Source-Free Domain Adaptation for Cross-Modality Cardiac Image Segmentation with Contrastive Class Relationship Consistency

Ao Ma[1,2,✉], Qingpeng Zhu[1], Jingjing Li[3], Mads Nielsen[4], and Xu Chen[5,6]

[1] School of Computing and Artificial Intelligence, Southwestern University of Finance and Economics, Chengdu, China
[2] Kash Institute of Electronics and Information Industry, Kashi, China
[3] School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, China
[4] Pioneer Centre for AI, Department of Computer Science, University of Copenhagen, Copenhagen, Denmark
[5] Department of Medicine, University of Cambridge, Cambridge, UK
[6] Digital Environment Research Institute (DERI), Queen Mary University of London, London, UK
aoge1993@hotmail.com

Abstract. This paper investigates source-free domain adaptation for cross-modality cardiac image segmentation. Source-free domain adaptation (SFDA) leverages a pretrained model from source domain knowledge and adapts it using target domain data to predict target image labels. While existing SFDA methods have demonstrated strong performance in various medical segmentation tasks, cross-modality cardiac segmentation remains challenging due to significant domain discrepancies between MRI and CT modalities, hindering effective knowledge transfer. Current SFDA approaches primarily focus on pseudo-label denoising through image-level and feature-level alignment, often overlooking class-level information derived from classifier outputs. This paper proposes a novel framework that constructs two class relationship matrices using predictions from a teacher-student model. These matrices are integrated into a contrastive learning framework through intra-view and inter-view pairs. The teacher-student architecture processes both original samples and their augmented counterparts, enforcing prediction consistency for robust adaptation. Simultaneously, our class-aware contrastive learning enhances discriminative capability for cardiac structures. Experimental results demonstrate that our method outperforms state-of-the-art approaches by significant margins, particularly on the challenging CT $\rightarrow$ MR adaptation task.

Keywords: Domain Adaptation · Source-Free · Contrastive Learning.

1 Introduction

Deep learning has demonstrated remarkable success across various domains, ranging from computer vision [8,15] to natural language processing [18]. How-

ever, its performance heavily relies on the availability of large-scale annotated datasets, which are often costly and time-consuming to acquire. For medical images, despite the absence of annotations, domain shifts [13,2] caused by variations in imaging protocols, scanners, or institutions make label prediction even harder under i.i.d condition. Unsupervised Domain adaptation (UDA) [12,6,3] emerges as a promising solution to address this challenge. By leveraging knowledge learned from a source domain with abundant labeled data, domain adaptation techniques enable the transfer of learned representations to a target domain with limited or no annotations.

Source-free domain adaptation (SFDA) [11,14,4] extends UDA by removing the need for direct access to source domain data during adaptation, relying solely on a pre-trained model. This setting is particularly advantageous in medical image segmentation, where data-sharing restrictions prevent access to raw source data due to privacy concerns. Without source data, the model cannot explicitly correct domain shifts by comparing distributions, making it harder to identify and mitigate discrepancies between the source and target domains. Consequently, SFDA methods must rely on more sophisticated strategies to extract meaningful supervision from the target domain alone, ensuring robust adaptation despite the absence of direct source-target alignment.

A classic SFDA methodology for medical image segmentation is pseudo label denoising [4,20,10,22,17] which has evolved significantly over time. [4] pioneered a creative pseudo-label denoising strategy in their work *DPL*, which inspired a wave of subsequent research in this area. Following this, [20] introduced *U-D4R*, proposing class-dependent thresholds for coarse pseudo-label selection and an uncertainty-rectified label soft self-correction mechanism for further refinement. Later, *CPR* [10] designed a pseudo-label refinement scheme based on context similarities to enhance label quality. [17] proposed *CBMT*, which leverages a teacher-student framework and a well-designed calibration loss to enforce robustness. These works achieve great performance on cross-modality fundus images and abdominal organ images, reaching a level close to human expertise. However, among publicly available datasets, cardiac images are still very difficult to be segmented under source-free setting.

The challenge of cross-modality cardiac image segmentation is caused by domain discrepancy between MRIs and CT images. Visually, the contrast ratio of CT images is lower than MR images, which means the foreground and background are deeply mixed up, making it difficult for CNN models to learn discriminative features especially on CT $\rightarrow$ MR tasks. Mainstream methods for cardiac image segmentation are image reconstruction [3,21] and introducing auxiliary knowledge [1,19,23]. The former one leverages CycleGAN [3], fourier style mining [21] to generate modality-like images for training, which essentially mitigates the difference between MR and CT images. The later one introduces prior knowledge like class-ratio [1], visual prompt [19] and MedSAM [23] for image segmentation. For image reconstruction way, we argue that the generated images lose image information during transformation more or less. Besides, it is still difficult for CNN model to learn discriminative features when transforming

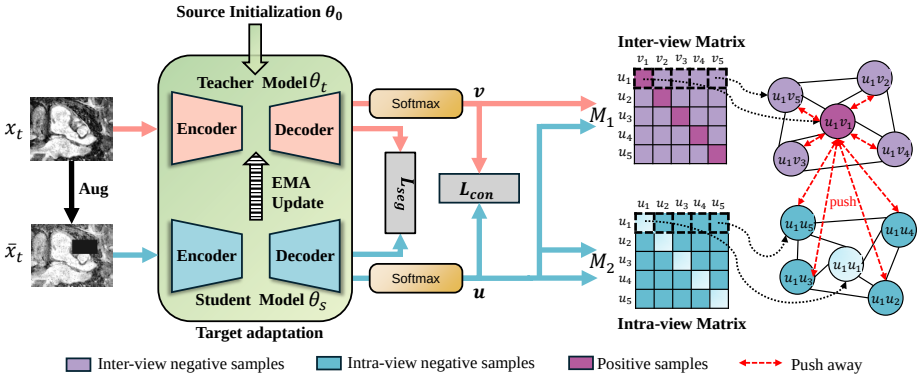


Fig. 1: An overview of our work. Better viewed in color.

MR images into CT-like images. Meanwhile, introducing auxiliary information of target domain [1,23] somehow changes the unsupervised source-free setting into weakly-supervised one, which is hard to accomplish in real-word applications.

Motivated by recent advancements in contrastive learning methods [22,7], we propose a novel contrastive learning framework that incorporates class-aware relationships to enhance the model's discriminative capabilities. We adopt a teacher-student model that utilizes both original and augmented samples as inputs to ensure model stability [17]. We refer to the raw images and their augmented counterparts as view 1 and view 2, respectively. The teacher and student classifiers generate distinct predictions based on these two views. We utilize these predictions to construct two class relationship matrices, selecting both diagonal and non-diagonal elements to form intra-view and inter-view positive and negative pairs embedded within a contrastive learning framework. Unlike existing SFDA methods, our contributions are threefold: First, we introduce a novel contrastive learning framework that emphasizes class relevance without necessitating the calculation of centroids [7] or prototypes [22] based on source classifiers. Second, we fully leverage data augmentation within the teacher-student model, not only to enhance model robustness but also to provide diverse views for constructing positive and negative pairs that improve discriminability. Third, our approach enhances existing denoising schemes by utilizing all soft predictions rather than discarding poor ones, thereby preventing information loss. The code is available at https://github.com/aoge1993/SFDA-CCRC.

2 Method

The source domain is denoted as $\mathcal{D}_s = \{X_s, Y_s\}$, consisting of N_s training samples $\mathbf{x}_s$ and their corresponding labels $\mathbf{y}_s \in \{0,1\}^{H\times W\times C}$, where H, W, and C represent the image height, width, and the number of classes, respectively. In contrast, the target domain, $\mathcal{D}_t = \{X_t, Y_t\}$, comprises N_t unlabeled samples,

with Y_t remaining unknown. Both domains share an identical label space. Under the source-free setting, the labeled source data is exclusively employed to initialize the target model, with subsequent learning relying solely on the target domain. Additionally, we define the augmented target samples as $\bar{\mathbf{x}}_t$, which are obtained through standard data augmentation techniques such as random erasure, contrast modulation, and the introduction of impulse noise. Motivated by famous contrastive learning framework [24], we take original $\mathbf{x}_t$ and augmented $\bar{\mathbf{x}}_t$ as different views of images. Fig.1 presents the overview of our work.

2.1 The Teacher-Student Segmentation with Pseudo Labels

First, the backbone network is trained with labeled source data $\mathcal{D}_s$ with a vanilla cross-entropy loss as follows:

$$\min_{\theta_0} \mathcal{L}_0 = \mathbb{E}_{\mathbf{x},\mathbf{y}\sim\mathcal{D}_s} \sum L_{ce}(\mathbf{y}_s, \hat{\mathbf{y}}_s). \tag{1}$$

In Eq.(1), $\hat{\mathbf{y}}_s$ is the pixel-wise softmax output of the network. θ_0 represents the network parameters. L_{ce} is a standard cross-entropy loss.

After being initialized with θ_0, the student model is further trained with augmented samples and pseudo labels. Specifically, for each target sample $\mathbf{x}_t$ and its corresponding augmented $\bar{\mathbf{x}}_t$, the teacher model takes $\mathbf{x}_t$ as input and obtains softmax prediction $\mathbf{v}$, the student model utilizes $\bar{\mathbf{x}}_t$ and gets $\mathbf{u}$. We set a threshold $\epsilon = 0.75$ to transfer the soft predictions $\mathbf{v}$ into hard labels $\hat{\mathbf{y}}$ with $\hat{\mathbf{y}}_k = \mathbb{I}[\mathbf{v}^k > \epsilon]$ where $\mathbb{I}$ is the indicator function. $\mathbf{v}^k$ and $\hat{\mathbf{y}}_k$ are the k-th dimension of $\mathbf{v}$ and $\hat{\mathbf{y}}$, respectively. For instance, $k \in \{0, 1, 2, 3, 4\}$ for the cardiac dataset (category 0 represents background). Then, we form a binary cross-entropy loss with the hard label $\hat{\mathbf{y}}$ and the soft predictions $\mathbf{u}$ to train the student model as

$$\mathcal{L}_{seg} = \mathbb{E}_{\mathbf{x}\sim\mathcal{D}_t} \sum [\hat{\mathbf{y}} log(\mathbf{u}) + (1 - \hat{\mathbf{y}}) log(1 - \mathbf{u})]. \tag{2}$$

In Eq.(2), a basic teacher-student model has been established. However, hard labels selected by the threshold-based filter is coarse and inaccurate, thus further refinement is necessary.

2.2 Contrastive Learning with Class Relevance

Generally, $\mathbf{u}$ and $\mathbf{v} \in \mathbb{R}^{C\times H\times W}$ are the softmax predictions of the two classifiers (the last FC layer of decoders). For simplicity, we abuse $\mathbf{u}$ and $\mathbf{v}$ as vectors with size $C \times 1$. As shown in Fig 1, we construct two class relevance matrix $\mathbf{M}_1 = \mathbf{u}\mathbf{v}^\top$ and $\mathbf{M}_2 = \mathbf{u}\mathbf{u}^\top$ to describe the class-wise relationship of different views (the original and augmented samples). Then, we define the contrastive learning loss as follows

$$\mathcal{L}_{cl} = -\log \frac{H(\sum_{i=1}^{k} \mathbf{M}_{1,ii})}{\underbrace{H(\sum_{i=1}^{k} \mathbf{M}_{1,ii})}_{\text{the positive pairs}} + \underbrace{H(\sum_{i\neq j} \mathbf{M}_{1,ij})}_{\text{inter-view negative pairs}} + \underbrace{H(\sum_{i\neq j} \mathbf{M}_{2,ij})}_{\text{intra-view negative pairs}}}. \tag{3}$$

In Eq.(3), $H() = exp()/\tau$ where $exp()$ denotes exponential function and $\tau = 0.07$ is the temperature parameter. Eq.(3) consists of three components: the diagonal elements of $\mathbf{M}_1$ as (inter-view) positive pairs, the non-diagonal elements of $\mathbf{M}_1$ as inter-view negative pairs and the non-diagonal elements of $\mathbf{M}_2$ as intra-view negative pairs. The core idea of this contrastive loss is to push away the negative pairs away from the positive pairs. Different from previous contrastive learning works [7,22] focusing on feature-level alignment, we design contrastive loss according to the class relevance of two different views.

Traditional contrastive learning methods emphasize pushing away positive and negative samples at feature space. Further, our work considers class-wise probability, which not only encodes the spatial coordinates of the pixel-wise embeddings but also quantifies their probabilistic distribution within the latent space, which directly corresponds to the confidence level of the semantic attributes allocated to each representation.

2.3 The Teacher and Student Classifiers Consistency

To maintain robustness, we enforce the two classifiers of the teacher and student models to be consistent. The consistency loss is defined as:

$$\mathcal{L}_{con} = \mathbb{E}_{\mathbf{x}\sim\mathcal{D}_t}\Big(\sum_{i,j=1}^{C} \mathbf{M}_{1,ij} - \sum_{i=1}^{C} \mathbf{M}_{1,ii}\Big). \tag{4}$$

Considering a ideal situation that $\mathbf{u} = \mathbf{v}$ with probability 1 for the predicted class, $\mathcal{L}_{con}$ is equal to 0. Therefore, we minimize $\mathcal{L}_{con}$ in Eq.(4) to enforce the teacher-student consistency. Note that Eq.(3) partly includes terms of Eq.(4), it is still necessary to underline the consistency loss for model stability (details are discussed in ablation study).

Finally, our overall loss during the target adaptation phase is

$$\min_{\theta_s} \mathcal{L}_{seg} + \alpha\mathcal{L}_{cl} + \beta\mathcal{L}_{con}, \tag{5}$$

where $\alpha = 1$, $\beta = 1$ are trade-off parameters. θ_s denotes the parameters of the student model. Accordingly, the parameters of teacher model θ_t are updated by EMA strategy as $\theta_t \leftarrow \lambda\theta_t + (1-\lambda)\theta_s$ with $\lambda = 0.98$.

3 Experiments

3.1 Experimental Settings

Dataset. We use the 2017 Multi-modal Whole Heart Segmentation Challenge (MMWHS) dataset, which comprises images including 20 MRI and 20 CT volumes from multiple clinical sites. Five categories of segmentation labels are for each slice in the cardiac dataset, including left ventricle blood cavity (LV), left

Table 1: Results on **MR** (source) → **CT** (target). The evaluation metric is 3D DSC (%) and ASSD (vox). The best result is marked as bold.

method	source-free	DSC					ASSD				
		AA	LA	LV	MYO	Avg	AA	LA	LV	MYO	Avg
NoDA	-	67.4	66.2	69.2	37.0	60.0	17.1	7.7	14.1	11.0	12.4
Oracle	-	91.9	88.3	91.0	85.8	89.2	3.1	3.4	3.6	2.2	3.0
SIFA [3]	×	81.3	79.5	73.8	61.6	74.1	7.9	6.2	5.5	8.5	7.0
DPL [4]	✓	72.8	77.2	61.5	60.2	67.9	8.1	7.2	7.5	8.8	7.9
CBMT [17]	✓	72.2	77.5	67.2	66.3	70.8	6.6	7.9	7.0	7.1	7.2
FSM [21]	✓	84.9	61.6	77.9	**67.3**	72.9	10.4	10.2	7.8	5.3	8.4
AdaMI [1]	✓	83.1	78.2	74.5	66.8	75.7	5.6	**4.2**	5.7	6.9	**5.6**
FVP [19]	✓	**85.6**	71.9	79.5	64.0	75.3	9.0	9.0	4.4	3.5	6.5
CCG [9]	✓	74.2	74.3	69.1	63.5	70.3	**4.9**	5.4	6.5	8.5	6.3
ours	✓	84.1	**85.6**	**85.2**	61.8	**79.2**	7.2	9.5	**3.0**	**3.3**	5.8

atrium blood cavity (LA), the myocardium of the left ventricle (MYO), ascending aorta (AA) and background. We follow the previous dataset pre-processing method of [3,1]. The domain adaptation tasks are MR → CT and CT → MR.

Implementation details. Following the standard setting of SFDA, we utilize all labeled source data to obtain a pre-trained model θ_0. In line with [22], we employ the DeepLabV3+ [5] with MobileNetV2 [16] backbone as our segmentation network. Both the student and teacher models are initialized with θ_0. The batch size is 16. The output probability threshold ϵ is set to 0.75. The initial learning rate is 5×10^{-6} (MR → CT) and 5×10^{-4} (CT → MR), and the model Exponential Moving Average (EMA) update rate λ is 0.98. All adaptation models are trained with the Adam optimizer. The following experiments are implemented by PyTorch and conducted on an NVIDIA 3090 GPU.

We compare our methods with several SOTAs: 1) SFDA methods: *DPL* [4], *CBMT* [17], *FSM* [21], *AdaMI* [1], *FVP* [19], *CCG* [9]; 2) Classic UDA method for medical imaging: *SIFA* [3]. The lower-bound denoted as *NoDA* is a vanilla segmentation network trained with Eq.(1) without adaptation. For all above-mentioned methods, we report the results in the original paper, otherwise we conduct experiments according to their codes. The upper-bound denoted as *Oracle* is a network with initial θ_0 which is updated according to ground-truth labels. For evaluation, we employed two widely utilized metrics: the 3D Dice score and the Average Symmetric Surface Distance (ASSD).

3.2 Quantitative and Qualitative Discussion

In Table 1, *SIFA* [3] mitigates domain shifts by reconstructing CT-like MR images, achieving favorable results. Pseudo-labeling denoising SFDA methods, including *DPL* [4], *CBMT* [17], and *CCG* [9], are primarily designed to handle fundus images that consist solely of background and foreground. Consequently,

Table 2: Results on **CT** (source) → **MR** (target). The evaluation metric is 3D DSC (%) and ASSD (vox). The best result is marked as bold.

method	source-free	DSC					ASSD				
		AA	LA	LV	MYO	Avg	AA	LA	LV	MYO	Avg
NoDA	-	1.6	14.6	46.9	17.0	20.0	37.6	28.7	10.6	15.3	23.0
Oracle	-	91.2	90.1	91.8	89.4	90.6	2.8	3.2	3.0	2.5	2.9
SIFA [3]	×	65.3	62.3	78.9	47.3	63.4	7.3	7.4	3.8	4.4	5.7
DPL [4]	✓	15.6	59.3	63.0	39.8	44.4	15.6	12.7	10.5	9.7	12.1
CBMT [17]	✓	25.2	77.2	73.2	25.3	50.3	26.2	10.2	5.5	7.6	12.4
FSM [21]	✓	35.4	60.5	62.6	42.3	50.2	11.1	8.4	7.2	9.0	8.9
FVP [19]	✓	38.5	44.8	57.8	**49.1**	47.6	19.0	24.6	18.9	14.6	19.3
CCG [9]	✓	62.9	73.7	45.8	35.7	54.5	**5.2**	5.8	5.5	5.8	5.6
ours	✓	**64.3**	**81.3**	**83.4**	38.6	**66.9**	5.4	**3.4**	**3.0**	**5.1**	**4.2**

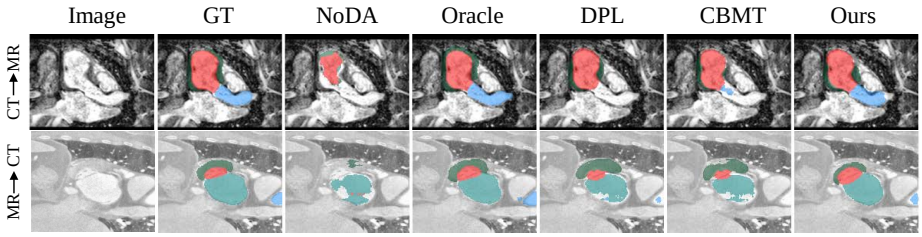


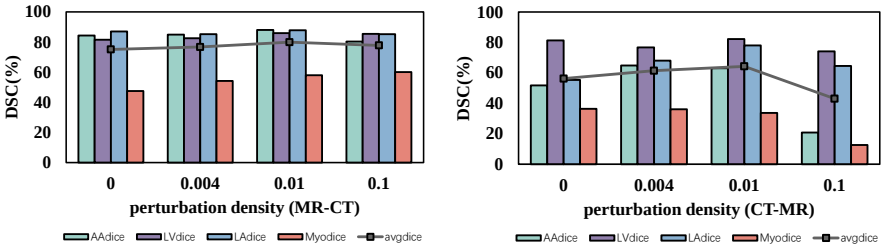
Fig. 2: Visualization on segmentation results. The cardiac structures of AA, LV, LA and MYO are shown in blue, red, cyan and green, respectively.

they struggle to differentiate between various categories of cardiac images. *FVP* [19] employs visual prompts utilizing low-frequency parameters, which may result in the loss of semantic information. *AdaMI* [1] introduces a ground-truth class-ratio prior as supervised information to facilitate target adaptation. However, *AdaMI* neglects to refine network predictions, leading to the potential for poor pseudo-labels that can cause negative transfer.

Table 2 demonstrates that our approach offers significant advantages in the CT → MR task. Previous SFDA studies primarily evaluate their methods on the MR → CT task, as convolutional CNN models can more readily learn domain-invariant features from MR images than from CT images. We contend that the contrast ratio of MR images is greater than that of CT images (as indicated in column 1 of Fig. 2), which facilitates the differentiation between foregrounds and backgrounds. In a source-free setting, the model is initialized with CT images, resulting in poor performance on target MR images during the first epoch. Due to the presence of inaccurate pseudo labels, the model cannot effectively update itself in the correct direction when relying solely on target data in an unsuper-

Table 3: Ablation study on CT → MR.

loss	DSC					ASSD				
	AA	LA	LV	MYO	Avg	AA	LA	LV	MYO	Avg
$\mathcal{L}_0/\mathcal{L}_{seg}$	15.8	54.9	61.1	27.1	39.7	12.7	18.4	5.5	6.6	10.8
$\mathcal{L}_0/\mathcal{L}_{seg} + \mathcal{L}_{cl}$	63.1	78.1	82.3	33.7	64.3	5.0	2.9	3.2	7.1	4.5
$\mathcal{L}_0/\mathcal{L}_{seg} + \mathcal{L}_{cl} + \mathcal{L}_{con}$	64.3	81.3	83.4	38.6	66.9	5.4	3.4	3.0	5.1	4.2



(a) MR→CT (b) CT→MR

Fig. 3: Evaluation on data augmentation of perturbation density. Different colors corresponds to different parts of cardiac images.

vised manner. Our work explores the class relationships of target images, which enables the model to learn discriminative features for these images.

In Fig.2, we can see that our result is the closest to upper-bound *Oracle*, appearing smoother edges than other methods. In Fig.3, we test the influence of data augmentation techniques. Limited by space, we only show the effectiveness of perturbation density. Obviously, a large perturbation density degrades the segmentation results.

3.3 Ablation Study

In Table 3, line 1 demonstrates that the teacher-student model is a good framework which can achieve passable results without other domain adaptation tricks. Then, class-wise contrastive learning loss boosts the basic model with discriminability for better pseudo labeling prediction. Finally, a consistency loss is proposed to regularize the model for robustness. Although $\mathcal{L}_0/\mathcal{L}_{seg} + \mathcal{L}_{cl}$ currently exhibits excellent performance, incorporation of consistency loss $\mathcal{L}_{con}$ can further enhance its effectiveness. Therefore, it is necessary to emphasize the consistency loss despite the fact that $\mathcal{L}_{con}$ is partly contained in $\mathcal{L}_{cl}$.

4 Conclusion

This paper proposes a contrastive SFDA method that considers class relationships with different views for cardiac image segmentation. To enhance the learning of pseudo labels, we construct two matrices to thoroughly explore the class

relevance of the original and augmented images. We establish positive and negative pairs based on the predictions from the teacher and student models, pushing the negative pairs away from the positive ones to improve discriminability. Although our method performs well on cardiac datasets compared to other SOTA approaches, it achieves only technically acceptable results on the MYO part. This indicates that our proposal still has limitations in handling structures with irregular shapes that are surrounded by or surround other structures, which will be the focus of our future improvements.

Acknowledgments. This study was funded by the Ministry of education of Humanities and Social Science project of China (grant No. 24YJCZH203), the Natural Science Foundation of Xinjiang Uyghur Autonomous Region of China (grant No. 2024D01B05), the Sichuan Province Science and Technology Support Program, China (grant No. 2025ZNSFSC1505), the Natural Science Foundation of Xinjiang Uyghur Autonomous Region of China (grant No. 2024D01B07).

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

References

1. Bateson, M., Kervadec, H., Dolz, J., Lombaert, H., Ayed, I.B.: Source-free domain adaptation for image segmentation. Medical Image Analysis **82**, 102617 (2022)
2. Ben-David, S., Blitzer, J., Crammer, K., Kulesza, A., Pereira, F., Vaughan, J.W.: A theory of learning from different domains. Machine learning **79**(1-2), 151–175 (2010)
3. 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 transactions on medical imaging **39**(7), 2494–2505 (2020)
4. 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)
5. Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H.: Encoder-decoder with atrous separable convolution for semantic image segmentation. In: ECCV. pp. 801–818 (2018)
6. Du, Z., Li, X., Li, F., Lu, K., Zhu, L., Li, J.: Domain-agnostic mutual prompting for unsupervised domain adaptation. In: CVPR. pp. 23375–23384 (2024)
7. Gu, M., Thies, M., Mei, S., Wagner, F., Fan, M., Sun, Y., Pan, Z., Vesal, S., Kosti, R., Possart, D., et al.: Unsupervised domain adaptation using soft-labeled contrastive learning with reversed monte carlo method for cardiac image segmentation. In: MICCAI. pp. 681–691. Springer (2024)
8. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: CVPR. pp. 770–778 (2016)
9. Hu, J., Yang, Y., Guo, X., Wang, J., Ma, T.: A chebyshev confidence guided source-free domain adaptation framework for medical image segmentation. JBHI (2024)
10. Huai, Z., Ding, X., Li, Y., Li, X.: Context-aware pseudo-label refinement for source-free domain adaptive fundus image segmentation. In: MICCAI. pp. 618–628. Springer (2023)

11. Li, J., Yu, Z., Du, Z., Zhu, L., Shen, H.T.: A comprehensive survey on source-free domain adaptation. TPAMI (2024)
12. Long, M., Cao, Z., Wang, J., Jordan, M.I.: Conditional adversarial domain adaptation. In: NeurIPS. pp. 1640–1650 (2018)
13. Long, M., Wang, J., Ding, G., Sun, J., Yu, P.S.: Transfer joint matching for unsupervised domain adaptation. In: CVPR. pp. 1410–1417 (2014)
14. Mitsuzumi, Y., Kimura, A., Kashima, H.: Understanding and improving source-free domain adaptation from a theoretical perspective. In: CVPR. pp. 28515–28524 (2024)
15. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: MICCAI. pp. 234–241. Springer (2015)
16. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.C.: Mobilenetv2: Inverted residuals and linear bottlenecks. In: CVPR. pp. 4510–4520 (2018)
17. Tang, L., Li, K., He, C., Zhang, Y., Li, X.: Source-free domain adaptive fundus image segmentation with class-balanced mean teacher. In: MICCAI. pp. 684–694. Springer (2023)
18. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. NeurIPS **30** (2017)
19. Wang, Y., Cheng, J., Chen, Y., Shao, S., Zhu, L., Wu, Z., Liu, T., Zhu, H.: Fvp: Fourier visual prompting for source-free unsupervised domain adaptation of medical image segmentation. IEEE Transactions on Medical Imaging (2023)
20. Xu, Z., Lu, D., Wang, Y., Luo, J., Wei, D., Zheng, Y., Tong, R.K.y.: Denoising for relaxing: unsupervised domain adaptive fundus image segmentation without source data. In: MICCAI. pp. 214–224. Springer (2022)
21. Yang, C., Guo, X., Chen, Z., Yuan, Y.: Source free domain adaptation for medical image segmentation with fourier style mining. Medical Image Analysis **79**, 102457 (2022)
22. Yu, Q., Xi, N., Yuan, J., Zhou, Z., Dang, K., Ding, X.: Source-free domain adaptation for medical image segmentation via prototype-anchored feature alignment and contrastive learning. In: MICCAI. pp. 3–12. Springer (2023)
23. Zhang, G., Qi, X., Yan, B., Wang, G.: Iplc: Iterative pseudo label correction guided by sam for source-free domain adaptation in medical image segmentation. In: MICCAI. pp. 351–360. Springer (2024)
24. Zhu, Y., Xu, Y., Yu, F., Liu, Q., Wu, S., Wang, L.: Deep graph contrastive representation learning. arXiv preprint arXiv:2006.04131 (2020)