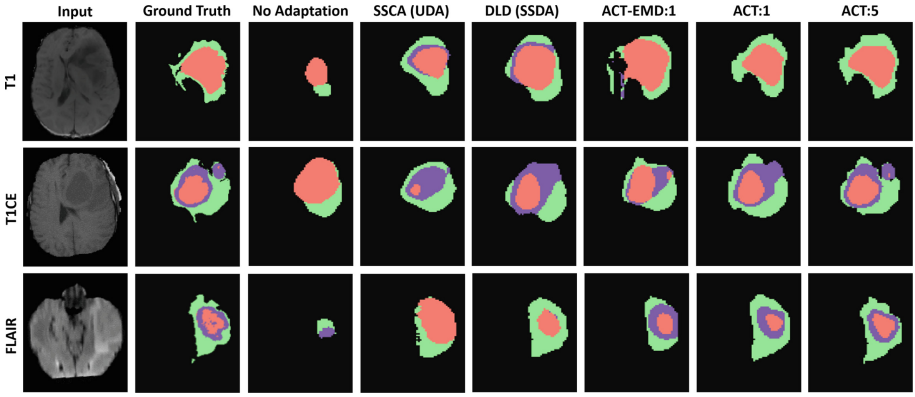

Fig. 2. Comparisons with other UDA/SSDA methods and ablation studies for the cross-modality tumor segmentation. We show target test slices of T1, T1ce, and FLAIR MRI from three subjects.

of U^{ϕ} and U^{θ}. We note that the mixup operates on the image level, which is indicated by i. The number of generated mixed samples depends on the scale of U^{ϕ} and U^{θ} in each iteration and batch size N. With the labeled $\mathcal{D}^s$, $\mathcal{D}^{lt}$, as well as the pseudo labeled sets with EMD $\tilde{U}^{\phi}$ and $\tilde{U}^{\phi}$, we update the parameters of the segmentors ϕ and θ, i.e., ω_{ϕ} and ω_{θ} with SGD as:

$$\omega_{\phi} \leftarrow \omega_{\phi} - \eta\nabla(\mathcal{L}(\omega_{\phi}, \mathcal{D}^s) + \mathcal{L}(\omega_{\phi}, \tilde{U}^{\theta})), \tag{5}$$

$$\omega_{\theta} \leftarrow \omega_{\theta} - \eta\nabla(\mathcal{L}(\omega_{\theta}, \mathcal{D}^{lt}) + \mathcal{L}(\omega_{\theta}, \tilde{U}^{\phi})), \tag{6}$$

where η indicates the learning rate, and $\mathcal{L}(\omega_{\phi}, \mathcal{D}^s)$ denotes the learning loss on $\mathcal{D}^s$ with the current segmentor ϕ parameterized by ω_{ϕ}. The training procedure is detailed in Algorithm 1. After training, only the target domain specific SSL segmentor θ is used for testing.

3 Experiments and Results

To demonstrate the effectiveness of our proposed SSDA method, we evaluated our method on T2-weighted MRI to T1-weighted/T1ce/FLAIR MRI brain tumor segmentation using the BraTS2018 database [21]. We denote our proposed method as ACT, and used ACT-EMD for an ablation study of an EMD-based pseudo label exploration.

Of note, the BraTS2018 database contains a total of 285 patients [21] with the MRI scannings, including T1-weighted (T1), T1-contrast enhanced (T1ce), T2-weighted (T2), and T2 Fluid Attenuated Inversion Recovery (FLAIR) MRI. For the segmentation labels, each pixel belongs to one of four classes, i.e., enhancing tumor (EnhT), peritumoral edema (ED), necrotic and non-enhancing tumor core (CoreT), and background. In addition, the whole tumor covers CoreT, EnhT,

Table 1. Whole tumor segmentation performance of the cross-modality UDA and SSDA. The supervised joint training can be regarded as an "upper bound".

Method		DICE	Score	(DSC)	[%] ↑	Hausdorff	Distance	(HD)	[mm] ↓
	Task	**T1**	**FLAIR**	**T1CE**	**Ave**	**T1**	**FLAIR**	**T1CE**	**Ave**
Source only	No DA	4.2	65.2	6.3	27.7 ± 1.2	55.7	28.0	49.8	39.6 ± 0.5
Target only	SSL:5	43.8	54.6	47.5	48.6 ± 1.7	31.9	29.6	35.4	32.3 ± 0.8
SIFA [5]	UDA	51.7	68.0	58.2	59.3 ± 0.6	19.6	16.9	15.0	17.1 ± 0.4
DSFN [31]	UDA	57.3	78.9	62.2	66.1 ± 0.8	17.5	13.8	15.5	15.6 ± 0.3
DSA [9]	UDA	57.7	81.8	62.0	67.2 ± 0.7	14.2	8.6	13.7	12.2 ± 0.4
SSCA [17]	UDA	59.3	82.9	63.5	68.6 ± 0.6	12.5	7.9	11.2	11.5 ± 0.3
SLA [26]	SSAD:1	64.7	82.3	66.1	71.0 ± 0.5	12.2	7.1	10.5	9.9 ± 0.3
DLD [6]	SSAD:1	65.8	81.5	66.5	71.3 ± 0.6	12.0	7.1	10.3	9.8 ± 0.2
ACT	SSAD:1	**69.7**	**84.5**	**69.7**	**74.6 ± 0.3**	**10.5**	**5.8**	**10.0**	**8.8 ± 0.1**
ACT-EMD	SSAD:1	67.4	83.9	69.0	73.4 ± 0.6	10.9	6.4	10.3	9.2 ± 0.2
ACT	SSAD:5	**71.3**	**85.0**	**70.8**	**75.7 ± 0.5**	**10.0**	**5.2**	**9.8**	**8.3 ± 0.1**
ACT-EMD	SSAD:5	70.3	84.4	69.8	74.8 ± 0.4	10.4	5.7	10.2	8.8 ± 0.2
Joint training	Supervised	73.2	85.6	72.6	77.1 ± 0.5	9.5	4.6	9.2	7.7 ± 0.2

and ED. We follow the conventional cross-modality UDA (i.e., T2-weighted to T1-weighted/T1ce/FLAIR) evaluation protocols [9,17,31] for 8/2 splitting for training/testing, and extend it to our SSDA task, by accessing the labels of 1–5 target domain subjects at the adaptation training stage. All of the data were used in a subject-independent and unpaired manner. We used SSDA:1 or SSDA:5 to denote that one or five target domain subjects are labeled in training.

For a fair comparison, we used the same segmentor backbone as in DSA [9] and SSCA [17], which is based on Deeplab-ResNet50. Without loss of generality, we simply adopted the cross-entropy loss as $\mathcal{L}$, and set the learning rate $\eta = 1\text{e}{-3}$ and confidence threshold $\epsilon = 0.5$. Both ϕ and θ have the same network structure. For the evaluation metrics, we adopted the widely used DSC (the higher, the better) and Hausdorff distance (HD: the lower, the better) as in [9,17]. The standard deviation was reported over five runs.

The quantitative evaluation results of the whole tumor segmentation are provided in Table 1. We can see that SSDA largely improved the performance over the compared UDA methods [9,17]. For the T2-weighted to T1-weighted MRI transfer task, we were able to achieve more than 10% improvements over [9,17] with only one labeled target sample. Recent SSDA methods for natural image segmentation [6,26] did not take the balance between the two labeled supervisions into consideration, easily resulting in a source domain-biased solution in case of limited labeled target domain data, and thus did not perform well on target domain data [23]. In addition, the depth estimation in [10] cannot be applied to the MRI data. Thus, we reimplemented the aforementioned methods [6,26] with the same backbone for comparisons, which is also the first attempt at the medical image segmentation. Our ACT outperformed [6,26] by a DSC of 3.3%

Table 2. Detailed comparison of Core/EnhT/ED segmentation. Results are averaged over three tasks including T2-weighted to T1-weighted, T1CE, and FLAIR MRI with the backbone as in [9,17].

Method		DICE	Score (DSC)	[%] ↑	Hausdorff	Distance (HD)	[mm] ↓
	Task	**CoreT**	**EnhT**	**ED**	**CoreT**	**EnhT**	**ED**
Source only	No DA	20.6 ± 1.0	39.5 ± 0.8	41.3 ± 0.9	54.7 ± 0.4	55.2 ± 0.6	42.5 ± 0.4
Target only	SSL:5	27.3 ± 1.1	38.0 ± 1.0	$40.2.3 \pm 1.3$	51.8 ± 0.7	52.3 ± 0.9	46.4 ± 0.6
DSA [9]	UDA	57.8 ± 0.6	44.0 ± 0.6	56.8 ± 0.5	25.8 ± 0.4	34.2 ± 0.3	25.6 ± 0.5
SSCA [17]	UDA	58.2 ± 0.4	44.5 ± 0.5	60.7 ± 0.4	26.4 ± 0.2	32.8 ± 0.2	23.4 ± 0.3
SLA [26]	SSDA:1	58.9 ± 0.6	48.1 ± 0.5	65.4 ± 0.4	24.5 ± 0.1	27.6 ± 0.3	20.3 ± 0.2
DLD [6]	SSDA:1	60.3 ± 0.6	48.2 ± 0.5	66.0 ± 0.3	24.2 ± 0.2	27.8 ± 0.1	19.7 ± 0.2
ACT	SSDA:1	$\mathbf{64.5 \pm 0.3}$	$\mathbf{52.7 \pm 0.4}$	$\mathbf{69.8 \pm 0.6}$	$\mathbf{20.0 \pm 0.2}$	$\mathbf{24.6 \pm 0.1}$	$\mathbf{16.2 \pm 0.2}$
ACT	SSDA:5	$\mathbf{66.9 \pm 0.3}$	$\mathbf{54.0 \pm 0.3}$	$\mathbf{71.2 \pm 0.5}$	$\mathbf{18.4 \pm 0.4}$	$\mathbf{23.7 \pm 0.2}$	$\mathbf{15.1 \pm 0.2}$
Joint training	Supervised	70.4 ± 0.3	62.5 ± 0.2	75.1 ± 0.4	15.8 ± 0.2	22.7 ± 0.1	13.0 ± 0.2

w.r.t. the averaged whole tumor segmentation in SSDA:1 task. The better performance of ACT over ACT-EMD demonstrated the effectiveness of our EMD scheme for smooth adaptation with pseudo-label. We note that we did not manage to outperform the supervised joint training, which accesses all of the target domain labels, which can be considered an "upper bound" of UDA and SSDA. Therefore, it is encouraging that our ACT can approach joint training with five labeled target subjects. In addition, the performance was stable for the setting of λ from 1 to 10.

In Table 2, we provide the detailed comparisons for more fine-grained segmentation w.r.t. CoreT, EnhT, and ED. The improvements were consistent with the whole tumor segmentation. The qualitative results of three target modalities in Fig. 2 show the superior performance of our framework, compared with the comparison methods.

In Fig. 3(a), we analyzed the testing pixel proportion change along with the training that has both, only one, and none of two segmentor pseudo-labels, i.e., the maximum confidence is larger than ϵ as in Eq. (1). We can see that the consensus of the two segmentors keeps increasing, by teaching each other in the co-training scheme for knowledge integration. "Both" low rates, in the beginning, indicate ϕ and θ may provide a different view based on their asymmetric tasks, which can be complementary to each other. The sensitivity studies of using a different number of labeled target domain subjects are shown in Fig. 3(b). Our ACT was able to effectively use $\mathcal{D}^{lt}$. In Fig. 3(c), we show that using more EMD pairs improves the performance consistently.

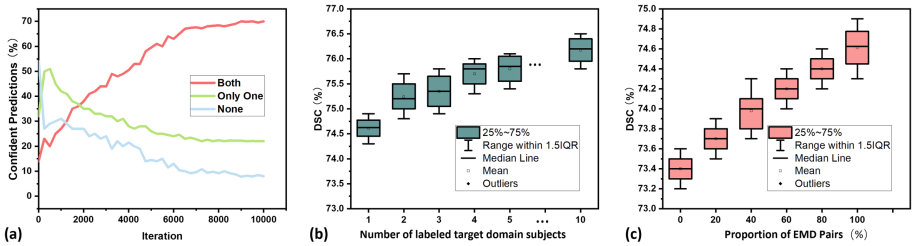

Fig. 3. Analysis of our ACT-based SSDA on the whole tumor segmentation task. (a) The proportion of testing pixels that both, only one, or none of the segmentors have high confidence on (b) the performance improvements with a different number of labeled target domain training subjects, and (c) a sensitivity study of changing different proportion of EMD pairs of $|\tilde{U}^{\phi}| \times N$ and $|\tilde{U}^{\theta}| \times N$.

4 Conclusion

This work proposed a novel and practical SSDA framework for the segmentation task, which has the great potential to improve a target domain generalization with a manageable labeling effort in clinical practice. To achieve our goal, we resorted to a divide-and-conquer strategy with two asymmetric sub-tasks to balance between the supervisions from source and target domain labeled samples. An EMD scheme is further developed to exploit the pseudo-label smoothly in SSDA. Our experimental results on the cross-modality SSDA task using the BraTS18 database demonstrated that the proposed method surpassed the state-of-the-art UDA and SSDA methods.

Acknowledgement. This work is supported by NIH R01DC018511, R01DE027989, and P41EB022544.

References

1. Balcan, M.F., Blum, A., Yang, K.: Co-training and expansion: towards bridging theory and practice. Adv. Neural Inf. Process. Syst. **17**, 89–96 (2005)
2. Berthelot, D., Carlini, N., Goodfellow, I., Papernot, N., Oliver, A., Raffel, C.: Mixmatch: a holistic approach to semi-supervised learning. arXiv preprint arXiv:1905.02249 (2019)
3. Blum, A., Mitchell, T.: Combining labeled and unlabeled data with co-training. In: Proceedings of the Eleventh Annual Conference on Computational Learning Theory, pp. 92–100 (1998)
4. Che, T., et al.: Deep verifier networks: verification of deep discriminative models with deep generative models. ArXiv (2019)
5. Chen, C., Dou, Q., Chen, H., Qin, J., Heng, P.A.: Synergistic image and feature adaptation: towards cross-modality domain adaptation for medical image segmentation. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, pp. 865–872 (2019)

6. Chen, S., Jia, X., He, J., Shi, Y., Liu, J.: Semi-supervised domain adaptation based on dual-level domain mixing for semantic segmentation. In: CVPR, pp. 11018–11027 (2021)
7. Der Kiureghian, A., Ditlevsen, O.: Aleatory or epistemic? Does it matter? Struct. Saf. **31**(2), 105–112 (2009)
8. Donahue, J., Hoffman, J., Rodner, E., Saenko, K., Darrell, T.: Semi-supervised domain adaptation with instance constraints. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 668–675 (2013)
9. Han, X., et al.: Deep symmetric adaptation network for cross-modality medical image segmentation. IEEE Trans. Med. Imaging (2022)
10. Hoyer, L., Dai, D., Wang, Q., Chen, Y., Van Gool, L.: Improving semi-supervised and domain-adaptive semantic segmentation with self-supervised depth estimation. arXiv preprint arXiv:2108.12545 (2021)
11. Hu, S., Worrall, D., Knegt, S., Veeling, B., Huisman, H., Welling, M.: Supervised uncertainty quantification for segmentation with multiple annotations. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11765, pp. 137–145. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32245-8_16
12. Kendall, A., Gal, Y.: What uncertainties do we need in Bayesian deep learning for computer vision? arXiv preprint arXiv:1703.04977 (2017)
13. Kim, T., Kim, C.: Attract, perturb, and explore: learning a feature alignment network for semi-supervised domain adaptation. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) ECCV 2020. LNCS, vol. 12359, pp. 591–607. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58568-6_35
14. Kong, L., Hu, B., Liu, X., Lu, J., You, J., Liu, X.: Constraining pseudo-label in self-training unsupervised domain adaptation with energy-based model. Int. J. Intell. Syst. (2022)
15. Liu, X., et al.: Domain generalization under conditional and label shifts via variational Bayesian inference. IJCAI (2021)
16. Liu, X., Li, S., Ge, Y., Ye, P., You, J., Lu, J.: Recursively conditional gaussian for ordinal unsupervised domain adaptation. In: International Conference on Computer Vision (ICCV), October 2021
17. Liu, X., Xing, F., Fakhri, G.E., Woo, J.: Self-semantic contour adaptation for cross modality brain tumor segmentation. In: IEEE International Symposium on Biomedical Imaging (ISBI) (2022)
18. Liu, X., et al.: Generative self-training for cross-domain unsupervised tagged-to-cine MRI synthesis. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12903, pp. 138–148. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87199-4_13
19. Liu, X., et al.: Unsupervised black-box model domain adaptation for brain tumor segmentation. Front. Neurosci., 341 (2022)
20. Liu, X., et al.: Deep unsupervised domain adaptation: a review of recent advances and perspectives. APSIPA Trans. Signal Inf. Process. (2022)
21. Menze, B.H., et al.: The multimodal brain tumor image segmentation benchmark (BRATS). IEEE Trans. Medical Imaging **34**(10), 1993–2024 (2014)
22. Qiao, S., Shen, W., Zhang, Z., Wang, B., Yuille, A.: Deep co-training for semi-supervised image recognition. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) ECCV 2018. LNCS, vol. 11219, pp. 142–159. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01267-0_9
23. Saito, K., Kim, D., Sclaroff, S., Darrell, T., Saenko, K.: Semi-supervised domain adaptation via minimax entropy. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 8050–8058 (2019)

24. Tajbakhsh, N., Jeyaseelan, L., Li, Q., Chiang, J.N., Wu, Z., Ding, X.: Embracing imperfect datasets: a review of deep learning solutions for medical image segmentation. Med. Image Anal. **63**, 101693 (2020)
25. van Engelen, J.E., Hoos, H.H.: A survey on semi-supervised learning. Mach. Learn. **109**(2), 373–440 (2019). https://doi.org/10.1007/s10994-019-05855-6
26. Wang, Z., et al.: Alleviating semantic-level shift: a semi-supervised domain adaptation method for semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp. 936–937 (2020)
27. Wei, C., Shen, K., Chen, Y., Ma, T.: Theoretical analysis of self-training with deep networks on unlabeled data. arXiv preprint arXiv:2010.03622 (2020)
28. Xia, Y., et al.: Uncertainty-aware multi-view co-training for semi-supervised medical image segmentation and domain adaptation. Med. Image Anal. **65**, 101766 (2021)
29. Yao, T., Pan, Y., Ngo, C.W., Li, H., Mei, T.: Semi-supervised domain adaptation with subspace learning for visual recognition. In: Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, pp. 2142–2150 (2015)
30. Zhang, H., Cisse, M., Dauphin, Y.N., Lopez-Paz, D.: mixup: Beyond empirical risk minimization. arXiv preprint arXiv:1710.09412 (2017)
31. Zou, D., Zhu, Q., Yan, P.: Unsupervised domain adaptation with dualscheme fusion network for medical image segmentation. In: Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI-20, International Joint Conferences on Artificial Intelligence Organization, pp. 3291–3298 (2020)
32. Zou, Y., Yu, Z., Liu, X., Kumar, B., Wang, J.: Confidence regularized self-training. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 5982–5991 (2019)