Unsupervised Domain Adaptive Fundus Image Segmentation with Category-level Regularization

Wei Feng[1,2,3], Lin Wang[1,2,3], Lie Ju[1,2,3], Xin Zhao[4], Xin Wang[4], Xiaoyu Shi[4], and Zongyuan Ge[1,2,3](✉)

[1] Monash Medical AI Group, Monash University, Melbourne, Australia
[2] Airdoc Monash Research Centre, Monash University, Clayton, Australia
[3] Monash eResearch Center, Monash University, Clayton, Australia
[4] Airdoc LLC, Beijing, China
wf02429@gmail.com, zongyuan.ge@monash.edu
https://www.monash.edu/mmai-group

Abstract. Existing unsupervised domain adaptation methods based on adversarial learning have achieved good performance in several medical imaging tasks. However, these methods focus only on global distribution adaptation and ignore distribution constraints at the category level, which would lead to sub-optimal adaptation performance. This paper presents an unsupervised domain adaptation framework based on category-level regularization that regularizes the category distribution from three perspectives. Specifically, for inter-domain category regularization, an adaptive prototype alignment module is proposed to align feature prototypes of the same category in the source and target domains. In addition, for intra-domain category regularization, we tailored a regularization technique for the source and target domains, respectively. In the source domain, a prototype-guided discriminative loss is proposed to learn more discriminative feature representations by enforcing intra-class compactness and inter-class separability, and as a complement to traditional supervised loss. In the target domain, an augmented consistency category regularization loss is proposed to force the model to produce consistent predictions for augmented/unaugmented target images, which encourages semantically similar regions to be given the same label. Extensive experiments on two publicly fundus datasets show that the proposed approach significantly outperforms other state-of-the-art comparison algorithms[1].

Keywords: Unsupervised domain adaptation · Category level regularization · Fundus image segmentation.

1 Introduction

Recently deep neural networks have dominated several medical image analysis tasks and have achieved good performance [18,10,7]. However, a well-trained

[1] Our code is available at https://github.com/fengweie/UDA_CLR.

model usually underperform when tested directly on an unseen dataset due to domain shift [15]. In clinical practice, this phenomenon is prevalent and remains unresolved. To this end, domain adaptation strategies have received a lot of attention, aiming to transfer knowledge from a label-rich source domain to a label-rare target domain. Recent adversarial training-based domain adaptation methods have shown promising performance, focusing mainly on global distribution adaptation at the input space [23], feature space [4] or output space [16]. Despite the significant performance gains achieved, they all ignore the category distribution constraints. This may result in a situation where although global distribution differences between domains have been reduced, the pixel features of different categories in the target domain are not well separated. This is because some categories are similar to others in terms of appearance and texture. There have been several studies try to address this issue. For example, Liu et al. [11] proposed a prototype alignment loss to reduce the mismatch between the source and target domains in the feature space. Xie et al. [19] proposed a semantic alignment loss to learn semantic feature representations by aligning the category centres of the labelled source domain and the category centres of the pseudo-labelled target domain. However, the shortcoming of these methods is that there is no explicit constraint on the distance between different category features, resulting in categories that look similar in the source domain also being distributed similarly in the target domain, which would potentially lead to incorrect prediction results, especially in edge regions and low-contrast regions.

In this paper, we propose an unsupervised domain adaptation framework based on category-level regularization to accurately segment the optic disc and cup from fundus images. We perform category regularization from both intra-domain and inter-domain perspectives. Specifically, for intra-domain category regularization, on the source domain side, we first propose a prototype-guided discriminative loss to enhance the separability of inter-class distributions and the compactness of intra-class distributions, thus learning more discriminative feature representations; on the target domain side, we propose an augmented consistency-based category regularization loss to constrain the model to produce consistent predictions for perturbed and unperturbed target images, thus encouraging semantically similar regions to have the same labels. For inter-domain category regularization, we propose an adaptive prototype alignment module to ensure that pixels from the same class but different domains can be mapped nearby in the feature space. Experiment results on two public fundus datasets and ablation studies demonstrate the effectiveness of our approach.

2 METHODOLOGY

Fig. 1 shows an overview of our proposed unsupervised domain adaptation framework based on category-level regularization. It consists of three main components, prototype-guided source domain category regularization, augmented consistency-based target domain category regularization, and inter-domain cat-

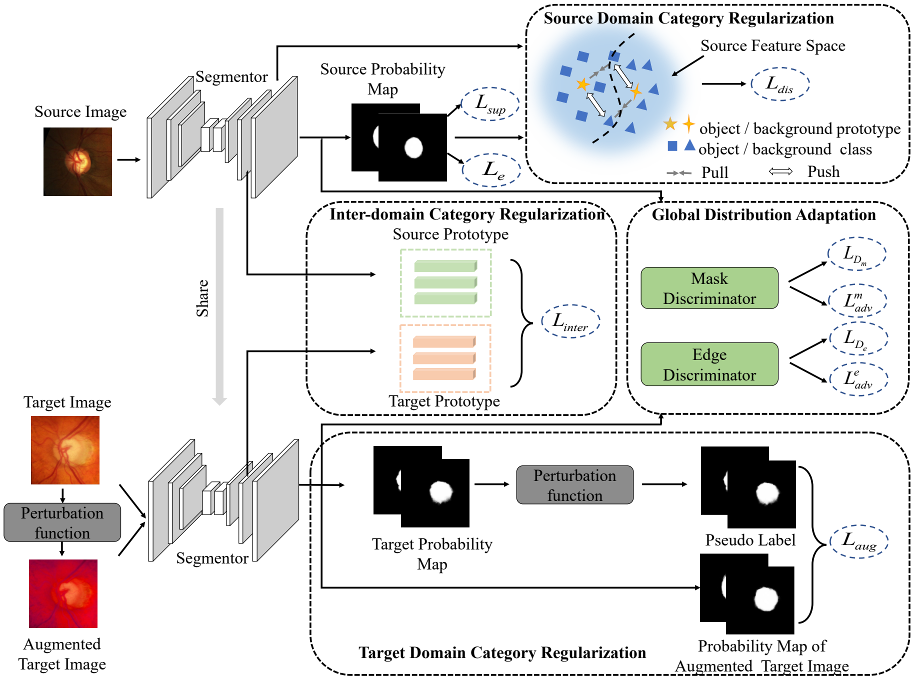

Fig. 1. Overview of an unsupervised domain adaptation framework based on category-level regularization. We first align global distributions between domains by adversarial learning. Then we perform fine-grained level category distribution adaptation from three perspectives: source domain, target domain and inter-domain, via three category regularization methods.

egory regularization, performing category-level regularization from different perspectives.

2.1 Inter-domain Category Regularization

In an typical unsupervised domain adaptation (UDA) setting, we are given a source domain image set $\{x_i^s\}_{i=1}^{N_s}$ and its corresponding pixel-wise annotation $\{y_i^s\}_{i=1}^{N_s}$, and a target domain image set $\{x_i^t\}_{i=1}^{N_t}$ without annotation. To regularize the category distributions between domains, we propose an adaptive prototype alignment module that aligns prototypes of pixels of the same category in the labelled source domain and the pseudo-labelled target domain, thus guaranteeing that features of the same category in different domains are mapped nearby.

Specifically, we feed the target images x^t into the segmentation model G to obtain the pseudo-label $\widehat{y}^t = \mathbb{1}[p^t \geqslant \beta]$, where p^t is the predicted probability and $\mathbb{1}$ is the indicator function. β is a probability threshold. We denote the feature map before the last convolution layer as h^t. Based on $\widehat{y}^t$ and h^t, we can obtain the object prototype f_{obj}^t of the target images as:

$$f_{obj}^t = \frac{1}{N_{obj}} \sum_k \mathbb{1}\left(\widehat{y}_k^t = 1\right) h_k^t, \tag{1}$$

where k represent the pixel index,N_{obj} is the number of pixels of the object class.

However, due to differences in distribution between domains, pseudo labels of target images may contain some incorrect predictions and these false pseudo labels may affect the prototype computation [11]. Inspired by [20], we introduce an uncertainty-guided noise-aware module to filter out those unreliable pseudo labels. We estimate the uncertainty of the pseudo labels based on the Monte Carlo dropout method [6]. Specifically, we enable dropout and perform M stochastic forward inferences to obtain M predicted outputs $\{p_m^t\}_{m=1}^M$. We are then able to obtain the uncertainty estimate for each pixel based on the standard deviation $S = std(\{p_m^t\}_{m=1}^M)$ of the multiple model predictions. We then filter out the pseudo labels of those unreliable pixels by setting an uncertainty threshold ξ to avoid their impact on the prototype computation. The object prototype f_{obj}^s of the source domain is calculated in a similar way[2]. Then the inter-domain category regularization loss can be formulated as:

$$L_{inter} = D(f_{obj}^s - f_{obj}^t), \tag{2}$$

where D is the distance function, we use the Euclidean distance. Note that we only align the prototypes of object class between domains, as object class has more domain shared features than background class [22].

2.2 Intra-domain Category Regularization

In order to further regularize the distributions of the different categories in the feature space, we perform intra-domain category regularization to learn discriminative feature representations by using the category information within the source and target domains, which also works as a complement of inter-domain category regularization.

Source Domain Category Regularization On the source domain side, we propose a prototype-guided discriminative loss to regularize the category distribution. Specifically, we use the category feature prototype to provide supervised signals, explicitly constraining pixel features to be closer to their corresponding category prototypes, while being farther away from other category prototypes. The prototype-guided discriminative loss is formulated as:

$$\begin{aligned} L_{dis} = & \sum_k \mathbb{1}\left(y_k^s = 1\right) \max\left(\left\|h_k^s - f_{obj}^s\right\| - \left\|h_k^s - f_{bg}^s\right\| + \delta, 0\right) \\ & + \sum_k \mathbb{1}\left(y_k^s = 0\right) \max\left(\left\|h_k^s - f_{bg}^s\right\| - \left\|h_k^s - f_{obj}^s\right\| + \delta, 0\right), \end{aligned} \tag{3}$$

[2] Note that since source domain annotation information is available, we use ground truth labels to compute the source domain prototypes.

where δ is a predefined distance margin. $f_{bg}^{s} = \frac{1}{N_{bg}} \sum_{k} \mathbb{1}\left(y_{k}^{s}=0\right) h_{k}^{s}$ is the prototype of the background class. h^{s} is the pixel-wise deep feature of the source domain images. This loss would be 0 when the distance between each pixel feature and its corresponding prototype is less than its distance from other classes of prototypes by a margin δ.

Target Domain Category Regularization In the target domain, since we do not have access to the ground truth labels and therefore the discriminative loss can not be applied as the same way as in the source domain. To perform category-level feature regularization, inspired by the dominant consistency training strategy in semi-supervised learning [17], we propose an augmented consistency-based regularization method that constrains the predictions of the augmented target images to be consistent with the pseudo labels of the original target images, which encourages semantically similar parts of the target images to have the same labels and thus regularize the category-level feature distributions.

Specifically, we apply a perturbation function to $(x^{t}, \widehat{y}^{t})$ to generate a perturbed pair $(x^{pert}, \hat{y}^{pert})$.The augmented consistency loss can be formulated as:

$$L_{aug} = -\sum_{k} \mathbb{1}\left(S_{k} < \mu\right) \ell\left(G(x_{k}^{pert}), \hat{y}_{k}^{pert}\right), \tag{4}$$

where $\ell(\cdot)$ is the cross-entropy loss. Note that here we only calculate the augmented consistency loss for those pseudo-labelled pixels for which the uncertainty estimate S_{k} is less than a threshold μ to avoid error accumulation due to incorrect pseudo labels [14].

2.3 Training Procedure

In addition to category-level regularization, we also perform global distribution alignment by adversarial learning. Following [15], we build two discriminators, D_{m} and D_{e}, to align the predicted probability distribution (p_{m}^{s}, p_{m}^{t}) and the edge structure distribution (p_{e}^{s}, p_{e}^{t}) of the source and target domains respectively. At the same time the training goal of the segmentation model is to learn domain invariant features to deceive the discriminators. In summary, the training objective of the segmentation network can be formulated as:

$$\begin{aligned} L_{total} &= L_{sup} + L_{e} + \lambda_{1} L_{d} + \lambda_{2} L_{inter} + \lambda_{3} L_{dis} + \lambda_{4} L_{aug} \\ L_{e} &= \sum_{k} \left(y_{e,k}^{s} - p_{e,k}^{s}\right)^{2} \\ L_{d} &= L_{adv}^{m} + L_{adv}^{e} = \frac{1}{N_{t}} \sum_{i=1}^{N_{t}} L_{D}\left(p_{m,i}^{t}, 1\right) + \frac{1}{N_{t}} \sum_{i=1}^{N_{t}} L_{D}\left(p_{e,i}^{t}, 1\right), \end{aligned} \tag{5}$$

where L_{sup} is the supervised loss on the labelled source domain image. L_{e} and L_{d} are the edge regression loss and the adversarial loss respectively, L_{D} is the binary cross-entropy loss. y_e^s is the edge ground truth labels. $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ are balance coefficients.

The training objectives of the two discriminators are:

$$\begin{aligned} L_{D_m} &= \frac{1}{N_s} \sum_{i=1}^{N_s} L_D\left(p_{m,i}^s, 1\right) + \frac{1}{N_t} \sum_{i=1}^{N_t} L_D\left(p_{m,i}^t, 0\right) \\ L_{D_e} &= \frac{1}{N_s} \sum_{i=1}^{N_s} L_D\left(p_{e,i}^s, 1\right) + \frac{1}{N_t} \sum_{i=1}^{N_t} L_D\left(p_{e,i}^t, 0\right), \end{aligned} \tag{6}$$

where L_{D_m} and L_{D_e} are the adversarial loss of the mask discriminator and the adversarial loss of the edge discriminator, respectively.

3 Experiments

Dataset and evaluation metric. In order to evaluate the proposed method, we use three datasets: the training part of the REFUGE challenge[3] [12], RIMONE-r3 [5] and Drishti-GS [13]. Following [15], we choose the REFUGE challenge as the source domain and RIMONE-r3 and Drishti-GS as the target domains, respectively. The training set of the REFUGE challenge contains 400 images with annotations, and the RIMONE-r3 and Drishti-GS contain 99/60 and 50/51 training/testing images respectively. Following [15], we crop a 512x512 optic disc region as input of the model. In addition, we use the commonly used Dice coefficient to evaluate the segmentation performance of the optic disc and cup [15].

Implementation details. We use the MobileNetV2 modified Deeplabv3+ [2] network as the segmentation backbone [15]. The Adam algorithm is used to optimize the segmentation model and SGD algorithm is used to optimize the two discriminators [15]. The learning rate of the segmentation network is set as $1e-3$ and divided by 0.2 every 100 epochs and we train a total of 500 epochs. The learning rate of the two discriminators is set as $2e-5$. The probability threshold β is set as 0.75 [15]. In the uncertainty estimation part, we perform 10 stochastic forward passes, and the uncertainty threshold μ is set as 0.05 [1]. We empirically set the distance margin δ to 0.01 and found that it worked well on different datasets. The loss balance coefficients $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ are set as 0.01, 0.01, 0.01,0.01. For the perturbation function, we use the perturbation function used in [3], which includes: color jittering and gaussian blur. We use the feature map of the previous layer of the last convolutional layer to calculate the prototype. All experiments are performed using the Pytorch framework and 8 RTX 3090 GPUs.

[3] https://refuge.grand-challenge.org/

Table 1. Comparison of different methods on the target domain datasets

Method	RIM-ONE-r3 [5]		Drishti-GS [13]	
	Dice disc	Dice cup	Dice disc	Dice cup
Oracle	0.968	0.856	0.974	0.901
Baseline	0.779	0.744	0.944	0.836
TD-GAN [21]	0.853	0.728	0.924	0.747
Hoffman et al. [8]	0.852	0.755	0.959	0.851
Javanmardi et al. [9]	0.853	0.779	0.961	0.849
OSAL-pixel [16]	0.854	0.778	0.962	0.851
pOSAL [16]	0.865	0.787	0.965	0.858
BEAL [15]	0.898	0.810	0.961	0.862
Ours	**0.905**	**0.841**	**0.966**	**0.892**

Comparison with state-of-the-art methods. We compare the proposed method with Baseline method (without adaptation), fully supervised methods (Oracle), and several state-of-the-art unsupervised domain adaptation algorithms, including TD-GAN [21], high-level alignment [8], output space adaptation [9,16], BEAL [15]. As can be seen from the experimental results in Tabel 1, the proposed method achieves significant performance gains on both datasets, especially for the segmentation of the optic cup. Compared to the best comparison algorithm BEAL, our method achieves 3.1% and 3% Dice improvement for the optic cup segmentation on the RIM-ONE-r3 and Drishti-GS datasets, respectively. Furthermore, the segmentation performance of our method is very close to that of fully supervised performance. This indicates that our method is able to achieve good performance in scenarios with varying degrees of domain shift.

We also show the segmentation results of the different algorithms on the two datasets in Fig. 2. It can be seen that in some regions that are obscured or blurred by blood vessels, the segmentation results of other comparison algorithms are poor, while our method is able to accurately identify the boundaries of the optic cup and optic disc, while being very close to the ground truth labels.

Ablation study. We conduct ablation experiments to investigate the effectiveness of each component of the proposed method. In Tabel 2, +src_reg represents source domain category regularization, +trg_reg denotes target domain category regularization, and +inter_reg represents inter-domain category regularization. As seen in Tabel 2, inter-domain category regularization and both intra-domain regularization techniques lead to performance gains, which justifies the need for performing global distribution regularization and category regularization simultaneously. In addition, from Tabel 3 we can also observe that using uncertainty-guided noise-aware (UGNA) modules to filter out unreliable pseudo-labels can benefit inter-domain category distribution regularization. By combining multiple category regularization techniques, our approach further improves segmentation performance on both datasets.

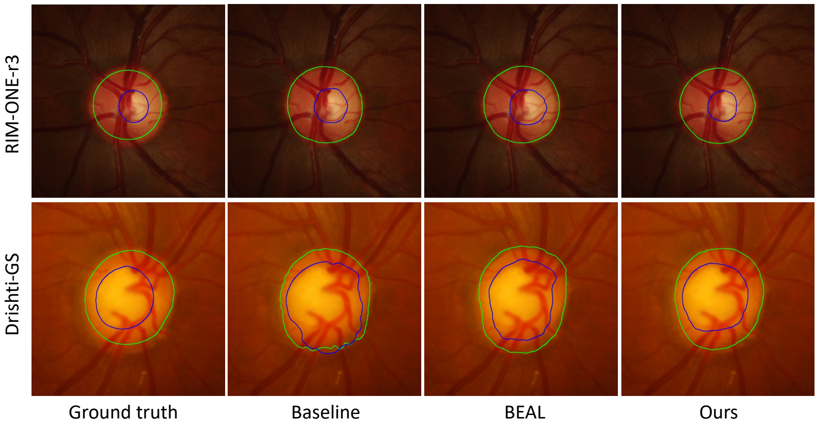

Fig. 2. Quantitative comparison of segmentation results of different methods

Table 2. Ablation study of different components of our method

Method				Target domain			
				RIM-ONE-r3[5]		Drishti-GS[13]	
baseline	+src_reg	+trg_reg	+inter_reg	Dice disc	Dice cup	Dice disc	Dice cup
✓				0.779	0.744	0.944	0.836
✓	✓			0.899	0.829	0.958	0.871
✓		✓		0.898	0.837	0.963	0.875
✓			✓	0.901	0.833	0.961	0.881
✓	✓	✓		0.900	0.839	0.965	0.880
✓	✓		✓	0.903	0.836	0.964	0.883
✓		✓	✓	0.902	0.838	0.963	0.887
✓	✓	✓	✓	**0.905**	**0.841**	**0.966**	**0.892**

Table 3. The impact of UGNA in inter-domain category regularization

Method	RIM-ONE-r3 [5]		Drishti-GS [13]	
	Dice disc	Dice cup	Dice disc	Dice cup
+inter_reg(W/o UGNA)	0.898	0.824	0.959	0.870
+inter_reg	**0.901**	**0.833**	**0.961**	**0.881**

4 Conclusion

In this paper, we propose an unsupervised domain adaptation framework based on category-level regularization for cross-domain fundus image segmentation. Three category regularization methods are developed to simultaneously regularize the category distribution from three perspectives: inter-domain, source and target domains, thus making the model better adapted to the target domain. Our method significantly outperforms state-of-the-art comparison algorithms on two public fundus datasets, demonstrating its effectiveness, and it can be applied to other unsupervised domain adaptation tasks as well.

References

1. Chen, C., Liu, Q., Jin, Y., Dou, Q., Heng, P.A.: Source-free domain adaptive fundus image segmentation with denoised pseudo-labeling. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 225–235. Springer (2021)
2. Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H.: Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proceedings of the European conference on computer vision (ECCV). pp. 801–818 (2018)
3. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A simple framework for contrastive learning of visual representations. In: International conference on machine learning. pp. 1597–1607. PMLR (2020)
4. Dou, Q., Ouyang, C., Chen, C., Chen, H., Heng, P.A.: Unsupervised cross-modality domain adaptation of convnets for biomedical image segmentations with adversarial loss. arXiv preprint arXiv:1804.10916 (2018)
5. Fumero, F., Alayón, S., Sanchez, J.L., Sigut, J., Gonzalez-Hernandez, M.: Rimone: An open retinal image database for optic nerve evaluation. In: 2011 24th international symposium on computer-based medical systems (CBMS). pp. 1–6. IEEE (2011)
6. Gal, Y., Ghahramani, Z.: Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In: international conference on machine learning. pp. 1050–1059. PMLR (2016)
7. Hang, W., Feng, W., Liang, S., Yu, L., Wang, Q., Choi, K.S., Qin, J.: Local and global structure-aware entropy regularized mean teacher model for 3d left atrium segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 562–571. Springer (2020)
8. Hoffman, J., Wang, D., Yu, F., Darrell, T.: Fcns in the wild: Pixel-level adversarial and constraint-based adaptation. arXiv preprint arXiv:1612.02649 (2016)
9. Javanmardi, M., Tasdizen, T.: Domain adaptation for biomedical image segmentation using adversarial training. In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). pp. 554–558. IEEE (2018)
10. Ju, L., Wang, X., Wang, L., Mahapatra, D., Zhao, X., Zhou, Q., Liu, T., Ge, Z.: Improving medical images classification with label noise using dual-uncertainty estimation. IEEE transactions on medical imaging (2022)
11. Liu, Y., Deng, J., Gao, X., Li, W., Duan, L.: Bapa-net: Boundary adaptation and prototype alignment for cross-domain semantic segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 8801–8811 (2021)
12. Orlando, J.I., Fu, H., Breda, J.B., van Keer, K., Bathula, D.R., Diaz-Pinto, A., Fang, R., Heng, P.A., Kim, J., Lee, J., et al.: Refuge challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs. Medical image analysis **59**, 101570 (2020)
13. Sivaswamy, J., Krishnadas, S., Chakravarty, A., Joshi, G., Tabish, A.S., et al.: A comprehensive retinal image dataset for the assessment of glaucoma from the optic nerve head analysis. JSM Biomedical Imaging Data Papers **2**(1), 1004 (2015)

14. Sohn, K., Berthelot, D., Carlini, N., Zhang, Z., Zhang, H., Raffel, C.A., Cubuk, E.D., Kurakin, A., Li, C.L.: Fixmatch: Simplifying semi-supervised learning with consistency and confidence. Advances in Neural Information Processing Systems **33**, 596–608 (2020)
15. Wang, S., Yu, L., Li, K., Yang, X., Fu, C.W., Heng, P.A.: Boundary and entropy-driven adversarial learning for fundus image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 102–110. Springer (2019)
16. Wang, S., Yu, L., Yang, X., Fu, C.W., Heng, P.A.: Patch-based output space adversarial learning for joint optic disc and cup segmentation. IEEE transactions on medical imaging **38**(11), 2485–2495 (2019)
17. Wang, X., Chen, H., Xiang, H., Lin, H., Lin, X., Heng, P.A.: Deep virtual adversarial self-training with consistency regularization for semi-supervised medical image classification. Medical image analysis **70**, 102010 (2021)
18. Wu, Y., Xia, Y., Song, Y., Zhang, Y., Cai, W.: Multiscale network followed network model for retinal vessel segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 119–126. Springer (2018)
19. Xie, S., Zheng, Z., Chen, L., Chen, C.: Learning semantic representations for unsupervised domain adaptation. In: International conference on machine learning. pp. 5423–5432. PMLR (2018)
20. Yu, L., Wang, S., Li, X., Fu, C.W., Heng, P.A.: Uncertainty-aware self-ensembling model for semi-supervised 3d left atrium segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 605–613. Springer (2019)
21. Zhang, Y., Miao, S., Mansi, T., Liao, R.: Task driven generative modeling for unsupervised domain adaptation: Application to x-ray image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 599–607. Springer (2018)
22. Zheng, Y., Huang, D., Liu, S., Wang, Y.: Cross-domain object detection through coarse-to-fine feature adaptation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 13766–13775 (2020)
23. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE international conference on computer vision. pp. 2223–2232 (2017)