An Uncertainty-guided Tiered Self-training Framework for Active Source-free Domain Adaptation in Prostate Segmentation

Zihao Luo[1,3], Xiangde Luo[1,2], Zijun Gao[4], and Guotai Wang[1,2]

[1]School of Mechanical and Electrical Engineering, University of Electronic Science and Technology of China, Chengdu, China
[2]Shanghai AI Lab, Shanghai, China
[3]School of Mathematical Sciences, Harbin Engineering University, Harbin, China.
[4]Department of Computer Science and Engineering, The Chinese University of Hong Kong, Sha Tin, Hong Kong.
guotai.wang@uestc.edu.cn

Abstract. Deep learning models have exhibited remarkable efficacy in accurately delineating the prostate for diagnosis and treatment of prostate diseases, but challenges persist in achieving robust generalization across different medical centers. Source-free Domain Adaptation (SFDA) is a promising technique to adapt deep segmentation models to address privacy and security concerns while reducing domain shifts between source and target domains. However, recent literature indicates that the performance of SFDA remains far from satisfactory due to unpredictable domain gaps. Annotating a few target domain samples is acceptable, as it can lead to significant performance improvement with a low annotation cost. Nevertheless, due to extremely limited annotation budgets, careful consideration is needed in selecting samples for annotation. Inspired by this, our goal is to develop Active Source-free Domain Adaptation (ASFDA) for medical image segmentation. Specifically, we propose a novel **U**ncertainty-**g**uided **T**iered **S**elf-**t**raining (UGTST) framework, consisting of efficient active sample selection via entropy-based primary local peak filtering to aggregate global uncertainty and diversity-aware redundancy filter, coupled with a tiered self-learning strategy, achieves stable domain adaptation. Experimental results on cross-center prostate MRI segmentation datasets revealed that our method yielded marked advancements, with a mere 5% annotation, exhibiting an average Dice score enhancement of 9.78% and 7.58% in two target domains compared with state-of-the-art methods, on par with fully supervised learning. Code is available at: https://github.com/HiLab-git/UGTST.

1 Introduction

Automatic and accurate delineation of the prostate plays an important role in assisting the diagnosis and treatment of prostate diseases. Despite that deep

Z. Luo and X. Luo contributed equally to this work.

learning models have achieved remarkable performance on this task [12,15], they often struggle to generalize well when confronted with gaps between training and testing data [18]. To tackle this issue, Domain Adaptation (DA) methods emerge as a promising solution [5]. Unsupervised Domain Adaptation (UDA) has demonstrated considerable efficacy by leveraging knowledge from labeled source domain data to facilitate segmentation on unlabeled target domain [27, 30]. Moreover, given the constraints posed by privacy and security concerns, the unavailability of source domain necessitates extensive exploration of Source-Free Domain Adaptation (SFDA) techniques in medical image segmentation [**?**,6,17,31]. Nonetheless, owing to the unforeseeable domain discrepancies, both UDA and SFDA face challenges in achieving satisfactory results.

Recently, a few works [2, 16] have confirmed that a small amount of labeled images in the target domain can significantly improve the model's generalizability in the Semi-supervised Domain Adaptation (SSDA) scenario. Despite its performance, SSDA still requires a considerable amount of annotations for DA and still needs to access the source domain. In addition, SSDA overlooks the strategic selection of annotated samples and uses random sample selection with a given annotation budget, which may not select the most valuable images for annotation, leading to sub-optimal performance. In this work, we explore using active learning strategies for effectively selecting valuable samples for annotation [4], which is promising to further reduce the annotation cost, leading to active SFDA (ASFDA). Presently, there is widespread exploration of active sample selection methods grounded in uncertainty-guided approaches [10,11], feature space diversity [22], and their amalgamation [13]. However, due to the complex and dense nature of inherent predictions, along with domain gaps leading to unreliable model features or predictions, conventional active learning methods are unsuitable for ASFDA scenarios. Moreover, as active samples are commonly assumed to harbor the most informative and representative data, they ideally should play a dominant role in the training process. However, this aspect has been neglected by current methods [25, 26, 28, 31].

To mitigate the aforementioned limitations, we propose a practical active learning method Uncertainty-guided Tiered Self-training (UGTST), tailored for ASFDA scenarios in medical image segmentation. In contrast to traditional active learning methods, which often require multiple rounds and utilize only annotated active samples, our approach involves just one round of inference by the source model on the target domain and utilizes unlabeled data in adaptation. We proposed a novel entropy-based slice-level uncertainty estimation method termed global aleatoric uncertainty aggregation and incorporated a diversity-aware redundancy filter for the active sample selection. In response to active samples being undervalued, we developed a Tiered Self-training (TST) DA strategy, by obtaining assumed stable sets to cooperate with active sample dominated DA.

The contributions of this work can be summarized as follows: (1)We present a novel and efficient ASFDA framework called UGTST for prostate segmentation tasks, aiming to improve target domain generalizability through efficient annotation efforts manageable in clinical practice. (2)A global uncertainty es-

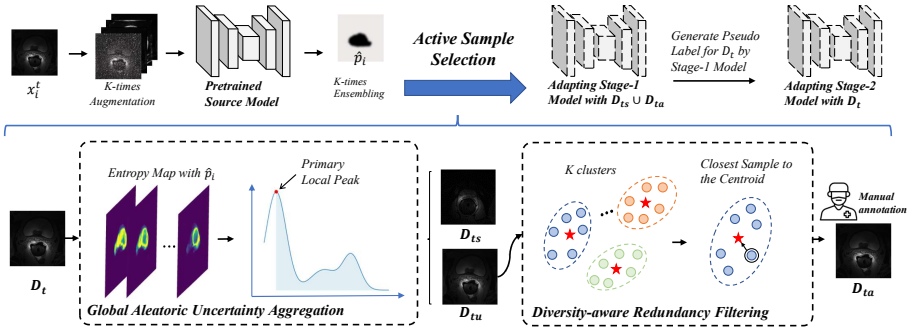

Fig. 1. Overview of our Uncertainty-guided Tiered Self-training Framework, where the D_t, D_{tu}, D_{ts} and D_{ta} are the target domain set, uncertainty candidate set, assumed stable set, and active sample set, respectively. Our method uses the augmentation-based perturbations output for active sample selection via uncertainty and diversity, then employs a tiered self-training strategy for domain adaptation.

timation method for active sample selection in medical image segmentation is designed, along with a diversity-aware redundancy filter to achieve stable and efficient active sample selection. (3)We proposed a practical DA strategy TST for ASFDA, ensuring dominant learning of active samples while progressively utilizing pseudo-labels of unlabeled images. Our method has achieved better performance on the prostate segmentation task than existing ASFDA approaches and was comparable to fully supervised learning with 5% annotation costs.

2 Method

We consider a scenario where a segmentation model trained on a source domain dataset is deployed to a target domain dataset $D_t = \{(x_i^t)\}_{i=1}^{N_t}$, where D_t is unlabeled at the beginning. The objective of ASFDA is, under a controllable small labeling budget M ($M \ll N_t$), to select a labeled subset of samples $D_{at} = \{(x_i^t, y_i^t)\}_{i=1}^{M}$ for one round, and utilize D_t to adjust the pre-trained source model $\mathbf{M}_s$ to achieve good dense predictions on the target domain. Our proposed UGTST is depicted in Fig.1, for the active sample selection stage, given a labeling budget M, we partition the target domain set D_t into an uncertainty candidate set D_{tu} and an assumed stable set D_{ts} based on global aleatoric uncertainty aggregation in entropy map. To ensure the diversity of active samples, we further select the active sample set D_{ta} from the uncertainty candidate set D_{tu} through diversity-aware redundancy filtering. Then, a tiered self-training strategy was employed for adaptation.

2.1 Active Sample Selection via Uncertainty and Diversity

To highlight the most valuable and informative samples, the entropy-based uncertainty estimation method is a common approach in active learning [4]. How-

ever, the source model's limited generalizability leads to highly confident yet unstable predictions on the target domain, making direct computation of entropy maps unreliable. To address this, we adopt a test-time augmentation approach, combining predictions with perturbations from diverse augmentation [9, 24], to diminish confidence in unstable regions and yield more stable predictions [8, 19]. For $x_t^i \in D_t$, we design intensity augmentation $\mathcal{I}$ and spatial augmentation $\mathcal{T}$, with K-times random perturbation, the ensemble segmentation result of x_t^i is:

$$\hat{p}^i = \frac{1}{K}\sum_{k=1}^{K}(\mathcal{T}_k^{-1} \circ \mathbf{M}_S\big(\mathcal{I}_k(\mathcal{T}_k \circ x_t^i))) \tag{1}$$

where $\mathcal{I}_k$, $\mathcal{T}_k$ is k-th random intensity and spatial transformation and $\mathcal{T}_k^{-1}$ is the corresponding inverse spatial transformation. And for $\hat{p}^i \in \mathbb{R}^{C\times H\times W}$ of x_t^i, the entropy map $\boldsymbol{H}(\hat{p}^i) \in \mathbb{R}^{H\times W}$ is calculated as:

$$\boldsymbol{H}(\hat{p}^i) = -\sum_{c=1}^{C}\hat{p}^i(c)\log(\hat{p}^i(c)) \tag{2}$$

Global Aleatoric Uncertainty Aggregation. As mentioned earlier, the entropy map $\boldsymbol{H}(\hat{p}^i)$ cannot be directly used for active sample selection. Due to the imbalance between foreground and background, taking an average of pixel-level uncertainty across the image will be biased to the background. To identify the uncertain region, we design an adaptive threshold to exclude this portion from the output, aiming to aggregate pixels to obtain an unbiased global uncertainty estimation. Hence, we introduce a novel slice-wise uncertainty estimation method called Global Aleatoric Uncertainty Aggregation (GAUA) specifically tailored for medical image segmentation tasks. The discrete density distribution $h^i[n] \in \bar{\boldsymbol{H}}(\hat{p}^i)_{n=1}^{\mathbb{N}}$ is obtained by partitioning the data into bins of size $\mathbb{N} = 100$, arranged from small to large, we can compute the primary local peak value $\boldsymbol{T}_i$ of x_t^i using the discrete difference method, as the self-adaptive threshold to aggregate pixels with relatively high entropy:

$$\boldsymbol{T}_i = \min\{h^i[n] | h^i[n] \in \bar{\boldsymbol{H}}(\hat{p}^i)_{n=1}^{\mathbb{N}}, |\Delta h^i[n]| < \delta, \Delta^2 h^i[n] < 0\} \tag{3}$$

where $\Delta h^i[n]$ is the first-order discrete difference of $h^i[n]$, $\Delta^2 h^i[n]$ is the second-order one. δ is a small adaptive bias for approximation. Then, we compute the mean on pixels with relatively high entropy as the GAUA uncertainty $\boldsymbol{U}_i$ for x_t^i:

$$\boldsymbol{U}_i = \frac{\sum_{n=1}^{N} h^i[n] \cdot \mathbb{I}(h^i[n] > \boldsymbol{T}_i)}{\sum_{n=1}^{N} \mathbb{I}(h^i[n] > \boldsymbol{T}_i)} \tag{4}$$

where $\mathbb{I}$ is the indicator function. Then, we divide D_t into two parts:

$$D_{tu} = \{x_t^i | x_t^i \in D_t, \boldsymbol{U}_i \geq \boldsymbol{U}_i^{N_{tu}}\}; D_{ts} = D_t \setminus D_{tu} \tag{5}$$

where $\boldsymbol{U}_i^{N_{tu}}$ is the N_{tu}-th largest value in $\boldsymbol{U}_i$ corresponding to D_t, the capacity N_{tu} of D_{tu} is a hyper-parameter for balancing uncertainty and diversity.

Diversity-aware Redundancy Filtering. In the uncertainty candidate set D_{tu}, neighboring slices often have similarly high uncertainties. Labeling them would inevitably introduce redundancy, leading to wasted annotation. To deal with this, we take the feature representation $\bar{f}_{x_t^i}$ of slice x_t^i from the encoder of $\mathbf{M}_S$, and we use K-means++ [1] to cluster D_{tu} into M clusters, which M is the annotation budget, and select the samples closest to the cluster centroids:

$$D_{ta} = \{\arg \min_{x_i^{tu} \in D_{tu}} ||\bar{f}_{x_i^{tu}} - C_k||^2; k = 1, 2, ..., M\} \tag{6}$$

where C_k is the centroid of the k-th cluster. $||\cdot||^2$ is the Euclidean distance. $\bar{f}_{x_i^{tu}}$ is the feature representation of x_i^{tu}. Then, annotators are requested to provide manual annotations for selected samples, leading to an annotated subset $D_{ta} = \{(x_i^{ta}, y_i^{ta})\}_{i=1}^{M}$.

2.2 Tiered Self-training for Adaptation

To mitigate the impact of noisy pseudo-labels on active sample learning and make active samples dominant in training, we propose a Tiered Self-training(TST) strategy. We first train a stage-1 model $\mathbf{M}_{t1}$ initialized with parameters from $\mathbf{M}_S$ on $D_{ta} \cup D_{ts}$, where D_{ta} with labeled samples, D_{ts} with pseudo labels. Then, using the trained $\mathbf{M}_{t1}$, we regenerate pseudo-labels for the unlabeled subset of target domain dataset $D_t \setminus D_{ta}$ with the same strategy in Eq.1. Subsequently, we train a stage-2 model $\mathbf{M}_{t2}$ on D_t, progressively achieving domain adaptation across samples with varying degrees of stability. The average of Dice loss and Cross-Entropy loss is used for self-training.

3 Experiment and Results

3.1 Experimental Details

Dataset. To demonstrate the effectiveness of our UGTST method, we employ publicly available prostate T2-weighted MRI images from various clinical centers to evaluate cross-center DA. We select 60 MRI samples comprising a total of 1544 slices from the NCI-ISBI 2013 dataset [3] as the source domain. Additionally, we choose a total of 512 slices from 12 MRI samples acquired from Beth Israel Deaconess Medical Center (BIDMC) and a total of 288 slices from 12 MRI samples obtained from Haokland University Hospital (HK) as two target domains from the PROMISE 12 dataset [14]. In the preprocessing stage, we resized all samples to 384×384 in the axial plane and applied min-max normalization to the volume, following previous studies [15]. Data from each site were divided into four folds at the case level for cross-validation. We only open the labels of the training set in the target domain during the active sample selection stage, simulating the annotation in clinical practice with a labeling budget of 5%.
Implementation Details. We tackled the challenge of large inter-slice spacing by employing slice-by-slice segmentation with 2D CNNs, followed by stacking

Table 1. Quantitative comparison of different domain adaptation methods on prostate segmentation. The best results are in bold, and the second-best are underlined. * indicates p-value < 0.01, and † (p-value < 0.05) (paired t-test) compared to the second-best.

Task	Method	Target Domain BIDMC		Target Domain HK	
		DSC(%)↑	HD_{95}(mm)↓	DSC(%)↑	HD_{95}(mm)↓
No DA	Source only	45.08±32.63	44.20±63.30	42.00±30.26	21.90±24.51
No DA	Target-only	80.59±9.27	7.94±8.24	81.21±8.10	4.14±3.07
Fully DA	Fine-tune	84.28±4.29	5.24±1.58	84.83±5.27	2.85±0.74
SFDA	DPL [6]	67.17±16.03	9.26±6.63	64.55±18.04	6.89±4.94
	FSM [31]	72.17±11.21	6.05±2.77	72.83±12.72	4.59±1.32
	UPL [28]	70.21±12.64	6.21±3.74	<u>73.59±11.51</u>	<u>4.36±1.93</u>
ASFDA	Random	65.14±18.23	8.17±5.04	60.97±26.59	11.50±16.42
	CTC [20]	62.14±21.07	10.14±6.87	64.60±23.92	10.02±13.92
	LC [11]	<u>73.68±11.10</u>	<u>5.89±1.94</u>	69.61±13.93	9.11±7.03
	Core-set [22]	70.07±15.30	6.25±2.93	72.12±11.33	5.29±1.49
	SALAD [13]	73.22±11.85	6.02±1.63	71.48±12.52	5.42±2.07
	UGTST(Ours)	**83.46±4.39***	**5.16±1.73**	**81.17±7.65***	**3.37±1.15†**

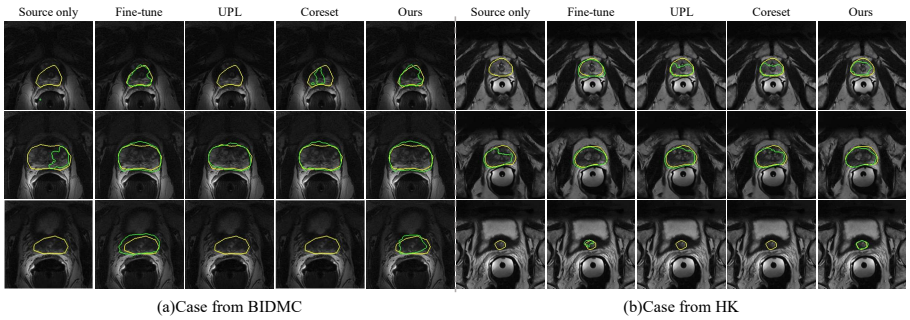

Fig. 2. Qualitative comparison of different Domain Adaptation methods. The ground truth and prediction are displayed in yellow and green contours respectively.

the results into a 3D volume. Our approach utilizes the widely adopted classic 2D U-Net segmentation network [21], with its encoder and decoder serving as the feature extractor and prediction head, respectively. Experiments were conducted using PyTorch on an NVIDIA RTX 2080Ti GPU. For the source model, we trained a segmentation network on annotated source data with a batch size of 24 for 400 epochs, using SGD optimization with an initial learning rate of 0.01 and polynomial decay with a power of 0.9. During the adaptation phase, training was conducted for 100 epochs with a batch size of 24, using the same SGD with an initial learning rate of 0.001. We used data augmentation including random spatial transformations (flips and rotations) and intensity transformations (gamma correction, contrast enhancement, Gaussian blur and noise) during training. Dice Similarity Coefficient (DSC) and 95% Hausdorff Distance (HD_{95}) were used as quantitative evaluation metrics in 3D volumes.

3.2 Comparison with State-of-the-art Methods.

Firstly, we investigated the performance of three state-of-the-art SFDA methods: 1)**DPL** [6], 2)**FSM** [31], and 3)**UPL** [28]. Next, our method was compared with five other sample selection methods for annotation with the same budget: 1)**Random**: randomly select the samples, 2)**CTC** [20]: select the samples closest to the cluster centers, 3)**LC** [11]: samples with smallest probability, 4)**Coreset** [22] samples selected by a set-cover problem and 5)**SALAD** [13]: an ASFDA method employing active learning strategy and guided attention transfer network. These methods were also compared: 1)**Source only**: The pretrained source model, serving as the lower bound. 2)**Target only**: The model was trained solely with annotated images from the target domain. 3)**Fine-tune**: finetuning the source model with full annotations of the target dataset, serving as the upper bound. For a fair comparison, all methods utilized the same backbone architecture [21] with post-processing by retaining the largest connected component in a 3D volume.

The quantitative results based on 4-fold cross-validation of adaptation in two target domains are shown in Table 1. "Source only" and "Target only" achieved an average Dice of 45.08% and 80.59%, respectively in BIDMC domain and 42.00% and 81.21% for HK domain. In observation, SFDA methods demonstrate an enhancement in performance compared to the "Source only", FSM [31] and UPL [28] respectively achieved results of 72.17% and 73.59% average DSC as the best SFDA method. However, there still exists a considerable gap from the upper bound, underscoring the necessity of ASFDA. In the ASFDA with 5% labeled data, Random selection achieved an average DSC of 65.14% and 60.97%, the corresponding values for the best existing method were 73.68% and 72.12%, respectively. Our method achieved DSC of 83.46% and 81.17%, significantly improving performance, and achieved comparable results with an upper bound with "Fine-tune". Fig.2 shows qualitative results between different methods in both two target domains. In the central region where the prostate boundary is prominent, most methods show considerable improvement than "Source only". However, due to the effective integration of uncertainty and diversity, only our approach achieves high-accuracy segmentation of the prostate region in areas where the boundary is less distinct like the apex and base of the prostate.

3.3 Ablation Study.

To further investigate each component's contribution, we conducted ablation and sensitivity study on the first fold. The capacity N_{tu} of the uncertainty candidate set during active sample selection is a hyper-parameter of our method. We set it to M, $2M$, $4M$, and $8M$ to investigate how it affects the performance, where $M = 5\%$, in Fig.3(a). The results from one fold of cross-validation in both domains show that $4M$ is the best hyper-parameter to trade off performance and computational overhead. Further, in Fig.3(b), we also validate the effectiveness of our GAUA compared to other uncertainty estimation methods,

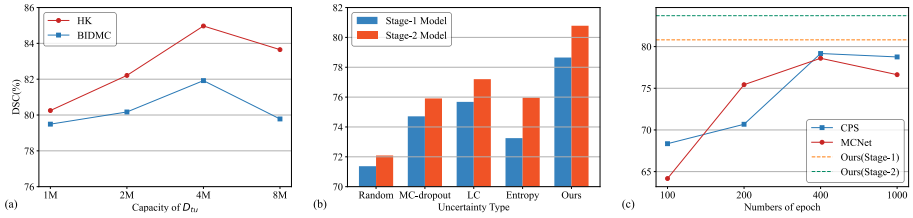

Fig. 3. Ablation study on the validation set. (a). The effect of capacity of D_{tu} with $M = 5\%$, (b). Comparison between different uncertainty estimation methods with two-stage results, and (c). Comparison of semi-supervised learning and our method.

Table 2. Ablation study of the proposed UGTST method. BIDMC was used as the target domain. The baseline was using the pre-trained model's predictions as pseudo labels for adaptation. GAUA: Global Aleatoric Uncertainty Aggregation, DARF: Diversity-aware Redundancy Filtering, TST: Tiered Self-training.

Components				DSC(%)	HD_{95}(mm)
Augmentation	GAUA	DARF	TST		
				49.08±16.77	8.55±3.67
✓				63.21±5.25	6.43±1.74
✓	✓			73.79±2.77	5.23±1.15
✓	✓	✓		77.12±5.78	4.67±1.25
✓	✓	✓	✓	**80.78±4.43**	**4.37±0.97**

including random, MC-dropout [10], Least Confidence(LC) [11] and highest entropy(Entropy) [23] followed typical practice of averaging the uncertainty across all the pixels to obtain image-level uncertainty. Our GAUA has achieved the highest performance, and all the methods' performance have been buffed from TST. To demonstrate the necessity of utilizing the source model in the adaptation stage, we employed a few semi-supervised learning methods [7, 29], using 5% annotated data selected by our active sample selection technique for Semi-supervised Learning (SSL) in the HK domain under different training epochs. The results are presented in Fig.3(c). The performance of both stages of UGTST surpasses the existing SSL methods, demonstrating the priority and efficiency of DA in ASFDA.

Next, we further validated the contribution of each component of our method in the domain BIDMC. The baseline involved using the source model's predictions as pseudo-labels for adaptation. "Augmentation" means using the ensemble prediction as the pseudo-label to apply self-training process without annotation. When not using TST, we directly merge active samples with labels and unlabeled samples with pseudo labels for self-training. Experimental results shown in Table 2 show marked performance improvements by each component of UGTST, further confirming the effectiveness of our approach.

4 Conclusion

This work presented an ASFDA framework for accurate prostate segmentation. In the absence of source domain data, active samples are selected by relying on only one round of predictions from a pre-trained source model on the target domain. We present a novel uncertainty-based active sample selection method in medical image segmentation tasks. It utilizes entropy-based primary local peak filtering to aggregate global uncertainty, along with diversity-aware redundancy filters, thus selecting both informative and representative samples for annotation. Then we designed the tiered self-training DA strategy, stabilizing the active learning while progressively leveraging pseudo labels. Our experimental results show that our method achieves comparable performance to fully supervised training with an annotation budget of 5%, which is manageable in clinical practice.

Acknowledgments. This work was supported by the National Natural Science Foundation of China under grant 62271115.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Arthur, D., Vassilvitskii, S., et al.: k-means++: The advantages of careful seeding. In: Soda. vol. 7, pp. 1027–1035 (2007)
2. Basak, H., Yin, Z.: Semi-supervised domain adaptive medical image segmentation through consistency regularized disentangled contrastive learning. In: MICCAI. pp. 260–270. Springer (2023)
3. Bloch, N., Madabhushi, A., Huisman, H., Freymann, J., Kirby, J., Grauer, M., Enquobahrie, A., Jaffe, C., Clarke, L., Farahani, K.: Nci-isbi 2013 challenge: Automated segmentation of prostate structures. The Cancer Imaging Archive (2015), http://doi.org/10.7937/K9/TCIA.2015.zF0vlOPv
4. Budd, S., Robinson, E.C., Kainz, B.: A survey on active learning and human-in-the-loop deep learning for medical image analysis. MedIA **71**, 102062 (2021)
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: AAAI. vol. 33, pp. 865–872 (2019)
6. Chen, C., Liu, Q., Jin, Y., Dou, Q., Heng, P.A.: Source-free domain adaptive fundus image segmentation with denoised pseudo-labeling. In: MICCAI. pp. 225–235. Springer (2021)
7. Chen, X., Yuan, Y., Zeng, G., Wang, J.: Semi-supervised semantic segmentation with cross pseudo supervision. In: CVPR. pp. 2613–2622 (2021)
8. Fu, J., Lu, T., Zhang, S., Wang, G.: Um-cam: Uncertainty-weighted multi-resolution class activation maps for weakly-supervised fetal brain segmentation. In: MICCAI. pp. 315–324. Springer (2023)

9. Gaillochet, M., Desrosiers, C., Lombaert, H.: Taal: Test-time augmentation for active learning in medical image segmentation. In: MICCAI Workshop on Data Augmentation, Labelling, and Imperfections. pp. 43–53. Springer (2022)
10. Gal, Y., Islam, R., Ghahramani, Z.: Deep bayesian active learning with image data. In: ICML. pp. 1183–1192. PMLR (2017)
11. He, T., Jin, X., Ding, G., Yi, L., Yan, C.: Towards better uncertainty sampling: Active learning with multiple views for deep convolutional neural network. In: ICME. pp. 1360–1365. IEEE (2019)
12. Jia, H., Song, Y., Huang, H., Cai, W., Xia, Y.: Hd-net: hybrid discriminative network for prostate segmentation in mr images. In: MICCAI. pp. 110–118. Springer (2019)
13. Kothandaraman, D., Shekhar, S., Sancheti, A., Ghuhan, M., Shukla, T., Manocha, D.: Salad: Source-free active label-agnostic domain adaptation for classification, segmentation and detection. In: WACV. pp. 382–391 (2023)
14. Litjens, G., Toth, R., Van De Ven, W., Hoeks, C., Kerkstra, S., Van Ginneken, B., Vincent, G., Guillard, G., Birbeck, N., Zhang, J., et al.: Evaluation of prostate segmentation algorithms for mri: the promise12 challenge. MedIA **18**(2), 359–373 (2014)
15. Liu, Q., Dou, Q., Yu, L., Heng, P.A.: Ms-net: multi-site network for improving prostate segmentation with heterogeneous mri data. TMI **39**(9), 2713–2724 (2020)
16. Liu, X., Xing, F., Shusharina, N., Lim, R., Jay Kuo, C.C., El Fakhri, G., Woo, J.: Act: Semi-supervised domain-adaptive medical image segmentation with asymmetric co-training. In: MICCAI. pp. 66–76. Springer (2022)
17. Liu, X., Xing, F., Yang, C., El Fakhri, G., Woo, J.: Adapting off-the-shelf source segmenter for target medical image segmentation. In: MICCAI. pp. 549–559. Springer (2021)
18. Luo, X., Liao, W., He, Y., Tang, F., Wu, M., Shen, Y., Huang, H., Song, T., Li, K., Zhang, S., et al.: Deep learning-based accurate delineation of primary gross tumor volume of nasopharyngeal carcinoma on heterogeneous magnetic resonance imaging: A large-scale and multi-center study. Radiotherapy and Oncology **180**, 109480 (2023)
19. Luo, X., Wang, G., Liao, W., Chen, J., Song, T., Chen, Y., Zhang, S., Metaxas, D.N., Zhang, S.: Semi-supervised medical image segmentation via uncertainty rectified pyramid consistency. MedIA **80**, 102517 (2022)
20. Nguyen, H.T., Smeulders, A.: Active learning using pre-clustering. In: ICML. p. 79 (2004)
21. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: MICCAI. pp. 234–241. Springer (2015)
22. Sener, O., Savarese, S.: Active learning for convolutional neural networks: A core-set approach. arXiv preprint arXiv:1708.00489 (2017)
23. Wang, D., Shang, Y.: A new active labeling method for deep learning. In: IJCNN. pp. 112–119. IEEE (2014)
24. Wang, G., Li, W., Aertsen, M., Deprest, J., Ourselin, S., Vercauteren, T.: Aleatoric uncertainty estimation with test-time augmentation for medical image segmentation with convolutional neural networks. Neurocomputing **338**, 34–45 (2019)
25. Wang, H., Chen, J., Zhang, S., He, Y., Xu, J., Wu, M., He, J., Liao, W., Luo, X.: Dual-reference source-free active domain adaptation for nasopharyngeal carcinoma tumor segmentation across multiple hospitals. TMI (2024)
26. Wang, H., Luo, X., Chen, W., Tang, Q., Xin, M., Wang, Q., Zhu, L.: Advancing uwf-slo vessel segmentation with source-free active domain adaptation and a novel multi-center dataset. arXiv preprint arXiv:2406.13645 (2024)

27. Wu, J., Guo, D., Wang, G., Yue, Q., Yu, H., Li, K., Zhang, S.: Fpl+: Filtered pseudo label-based unsupervised cross-modality adaptation for 3d medical image segmentation. TMI (2024)
28. Wu, J., Wang, G., Gu, R., Lu, T., Chen, Y., Zhu, W., Vercauteren, T., Ourselin, S., Zhang, S.: Upl-sfda: Uncertainty-aware pseudo label guided source-free domain adaptation for medical image segmentation. TMI **42**(12), 3932–3943 (2023)
29. Wu, Y., Xu, M., Ge, Z., Cai, J., Zhang, L.: Semi-supervised left atrium segmentation with mutual consistency training. In: MICCAI. pp. 297–306. Springer (2021)
30. Xu, X., Chen, Y., Wu, J., Lu, J., Ye, Y., Huang, Y., Dou, X., Li, K., Wang, G., Zhang, S., et al.: A novel one-to-multiple unsupervised domain adaptation framework for abdominal organ segmentation. MedIA **88**, 102873 (2023)
31. Yang, C., Guo, X., Chen, Z., Yuan, Y.: Source free domain adaptation for medical image segmentation with fourier style mining. MedIA **79**, 102457 (2022)