Adapting Off-the-Shelf Source Segmenter for Target Medical Image Segmentation

Xiaofeng Liu[1], **Fangxu Xing**[1], **Chao Yang**[2], **Georges El Fakhri**[1], **Jonghye Woo**[1]

[1]Gordon Center for Medical Imaging, Department of Radiology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, 02114

[2]Facebook Artificial Intelligence, Boston, MA, 02142

Abstract

Unsupervised domain adaptation (UDA) aims to transfer knowledge learned from a labeled source domain to an unlabeled and unseen target domain, which is usually trained on data from both domains. Access to the source domain data at the adaptation stage, however, is often limited, due to data storage or privacy issues. To alleviate this, in this work, we target source free UDA for segmentation, and propose to adapt an "off-the-shelf" segmentation model pre-trained in the source domain to the target domain, with an adaptive batch-wise normalization statistics adaptation framework. Specifically, the domain-specific low-order batch statistics, i.e., mean and variance, are gradually adapted with an exponential momentum decay scheme, while the consistency of domain shareable high-order batch statistics, i.e., scaling and shifting parameters, is explicitly enforced by our optimization objective. The transferability of each channel is adaptively measured first from which to balance the contribution of each channel. Moreover, the proposed source free UDA framework is orthogonal to unsupervised learning methods, e.g., self-entropy minimization, which can thus be simply added on top of our framework. Extensive experiments on the BraTS 2018 database show that our source free UDA framework outperformed existing source-relaxed UDA methods for the cross-subtype UDA segmentation task and yielded comparable results for the cross-modality UDA segmentation task, compared with a supervised UDA methods with the source data.

1   Introduction

Accurate tumor segmentation is a critical step for early tumor detection and intervention, and has been significantly improved with advanced deep neural networks (DNN) [25,18,10,9,17]. A segmentation model trained in a source domain, however, usually cannot generalize well in a target domain, e.g., data acquired from a new scanner or different clinical center, in implementation. Besides, annotating data in the new target domain is costly and even infeasible [11]. To address this, unsupervised domain adaptation (UDA) was proposed to transfer knowledge from a labeled source domain to unlabeled target domains [13].

The typical UDA solutions can be classified into three categories: statistic moment matching, feature/pixel-level adversarial learning [15,14,12], and self-training [33,16]. These UDA methods assume that the source domain data are available and usually trained together with target data. The source data, however, are often inaccessible, due to data storage or privacy issues, for cross-clinical center implementation [1]. Therefore, it is of great importance to apply an "off-the-shelf" source domain model, without access to the source data. For source-free classification UDA, Liang et al. [8] proposed to enforce the diverse predictions, while the diversity of neighboring pixels is not suited for the segmentation purpose. In addition, the class prototype [13] and variational inference methods [11] are not scalable for pixel-wise classification based segmentation. More importantly, without distribution alignment, these methods relied on unreliable noisy pseudo labeling.

Recently, the source relaxed UDA [1] was presented to pre-train an additional class ratio predictor in the source domain, by assuming that the class ratio, i.e., pixel proportion in segmentation, is invariant between source and target domains. At the adaptation stage, the class ratio was used as the only transferable knowledge. However, that work [1] has two limitations. First, the class ratio can be different between the two domains, due to label shift [11,13]. For example, a disease incident rate could vary between different countries, and tumor size could vary between different subtypes and populations. Second, the pre-trained class ratio predictor used in [1] is not typical for medical image segmentation, thereby requiring an additional training step using the data in the source domain.

In this work, to address the aforementioned limitations, we propose a practical UDA framework aimed at the source-free UDA for segmentation, without an additional network trained in the source domain or the unrealistic assumption of class ratio consistency between source and target domains. More specifically, our framework hinges on the batch-wise normalization statistics, which are easy to access and compute. Batch Normalization (BN) [6] has been a default setting in the most of modern DNNs, e.g., ResNet [5] and U-Net [30], for faster and more stable training. Notably, the BN statistics of the source domain are stored in the model itself. The low-order batch statistics, e.g., mean and variance, are domain-specific, due to the discrepancy of input data. To gradually adapt the low-order batch statistics from the source domain to the target domain, we develop a momentum-based progression scheme, where the momentum follows an exponential decay w.r.t. the adaptation iteration. For the domain shareable high-order batch statistics, e.g., scaling and shifting parameters, a high-order batch statistics consistent loss is applied to explicitly enforce the discrepancy minimization. The transferability of each channel is adaptively measured first, from which to balance the contribution of each channel. Moreover, the proposed unsupervised self-entropy minimization can be simply added on top of our framework to boost the performance further.

Our contributions are summarized as follows:

- To our knowledge, this is the first source relaxed or source free UDA framework for segmentation. We do not need an additional source domain network, or the unrealistic assumption of the class ratio consistency [1]. Our method only relies

on an "off-the-shelf" pre-trained segmentation model with BN in the source domain.

- The domain-specific and shareable batch-wise statistics are explored via the low-order statistics progression with an exponential momentum decay scheme and transferability adaptive high-order statistics consistency loss, respectively.

- Comprehensive evaluations on both cross-subtype (i.e., HGG to LGG) and cross-modality (i.e., T2 to T1/T1ce/FLAIR) UDA tasks using the BraTS 2018 database demonstrate the validity of our proposed framework and its superiority to conventional source-relaxed/source-based UDA methods.

2 Methodology

We assume that a segmentation model with BN is pre-trained with source domain data, and the batch statistics are inherently stored in the model itself. At the adaptation stage, we fine-tune the model based on the batch-wise statistics and the self-entropy (SE) of target data prediction. The overview of the different setups of conventional UDA and our "off-the-shelf (OS)" UDA is shown in Fig. 1. Below, we briefly revisit the BN in Subsec. 2.1 first and then introduce our OSUDA in Subsec. 2.2. The added unsupervised SE minimization and the overall training protocol are detailed in Subsec. 2.3.

2.1 Preliminaries on Batch Normalization

As a default setting in the most of modern DNNs, e.g., ResNet [5] and U-Net [30], Batch Normalization (BN) [6] normalizes the input feature in the l-th layer $f_l \in \mathbb{R}^{B \times H_l \times W_l \times C_l}$ within a batch in a channel-wise manner to have zero mean and unit variance. B denotes the number of images in a batch, and H_l, W_l, and C_l are the height, width, and channels of layer l. We have samples in a batch, with index $b \in \{1, \cdots, B\}$, spatial index $n \in \{1, \cdots, H_l \times W_l\}$, and channel index $c \in \{1, \cdots, C_l\}$. BN calculates the mean of each channel $\mu_{l,c} = \frac{1}{B \times H_l \times W_l} \sum_b^B \sum_n^{H_l \times W_l} f_{l,b,n,c}$, where $f_{l,b,n,c} \in \mathbb{R}$ is the feature value. The variance $\{\sigma^2\}_{l,c} = \frac{1}{B \times H_l \times W_l} \sum_b^B \sum_n^{H_l \times W_l} (f_{l,b,n,c} - \mu_{l,c})^2$. Then, the input feature is normalized as

$$\widetilde{f}_{l,b,n,c} = \gamma_{l,c}(f_{l,b,n,c} - \mu_{l,c}) / \sqrt{\{\sigma^2\}_{l,c} + \epsilon} + \beta_{l,c}, \tag{1}$$

where $\epsilon \in \mathbb{R}^+$ is a small scalar for numerical stability. $\gamma_{l,c}$ and $\beta_{l,c}$ are learnable scaling and shifting parameters, respectively.

In testing, the input is usually a single sample rather than a batch with B samples. Therefore, BN stores the exponentially weighted average of the batch statistics at the training stage and used it in testing. Specifically, the mean and variance over the training are tracked progressively, i.e.,

$$\bar{\mu}_{l,c}^k = (1 - \eta) \cdot \bar{\mu}_{l,c}^{k-1} + \eta \cdot \mu_{l,c}^k; \quad \{\bar{\sigma}^2\}_{l,c}^k = (1 - \eta) \cdot \{\bar{\sigma}^2\}_{l,c}^{k-1} + \eta \cdot \{\sigma^2\}_{l,c}^k, \tag{2}$$

where $\eta \in [0, 1]$ is a momentum parameter. After K training iterations, $\bar{\mu}_{l,c}^{K}$, $\{\bar{\sigma}^2\}_{l,c}^{K}$, $\gamma_{l,c}^{K}$, and $\beta_{l,c}^{K}$ are stored and used for testing normalization [6].

2.2 Adaptive source-relaxed batch-wise statistics adaptation

Early attempts of BN for UDA simply added BN in the target domain, without the interaction with the source domain [7]. Recent studies [2,20,26,19] indicated that the low-order batch statistics, i.e., mean $\mu_{l,c}$ and variance $\{\sigma^2\}_{l,c}$, are domain-specific, because of the divergence of cross-domain representation distributions. Therefore, brute-forcing the same mean and variance across domains can lead to a loss of expressiveness [29]. In contrast, after the low-order batch statistics discrepancy is partially reduced, with domain-specific mean and variance normalization, the high-order batch statistics, i.e., scaling and shifting parameters $\gamma_{l,c}$ and $\beta_{l,c}$, are shareable across domains [20,26].

However, all of the aforementioned methods [2,20,29,26,19] require the source data at the adaptation stage. To address this, in this work, we propose to mitigate the domain shift via the adaptive low-order batch statistics progression with momentum, and explicitly enforce the consistency of the high-order statistics in a source-relaxed manner.

Low-order statistics progression with an exponential momentum decay scheme.—In order to gradually learn the target domain-specific mean and variance, we propose an exponential low-order batch statistics decay scheme. We initialize the mean and variance in the target domain with the tracked $\bar{\mu}_{l,c}^{K}$ and $\{\bar{\sigma}^2\}_{l,c}^{K}$ in the source domain, which is similar to applying a model with BN in testing [6]. Then, we progressively update the mean and variance in the t-th adaptation iteration in the target domain as

$$\bar{\mu}_{l,c}^{t} = (1 - \eta^t) \cdot \mu_{l,c}^{t} + \eta^t \cdot \bar{\mu}_{l,c}^{K}; \quad \{\bar{\sigma}^2\}_{l,c}^{t} = (1 - \eta^t) \cdot \{\sigma^2\}_{l,c}^{t} + \eta^t \cdot \{\bar{\sigma}^2\}_{l,c}^{K}, \quad (3)$$

where $\eta^t = \eta^0 \exp(-t)$ is a target adaptation momentum parameter with an exponential decay w.r.t. the iteration $\mu_{l,c}^{t}$ and $\{\sigma^2\}_{l,c}^{t}$ are the mean and variance of the current target batch. Therefore, the weight of $\bar{\mu}_{l,c}^{K}$ and $\{\bar{\sigma}^2\}_{l,c}^{K}$ are smoothly decreased along with the target domain adaptation, while $\mu_{l,c}^{t}$ and $\{\sigma^2\}_{l,c}^{t}$ gradually represent the batch-wise low-order statistics of the target data.

Transferability adaptive high-order statistics consistency.—For the high-order batch statistics, i.e., the learned scaling and shifting parameters, we explicitly encourage its consistency between the two domains with the following high-order batch statistics (HBS) loss:

$$\mathscr{L}_{HBS} = \sum_{l}^{L} \sum_{c}^{C_l} (1 + \alpha_{l,c})\{ \, | \, \gamma_{l,c}^{K} - \gamma_{l,c}^{t} \, | \, + \, | \, \beta_{l,c}^{K} - \beta_{l,c}^{t} \, | \, \}, \quad (4)$$

where $\gamma_{l,c}^K$ and $\beta_{l,c}^K$ are the learned scaling and shifting parameters in the last iteration of pre-training in the source domain. $\gamma_{l,c}^t$ and $\beta_{l,c}^t$ are the learned scaling and shifting parameters in the t-th adaptation iteration. $\alpha_{l,c}$ is an adaptive parameter to balance between the channels.

We note that the domain divergence can be different among different layers and channels, and the channels with smaller divergence can be more transferable [22]. Accordingly, we would expect that the channels with higher transferability contribute more to the adaptation. In order to quantify the domain discrepancy in each channel, a possible solution is to measure the difference between batch statistics. In the source-relaxed UDA setting, we define the channel-wise source-target distance in the t-th adaptation iteration as

$$d_{l,c} = \left| \frac{\bar{\mu}_{l,c}^K}{\sqrt{\{\bar{\sigma}^2\}_{l,c}^K + \epsilon}} - \frac{\mu_{l,c}^t}{\sqrt{\{\sigma^2\}_{l,c}^t + \epsilon}} \right| . \tag{5}$$

Then, the transferability of each channel can be measured by $\alpha_{l,c} = \frac{L \times C \times (1 + d_{l,c})^{-1}}{\sum_l \sum_c (1 + d_{l,c})^{-1}}$.

Therefore, the more transferable channels will be assigned with higher importance, i.e., with larger weight $(1 + \alpha_{l,c})$ in $\mathscr{L}_{l,c}$.

2.3   Self-entropy minimization and overall training protocol

The training in the unlabeled target domain can also be guided by an unsupervised learning framework. The SE minimization is a widely used objective in modern DNNs to encourage the confident prediction, i.e., the maximum softmax value can be high [4,8,24,1]. SE for pixel segmentation is calculated by the averaged entropy of the classifier's softmax prediction given by

$$\mathscr{L}_{SE} = \frac{1}{B \times H_0 \times W_0} \sum_b^B \sum_n^{H_0 \times W_0} \{\delta_{b,n} \log \delta_{b,n}\}, \tag{6}$$

where H_0 and W_0 are the height and width of the input, and $\delta_{b,n}$ is the histogram distribution of the softmax output of the n-th pixel of the b-th image in a batch. Minimizing $\mathscr{L}_{SE}$ leads to the output close to a one-hot distribution.

At the source-domain pre-training stage, we follow the standard segmentation network training protocol. At the target domain adaptation stage, the overall training objective can be formulated as $\mathscr{L} = \mathscr{L}_{HBS} + \lambda \mathscr{L}_{SE}$, where λ is used to balance between the BN statistics matching and SE minimization. We note that a trivial solution of SE minimization is that all unlabeled target data could have the same one-hot encoding [4]. Thus, to stabilize the training, we linearly change the hyper-parameter λ from 10 to 0 in training.

3 Experiments and Results

The BraTS2018 database is composed of a total of 285 subjects [21], including 210 high-grade gliomas (HGG, i.e., glioblastoma) subjects, and 75 low-grade gliomas (LGG) subjects. Each subject has T1-weighted (T1), T1-contrast enhanced (T1ce), T2-weighted (T2), and T2 Fluid Attenuated Inversion Recovery (FLAIR) Magnetic Resonance Imaging (MRI) volumes with voxel-wise labels for the enhancing tumor (EnhT), the peritumoral edema (ED), and the necrotic and non-enhancing tumor core (CoreT). Usually, we denote the sum of EnhT, ED, and CoreT as the whole tumor. In order to demonstrate the effectiveness and generality of our OSUDA, we follow two UDA evaluation protocols using the BraTS2018 database, including HGG to LGG UDA [23] and cross-modality (i.e., T2 to T1/T1ce/FLAIR) UDA [32].

For evaluation, we adopted the widely used Dice similarity coefficient and Hausdorff distance metrics as in [32]. The Dice similarity coefficient (the higher, the better) measures the overlapping part between our prediction results and the ground truth. The Hausdorff distance (the lower, the better) is defined between two sets of points in the metric space.

3.1 Cross-subtype HGG to LGG UDA

HGG and LGG have different size and position distributions for tumor regions [23]. Following the standard protocol, we used the HGG training set (source domain) to pre-train the segmentation model and adapted it with the LGG training set (target domain) [23]. The evaluation was implemented in the LGG testing set. We adopted the same 2D U-Net backbone in [23], sliced 3D volumes into 2D axial slices with the size of 128×128, and concatenated all four MRI modalities to get a 4-channel input.

The quantitative evaluation results are shown in Table 1. Since the pixel proportion of each class is different between HGG and LGG domains, the class ratio-based CRUDA [1] only achieved marginal improvements with its unsupervised learning objective. We note that the Dice score of the core tumor was worse than the pre-trained source-only model, which can be the case of negative transfer [27]. Our proposed OSUDA achieved the state-of-the-art performance for source-relaxed UDA segmentation, approaching the performance of SEAT [23] with the source data, which can be seen as an "upper-bound."

We used OSUDA-AC and OSUDA-SE to indicate the OSUDA without the adaptive channel-wise weighting and self-entropy minimization, respectively. The better performance of OSUDA over OSUDA-AC and OSUDA-SE demonstrates the effectiveness of adaptive channel-wise weighting and self-entropy minimization. The illustration of the segmentation results is given in Fig. 2. We can see that the predictions of our proposed OSUDA are better than the no adaptation model. In addition, CRUDA [1] had a tendency to predict a larger area for the tumor; and the tumor core is often predicted for the slices without the core.

3.2 Cross-modality T2 to T1/T1ce/FLAIR UDA

Because of large appearance discrepancies between different MRI modalities, we further applied our framework to the cross-modality UDA task. Since clinical annotation of the whole tumor is typically performed on T2-weighted MRI, the typical cross-modality UDA

setting is to use T2-weighted MRI as the labeled source domain, and T1/T1ce/FLAIR MRI as the unlabeled target domains [32]. We followed the UDA training (80% subjects) and testing (20% subjects) split as in [32], and adopted the same single-channel input backbone. We note that the data were used in an unpaired manner [32].

The quantitative evaluation results are provided in Table 2. Our proposed OSUDA outperformed CRUDA [1] consistently. In addition, in CRUDA, the additional class ratio prediction model was required to be trained with the source data, which is prohibitive in many real-world cases. Furthermore, our OSUDA outperformed several UDA methods trained with the source data, e.g., CycleGAN [31] and SIFA [3], for the two metrics. The visual segmentation results of three target modalities are shown in Fig. 3, showing the superior performance of our framework, compared with the comparison methods.

4. Discussion and Conclusion

This work presented a practical UDA framework for the tumor segmentation task in the absence of the source domain data, only relying on the "off-the-shelf" pre-trained segmentation model with BN in the source domain. We proposed a low-order statistics progression with an exponential momentum decay scheme to gradually learn the target domain-specific mean and variance. The domain shareable high-order statistics consistency is enforced with our HBS loss, which is adaptively weighted based on the channel-wise transferability. The performance was further boosted with the unsupervised learning objective via self-entropy minimization. Our experimental results on the cross-subtype and cross-modality UDA tasks demonstrated that the proposed framework outperformed the comparison methods, and was robust to the class ratio shift.
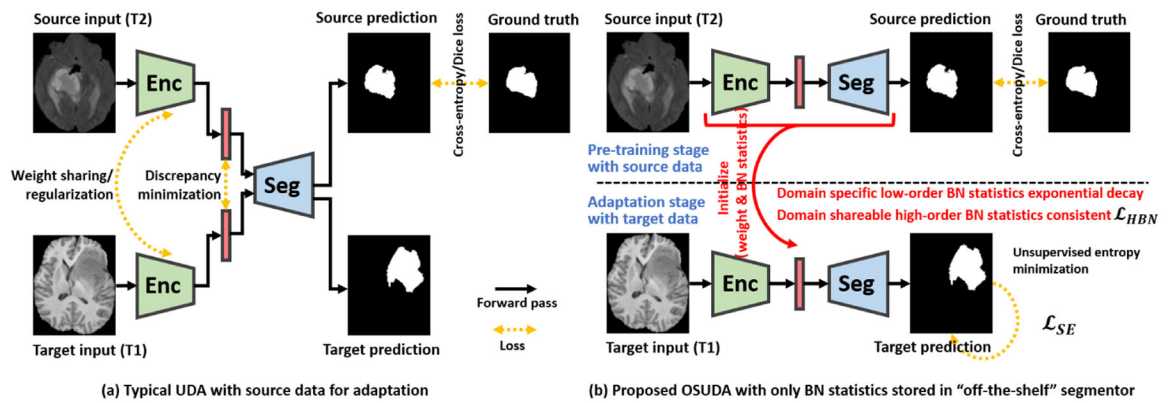
Acknowledgements

References

1. Bateson M, Kervadec H, Dolz J, Lombaert H, Ayed IB: Source-relaxed domain adaptation for image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 490–499. Springer (2020)

2. Chang WG, You T, Seo S, Kwak S, Han B: Domain-specific batch normalization for unsupervised domain adaptation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7354–7362 (2019)

3. Chen C, Dou Q, Chen H, Qin J, Heng PA: 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)

4. Grandvalet Y, Bengio Y: Semi-supervised learning by entropy minimization. In: NIPS (2005)

5. He K, Zhang X, Ren S, Sun J: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (6 2016)

6. Ioffe S, Szegedy C: Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: International conference on machine learning. pp. 448–456. PMLR (2015)

7. Li Y, Wang N, Shi J, Hou X, Liu J: Adaptive batch normalization for practical domain adaptation. Pattern Recognition 80, 109–117 (2018)

8. Liang J, Hu D, Feng J: Do we really need to access the source data? source hypothesis transfer for unsupervised domain adaptation. In: International Conference on Machine Learning. pp. 6028–6039. PMLR (2020)

9. Liu X, Fan F, Kong L, Diao Z, Xie W, Lu J, You J: Unimodal regularized neuron stick-breaking for ordinal classification. Neurocomputing 388, 34–44 (2020)

10. Liu X, Han X, Qiao Y, Ge Y, Li S, Lu J: Unimodal-uniform constrained wasserstein training for medical diagnosis. In: Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. pp. 0–0 (2019)

11. Liu X, Hu B, Jin L, Han X, Xing F, Ouyang J, Lu J, El Fakhri G, Woo J: Domain generalization under conditional and label shifts via variational bayesian inference. In: IJCAI (2021)

12. Liu X, Hu B, Liu X, Lu J, You J, Kong L: Energy-constrained self-training for unsupervised domain adaptation. ICPR (2020)

13. Liu X, Liu X, Hu B, Ji W, Xing F, Lu J, You J, Kuo CCJ, Fakhri GE, Woo J: Subtype-aware unsupervised domain adaptation for medical diagnosis. AAAI (2021)

14. Liu X, Xing F, El Fakhri G, Woo J: A unified conditional disentanglement framework for multimodal brain mr image translation. In: ISBI. pp. 10–14. IEEE (2021)

15. Liu X, Xing F, Prince JL, Carass A, Stone M, El Fakhri G, Woo J: Dual-cycle constrained bijective vae-gan for tagged-to-cine magnetic resonance image synthesis. In: ISBI. pp. 1448–1452. IEEE (2021)

16. Liu X, Xing F, Stone M, Zhuo J, Timothy R, Prince JL, El Fakhri G, Woo J: Generative self-training for cross-domain unsupervised tagged-to-cine mri synthesis. In: MICCAI (2021)

17. Liu X, Xing F, Yang C, Kuo CCJ, ElFakhri G, Woo J: Symmetric-constrained irregular structure inpainting for brain mri registration with tumor pathology. MICCAI BrainLes (2020)

18. Liu X, Zou Y, Song Y, Yang C, You J, K Vijaya Kumar B: Ordinal regression with neuron stick-breaking for medical diagnosis. In: Proceedings of the European Conference on Computer Vision (ECCV) Workshops. pp. 0–0 (2018)

19. Mancini M, Porzi L, Bulo SR, Caputo B, Ricci E: Boosting domain adaptation by discovering latent domains. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3771–3780 (2018)

20. Maria Carlucci F, Porzi L, Caputo B, Ricci E, Rota Bulo S: Autodial: Automatic domain alignment layers. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 5067–5075 (2017)

21. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, Burren Y, Porz N, Slotboom J, Wiest R, et al. : The multimodal brain tumor image segmentation benchmark (BRATS). IEEE transactions on medical imaging 34(10), 1993–2024 (2014) [PubMed: 25494501]

22. Pan X, Luo P, Shi J, Tang X: Two at once: Enhancing learning and generalization capacities via ibn-net. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 464–479 (2018)

23. Shanis Z, Gerber S, Gao M, Enquobahrie A: Intramodality domain adaptation using self ensembling and adversarial training. In: Domain Adaptation and Representation Transfer and Medical Image Learning with Less Labels and Imperfect Data, pp. 28–36. Springer (2019)

24. Wang D, Shelhamer E, Liu S, Olshausen B, Darrell T: Fully test-time adaptation by entropy minimization. arXiv preprint arXiv:2006.10726 (2020)

25. Wang J, Liu X, Wang F, Zheng L, Gao F, Zhang H, Zhang X, Xie W, Wang B: Automated interpretation of congenital heart disease from multi-view echocardiograms. Medical Image Analysis 69, 101942 (2021) [PubMed: 33418465]

26. Wang X, Jin Y, Long M, Wang J, Jordan M: Transferable normalization: Towards improving transferability of deep neural networks. arXiv preprint arXiv:2019 (2019)

27. Wang Z, Dai Z, Póczos B, Carbonell J: Characterizing and avoiding negative transfer. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11293–11302 (2019)

28. Wilson G, Cook DJ: A survey of unsupervised deep domain adaptation. ACM Transactions on Intelligent Systems and Technology (TIST) 11(5), 1–46 (2020)

29. Zhang J, Qi L, Shi Y, Gao Y: Generalizable semantic segmentation via model-agnostic learning and target-specific normalization. arXiv preprint arXiv:2003.12296 (2020)

30. Zhou XY, Yang GZ: Normalization in training u-net for 2-D biomedical semantic segmentation. IEEE Robotics and Automation Letters 4(2), 1792–1799 (2019)

31. Zhu JY, Park T, Isola P, Efros AA: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: ICCV (2017)

32. 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)

33. 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)

(a) Typical UDA with source data for adaptation

(b) Proposed OSUDA with only BN statistics stored in "off-the-shelf" segmentor

Fig. 1:
Comparison of (a) conventional UDA [28] and (b) our source-relaxed OSUDA segmentation framework based on the pre-trained "off-the-shelf" model with BN. We minimize the domain discrepancy based on the adaptively computed batch-wise statistics in each channel. The model consists of a feature encoder (Enc) and a segmentor (Seg) akin to [3,32].
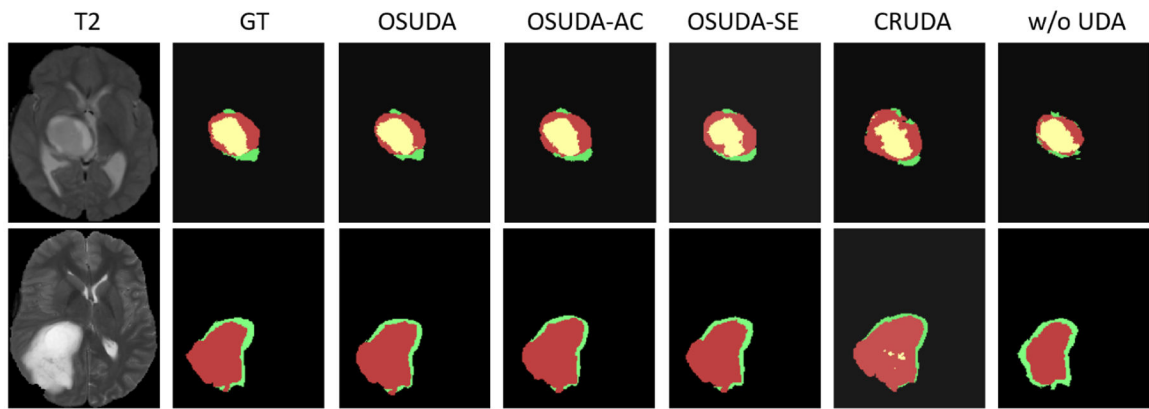
Fig. 2:
The comparison with the other UDA methods, and an ablation study of adaptive channel-wise weighting and SE minimization for HGG to LGG UDA.
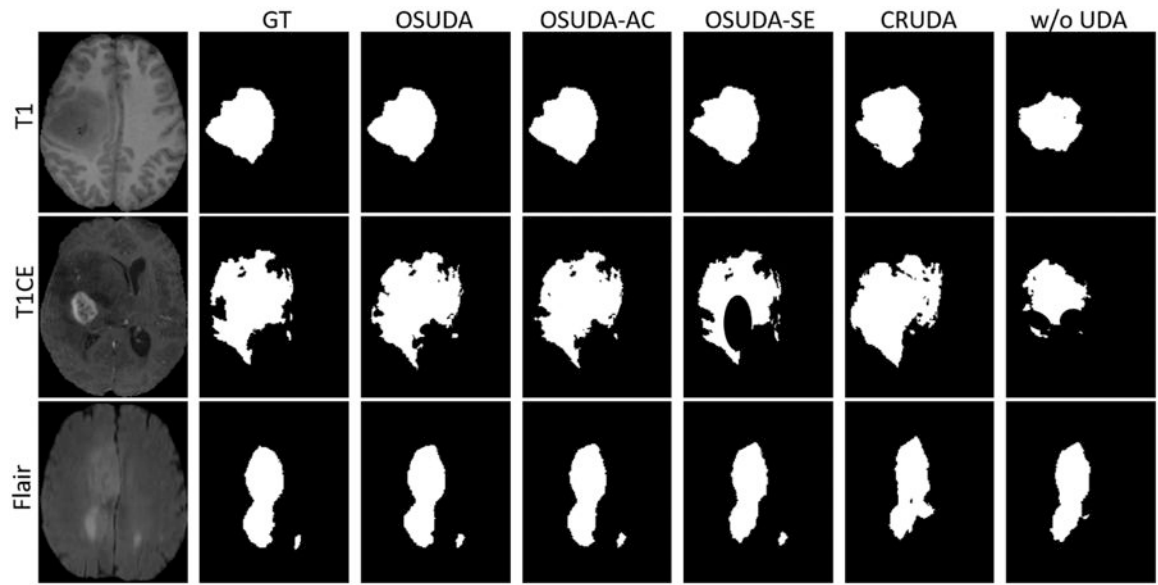
Fig. 3:
Comparison with the other UDA methods and an ablation study for the cross-modality whole tumor segmentation UDA task. From top to bottom, we show a target test slice of T1, T1ce, and FLAIR MRI.

Table 1:

Comparison of HGG to LGG UDA with the four-channel input for our four-class segmentation, i.e., whole tumor, enhanced tumor, core tumor, and background. ± indicates standard deviation. SEAT [23] with the source data for UDA training is regarded as an "upper bound."

Method	Source data	Dice Score [%] ↑				Hausdorff Distance [mm] ↓			
		WholeT	EnhT	CoreT	Overall	WholeT	EnhT	CoreT	Overall
Source only	no UDA	79.29	30.09	44.11	58.44±43.5	38.7	46.1	40.2	41.7±0.14
CRUDA [1]	Partial[3]	79.85	31.05	43.92	58.51±0.12	31.7	29.5	30.2	30.6±0.15
OSUDA	**no**	**83.62**	**32.15**	**46.88**	**61.94±0.11**	**27.2**	**23.4**	**26.3**	**25.6±0.14**
OSUDA-AC	no	82.74	32.04	46.62	60.75±0.14	27.8	25.5	27.3	26.5±0.16
OSUDA-SE	no	82.45	31.95	46.59	60.78±0.12	27.8	25.3	27.1	26.4±0.14
SEAT [23]	Yes	84.11	32.67	47.11	62.17±0.15	26.4	21.7	23.5	23.8±0.16

[3]An additional class ratio predictor was required to be trained with the source data.

Table 2:

Comparison of whole tumor segmentation for the cross-modality UDA. We used T2-weighted MRI as our source domain, and T1-weighted, FLAIR, and T1ce MRI as the unlabeled target domains.

Method	Source data	Dice Score [%] ↑				Hausdorff Distance [mm] ↓			
		T1	FLAIR	T1CE	Average	T1	FLAIR	T1CE	Average
Source only	no UDA	6.8	54.4	6.7	22.6±0.17	58.7	21.5	60.2	46.8±0.15
CRUDA [1]	Partial[4]	47.2	65.6	49.4	54.1±0.16	22.1	17.5	24.4	21.3±0.10
OSUDA	**no**	**52.7**	**67.6**	**53.2**	**57.8±0.15**	**20.4**	**16.6**	**22.8**	**19.9±0.08**
OSUDA-AC	no	51.6	66.5	52.0	56.7±0.16	21.5	17.8	23.6	21.0±0.12
OSUDA-SE	no	51.1	65.8	52.8	56.6±0.14	21.6	17.3	23.3	20.7±0.10
CycleGAN [31]	Yes	38.1	63.3	42.1	47.8	25.4	17.2	23.2	21.9
SIFA [3]	Yes	51.7	68	58.2	59.3	19.6	16.9	15.01	17.1
DSFN [32]	Yes	57.3	78.9	62.2	66.1	17.5	13.8	15.5	15.6