



Context-Aware Pseudo-label Refinement for Source-Free Domain Adaptive Fundus Image Segmentation

Zheang Huai, Xinpeng Ding, Yi Li, and Xiaomeng Li^(✉)

The Hong Kong University of Science and Technology, Kowloon, Hong Kong
eexmli@ust.hk

Abstract. In the domain adaptation problem, source data may be unavailable to the target client side due to privacy or intellectual property issues. Source-free unsupervised domain adaptation (SF-UDA) aims at adapting a model trained on the source side to align the target distribution with only the source model and unlabeled target data. The source model usually produces noisy and context-inconsistent pseudo-labels on the target domain, i.e., neighbouring regions that have a similar visual appearance are annotated with different pseudo-labels. This observation motivates us to refine pseudo-labels with context relations. Another observation is that features of the same class tend to form a cluster despite the domain gap, which implies context relations can be readily calculated from feature distances. To this end, we propose a context-aware pseudo-label refinement method for SF-UDA. Specifically, a context-similarity learning module is developed to learn context relations. Next, pseudo-label revision is designed utilizing the learned context relations. Further, we propose calibrating the revised pseudo-labels to compensate for wrong revision caused by inaccurate context relations. Additionally, we adopt a pixel-level and class-level denoising scheme to select reliable pseudo-labels for domain adaptation. Experiments on cross-domain fundus images indicate that our approach yields the state-of-the-art results. Code is available at <https://github.com/xmed-lab/CPR>.

Keywords: Source-free domain adaptation · Context similarity · Pseudo-label refinement · Fundus image

1 Introduction

Accurate segmentation of the optic cup and optic disc in fundus images is essential for the cup-to-disc ratio measurement that is critical for glaucoma screening and detection [6]. Although deep neural networks have achieved great advances in medical image segmentation, they are susceptible to data with

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-43990-2_58.

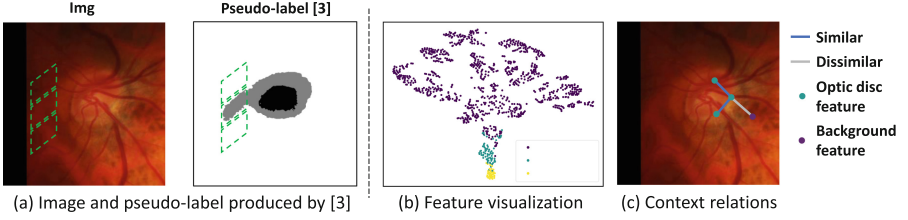


Fig. 1. (a) Example of context-inconsistent pseudo-labels. Due to domain gap, the pseudo-label of optic disc has irregular protuberance which is inconsistent with adjacent predictions. (b) t-SNE visualization of target pixel features produced by source model. Under domain shift, despite not aligning with source segmentor, target features of the same class still form a cluster. (c) Inspired by (b), context relations can be computed from feature distances.

domain shifts, such as those caused by using different scanning devices or different hospitals [24]. Unsupervised domain adaptation [11] is proposed to transfer knowledge to the target domain with access to the source and target data while not requiring any annotation in the target domain. Recently, source-free unsupervised domain adaptation (SF-UDA) has become a significant area of research [5, 10, 14, 15, 19, 20], where source data is inaccessible due to privacy or intellectual property concerns.

Existing SF-UDA solutions can be categorized into four main groups: batch normalization (BN) statistics adaptation [16, 17, 23], approximating source images [9, 26], entropy minimization [2], and pseudo-labeling [3, 25]. BN statistics adaptation methods aim to address the discrepancy of statistics between different domains. For example, [16, 17] update low-order and high-order BN statistics with distinct training objectives, while [23] adapts BN statistics to minimize the entropy of the model’s prediction. Approximating source images aims to generate source-like images. For example, [26] first attains a coarse source image by freezing the source model and training a learnable image, then refines the image via mutual Fourier Transform. The refined source-like image provides a representation of the source data distribution and facilitates domain alignment during the adaptation process. For another instance, [9] learns a domain prompt to add to a target domain image so that the sum simulates the source image. Entropy minimization methods aim to produce more confident model predictions. For example, [2] minimizes output entropy with a regularizer of class-ratio. The class-ratio is estimated by an auxiliary network that is pre-trained on the source domain. For pseudo-labeling [12, 29], erroneous pseudo-labels are either discarded or corrected. For example, [3] identifies low-confidence pseudo-labels at both the pixel-level and the class-level. On the other hand, [25] performs uncertainty-weighted soft label correction by estimating the class-conditional label error probability. However, all of these methods overlook context relations, which can enhance adaptation performance without the need to access the source data.

We observe in our experiments (see Fig. 1 (a)) that domain gaps can result in the source model making context-inconsistent predictions. For neighboring patches of an image with similar visual appearance, the source model can yield vastly

different predictions. This phenomenon can be explained by the observation in [15] that target data shifts in the feature space, causing some data points to shift across the boundary of the source domain segmentor. The issue of context inconsistency motivates us to utilize context relations in refining pseudo-labels. Moreover, it is observed in our experiments (as shown in Fig. 1(b)) that target features produced by the source model still form clusters, meaning that the features of target data points with the same class are closely located. This discovery led us to calculate context relations from feature distances; see Fig. 1(c).

In this paper, we present a novel context-aware pseudo-label refinement (**CPR**) framework for source-free unsupervised domain adaptation. Firstly, we develop a context-similarity learning module, where context relations are computed from distances of features via a context-similarity head. This takes advantage of the intrinsic clustered feature distribution under domain shift [27, 28], where target features generated by the source encoder are close for the same class and faraway for different classes (see Fig. 1 (b)). Secondly, context-aware revision is designed to leverage adjacent pseudo-labels for revising bad pseudo-labels, with aid of the learned context relations. Moreover, a calibration strategy is proposed, aiming to mitigate the negative effect brought about by the inaccurate learned context relations. Finally, the refined pseudo-labels are denoised with consideration of model knowledge and feature distribution [3, 13] to select reliable pseudo-labels for domain adaptation. Experiments on cross-domain fundus image segmentation demonstrate our proposed framework outperforms the state-of-the-art source-free methods [3, 25, 26].

2 Method

Figure 2 illustrates our SF-UDA framework via context-aware pseudo-label refinement. In this section, we first introduce the context-similarity learning scheme. Next, we propose the pseudo-label refinement strategy. Finally, we present the model training with the denoised refined pseudo-labels.

2.1 Context-Similarity Learning

In the SF-UDA problem, a source model $f^s : \mathcal{X}_s \rightarrow \mathcal{Y}_s$ is trained using the data $\{x_s^i, y_s^i\}_{i=1}^{n_s}$ from the source domain $\mathcal{D}_s = (\mathcal{X}_s, \mathcal{Y}_s)$, where $(x_s^i, y_s^i) \in (\mathcal{X}_s, \mathcal{Y}_s)$. f^s is typically trained with a supervision loss of cross-entropy. Also an unlabeled dataset $\{x_t^i\}_{i=1}^{n_t}$ from the target domain \mathcal{D}_t is given, where $x_t^i \in \mathcal{D}_t$. SF-UDA aims to learn a target model $f^t : \mathcal{X}_t \rightarrow \mathcal{Y}_t$ with only the source model f^s and the target dataset $\{x_t^i\}_{i=1}^{n_t}$. In our fundus segmentation problem, $y^i \in \{0, 1\}^{H \times W \times C}$, where C is the number of classes and $C = 2$ because there are two segmentation targets, namely optic cup and optic disc.

Architecture of Context-Similarity Head. Although the target features generated by source encoder do not align with the source segmentor, features of the same classes tend to be in the same cluster while those of different classes are far-away, as shown in Fig. 1 (b). This indicates the source feature encoder is useful for

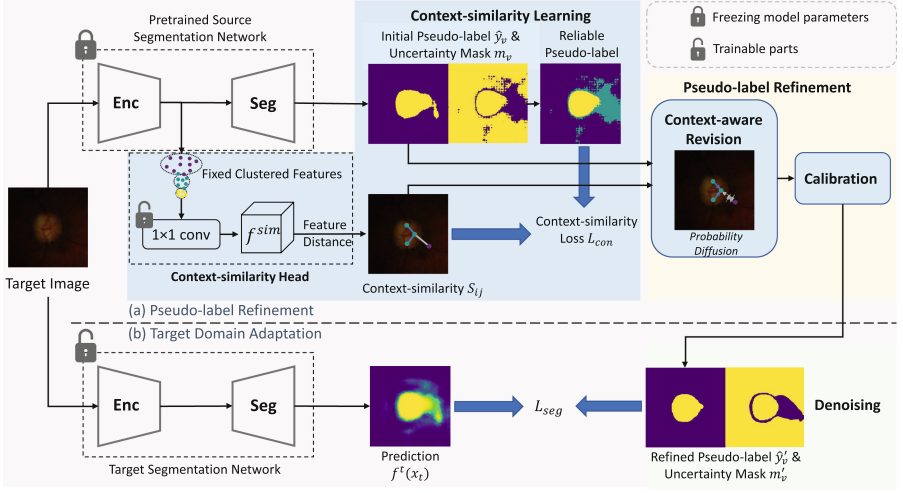


Fig. 2. Overview of the proposed context-aware pseudo-label refinement (CPR) framework for SF-UDA. It consists of two stages: (a) The context-similarity head for computing context relations is trained by reliable pseudo-labels. The learned context similarities are then used to refine the pseudo-labels; (b) Only the refined pseudo-labels with high confidence supervise the training of the segmentation network. The network consists of a feature encoder (Enc) and a segmentor (Seg).

computing context relations. Therefore, we freeze the source encoder and add an additional head to the encoder for learning context semantic relations, motivated by [1]. A side benefit of freezing the source encoder is the training time and required memory can be reduced, as backward propagation is not needed on the encoder. Specifically, the feature map f^{sim} is first obtained, where a 1×1 convolution is applied for adaptation to the target task. Then the semantic similarity between coordinate i and coordinate j on the feature map is defined as

$$S_{ij} = \exp \left\{ -\|f^{sim}(x_i, y_i) - f^{sim}(x_j, y_j)\|_1 \right\}. \quad (1)$$

Computing similarities between every pair of coordinates in a feature map is computationally costly. Thus, for each coordinate i , only similarities with coordinates j lying within the circle of radius r are considered in our implementation.

Training of Context-Similarity Head. Given a target image x_t , initial pseudo-labels and uncertainty mask can be obtained from the source model f^s and x_t , following previous work [3] as:

$$\begin{aligned} p_{v,k} &= f^s(x_t)_v, k = 1, \dots, K, \\ p_v &= \text{avg}(p_{v,1}, \dots, p_{v,K}), u_v = \text{std}(p_{v,1}, \dots, p_{v,K}), \\ \hat{y}_v &= \mathbb{1}[p_v \geq \gamma], \end{aligned} \quad (2)$$

$$z^\omega = \frac{\sum_v f_{l,v} \cdot \mathbb{1}[\hat{y}_v = \omega] \mathbb{1}[u_v < \eta] \cdot p_{v,\omega}}{\sum_v \mathbb{1}[\hat{y}_v = \omega] \mathbb{1}[u_v < \eta] \cdot p_{v,\omega}}, \quad \omega \in \{\text{foreground (fg), background (bg)}\},$$

$$d_v^\omega = \|f_{l,v} - z^\omega\|_2, \quad (3)$$

$$m_v = \mathbb{1}[u_v < \eta] (\mathbb{1}[\hat{y}_v = 1] \mathbb{1}[d_v^{fg} < d_v^{bg}] + \mathbb{1}[\hat{y}_v = 0] \mathbb{1}[d_v^{fg} > d_v^{bg}]).$$

In Eq. 2, Monte Carlo Dropout [8] is performed with K forward passes through the source model, thereby calculating pseudo-label \hat{y}_v and uncertainty u_v for the v -th pixel. Equation 3 first extracts the class-wise prototypes z^ω from the feature map $f_{l,v}$ of the layer before the last convolution, then uncertainty mask m_v is calculated by combining the distance to prototypes and uncertainty u_v . A pseudo-label for the v -th pixel is reliable if $m_v = 1$.

Binary similarity label is then obtained. For two coordinates i and j , similarity label S_{ij}^* is 1 if pseudo-labels $\hat{y}_i = \hat{y}_j$, and 0 otherwise. Note only reliable pseudo-labels are considered to provide less noisy supervision.

The context-similarity head is trained with S^* . To address the class imbalance issue, the loss of each type of similarity (fg-fg, bg-bg, fg-bg) is calculated and aggregated [1] as

$$\mathcal{L}_{con} = -\frac{1}{4} \underset{\substack{\hat{y}_i=\hat{y}_j=1 \\ m_i=m_j=1}}{avg} (\log S_{ij}) - \frac{1}{4} \underset{\substack{\hat{y}_i=\hat{y}_j=0 \\ m_i=m_j=1}}{avg} (\log S_{ij}) - \frac{1}{2} \underset{\substack{\hat{y}_i \neq \hat{y}_j \\ m_i=m_j=1}}{avg} (\log(1-S_{ij})). \quad (4)$$

2.2 Context-Similarity-Based Pseudo-Label Refinement

Context-Aware Revision. The trained context-similarity head is utilized to refine the initial coarse pseudo-labels. Specifically, context-similarities S_{ij} are computed by passing the target image through the source encoder and the trained head. Then the refined probability for the i -th coordinate is updated as the weighted average of the probabilities in a local circle around the i -th coordinate as

$$p_i^{re} = \sum_{d(i,j) \leq r} \frac{S_{ij}^\beta}{\sum_{d(i,j) \leq r} S_{ij}^\beta} \cdot p_j \quad (5)$$

where p_i^{re} is the revised probability and $d(\cdot)$ is the Euclidean distance. $\beta \geq 1$, in order to highlight the prominent similarities and ignore the smaller ones. By combining neighboring predictions based on context relations, revised probabilities are more robust. Equation 5 is performed iteratively for t rounds, since revised probabilities can be used for further revision.

Calibration. The probability update by Eq. 5 might be hurt by inaccurate context relations. We observe that for some classes (optic cup for fundus segmentation) with worse pseudo-labels, the context-similarity for “fg-bg” is not learned well. Consequently, the probability of background incorrectly propagates to that

of foreground, making the probability of foreground lower. To tackle this issue, the revised probability is calibrated as

$$p'_i = \frac{p_i^{re}}{\max_j(p_j^{re})}. \quad (6)$$

The decreased probability is rectified by the maximum value in the image, considering the maximum probability (e.g., in the center of a region) after calibration of a class should be close to 1.

2.3 Model Adaptation with Denoised Pseudo Labels

The refined pseudo-labels can be obtained by $\hat{y}'_v = \mathbb{1}[p'_v \geq \gamma]$. However, noisy pseudo-labels inevitably exist. The combination of model knowledge and target feature distribution shows the best estimation of sample confidence [13]. To this end, reliable pseudo-labels are selected at pixel-level and class-level [3] as

$$\begin{aligned} m'_{v,p} &= \mathbb{1}(p'_v < \gamma_{low} \text{ or } p'_v > \gamma_{high}) \\ m'_{v,c} &= \mathbb{1}(\hat{y}'_v = 1)\mathbb{1}(d_v^{fg} < d_v^{bg}) + \mathbb{1}(\hat{y}'_v = 0)\mathbb{1}(d_v^{fg} > d_v^{bg}), \end{aligned} \quad (7)$$

in which γ_{low} and γ_{high} are two thresholds for filtering out pseudo-labels without confident probabilities. d_v^{fg} and d_v^{bg} are the distances to feature prototypes as computed in Eq. 3. The final label selection mask is the intersection of $m'_{v,p}$ and $m'_{v,c}$, i.e., $m'_v = m'_{v,p} \cdot m'_{v,c}$. The target model is trained under the supervision of pseudo-labels selected by m'_v , with cross-entropy loss:

$$\mathcal{L}_{seg} = - \sum_v m'_v \cdot [\hat{y}'_v \cdot \log(f^t(x_t)_v) + (1 - \hat{y}'_v) \cdot \log(1 - f^t(x_t)_v)]. \quad (8)$$

3 Experiments

Datasets. For a fair comparison, we follow prior work [3] to select three mainstream datasets for fundus image segmentation, i.e., Drishti-GS [22], RIM-ONE-r3 [7], and the validation set of REFUGE challenge [18]. These datasets are split into 50/51, 99/60, and 320/80 for training/testing, respectively.

Implementation Details and Evaluation Metrics. Following prior works [3, 24, 25], our segmentation network is MobileNetV2-adapted [21] DeepLabv3+ [4]. The context-similarity head comprises two branches for optic cup and optic disc, respectively. Each branch includes a 1×1 convolution and a similarity feature map. The threshold γ for determining pseudo-labels is set to 0.75, referring to [24]. The radius r in Eq. 5, the β in Eq. 5 and the iteration number t are set to 4, 2 and 4 respectively. The two thresholds for filtering out unconfident refined pseudo-labels are empirically set as $\gamma_{low} = 0.4$ and $\gamma_{high} = 0.85$, respectively. Each image is pre-processed by clipping a 512×512 optic disc region [24]. The same augmentations as in [3, 25] are applied, including Gaussian noise, contrast adjustment, and

Table 1. Comparison with state-of-the-arts on two settings. “W/o adaptation” refers to directly evaluating the source model on the target dataset. “Upper bound” refers to training the model on the target dataset with labels.

Methods	Dice [%]↑			ASD [pixel]↓		
	Optic cup	Optic disc	Avg	Optic cup	Optic disc	Avg
Source: Drishti-GS; Target: RIM-ONE-r3						
W/o adaptation	70.84	89.94	80.39	13.44	10.76	12.10
Upper bound	83.81	96.61	90.21	6.92	2.96	4.94
DPL [3]	71.70	92.52	82.11	12.49	7.34	9.92
FSM [26]	74.34	91.41	82.88	14.52	10.30	12.41
U-D4R [25]	73.48	93.18	83.33	10.18	6.15	8.16
CPR (ours)	75.02	95.03	85.03	9.84	4.32	7.08
Source: REFUGE; Target: Drishti-GS						
W/o adaptation	79.80	93.89	86.84	13.25	6.70	9.97
Upper bound	89.63	96.80	93.22	6.65	3.55	5.10
DPL [3]	82.04	95.27	88.65	12.14	5.32	8.73
FSM [26]	79.30	94.34	86.82	13.79	5.95	9.87
U-D4R [25]	81.82	95.98	88.90	12.21	4.45	8.33
CPR (ours)	84.49	96.16	90.32	10.19	4.23	7.21

random erasing. The Adam optimizer is adopted with learning rates of $3e-2$ and $3e-4$ in the context-similarity learning stage and the target domain adaptation stage respectively. The momentum of the Adam optimizer is set to 0.9 and 0.99. The batch size is set to 8. The context-similarity head is trained for 16 epochs and the target model is trained for 10 epochs. The implementation is carried out via PyTorch on a single NVIDIA GeForce RTX 3090 GPU. For evaluation, we adopt the widely used Dice coefficient and Average Surface Distance (ASD).

Comparison with State-of-the-Arts. Table 1 shows the comparison of our method with the state-of-the-art SF-UDA methods. Besides three SOTA methods, i.e., DPL [3], FSM [26], and U-D4R [25], we also report the adaptation result without adaptation and the result with fully supervised learning (denoted as “upper bound”). The results show that our approach achieves clear improvements over the previous methods, owing to the proposed pseudo-label refinement scheme which takes advantage of the feature distribution property under domain shift to learn context relations and utilizes valuable context information to rectify pseudo-labels. Figure 3 (a) shows a qualitative comparison.

Ablation Study on Different Modules. Table 2 provides a quantitative analysis to investigate the function of each module. Each component shows its importance in improving the adaptation performance. Particularly, without our pseudo-label refinement, an obvious decrease of segmentation performance can be witnessed, revealing its necessity. Without calibration, the segmentation performance



Fig. 3. On the Drishti-GS to RIM-ONE-r3 adaptation: (a) Qualitative comparison of the optic cup and disc segmentation results with different methods. (b) An example of pseudo-label change with the proposed refinement scheme.

Table 2. Quantitative ablation study on the Drishti-GS to RIM-ONE-r3 adaptation.

Pseudo-label refinement		Denoising	Dice [%]		
Context-aware revision	Calibration		Optic cup	Optic disc	Avg
✗	✗	✗	67.25	93.73	80.49
✓	✗	✗	53.73	92.67	73.20
✓	✗	✓	69.80	94.95	82.38
✓	✓	✗	74.68	93.10	83.89
✗	✗	✓	72.34	94.13	83.23
✓	✓	✓	75.02	95.03	85.03

degrades significantly, which is because the probabilities without calibration do not have correct absolute values. This demonstrates calibration is a necessary step after the revision. Denoising filters out unreliable pseudo-labels by taking into account individual probabilities and feature distribution, thus providing more correct guidance. Integrating all the components completes our framework and yields the best result.

Ablation Study on Pseudo-label Refinement. Ablation study is conducted to verify the effectiveness of the pseudo-label refinement strategy. As shown in Table 3, after refinement, the quality of the pseudo-label is clearly promoted, leading to more accurate supervision for target domain adaptation. For the pseudo-label of optic disc which originally has high accuracy, our refinement scheme encouragingly achieves a boost of 3.5%, showing the robustness of our refinement scheme for different quality of initial pseudo-labels. Without calibration, the accu-

Table 3. Comparison of pseudo-label quality of the training set with different methods on the Drishti-GS to RIM-ONE-r3 adaptation.

Methods	Dice [%]	
	Optic cup	Optic disc
Initial pseudo-label [3]	67.66	90.01
Refined pseudo-label	72.01	93.51
Refined pseudo-label (w/o calibration)	58.34	93.40

racy of the pseudo-label of optic cup is substantially dropped, indicating it is an indispensable part of the overall scheme. Figure 3 (b) visualizes an example of the evolution of the pseudo-label. As can be seen, the context-inconsistent region is clearly improved.

4 Conclusion

This work presents a novel SF-UDA method for the fundus image segmentation problem. We propose to explicitly learn context semantic relations to refine pseudo-labels. Calibration is performed to compensate for the wrong revision caused by inaccurate context relations. The performance is further boosted via the denoising scheme, which provides reliable guidance for adaptation. Our experiments on cross-domain fundus image segmentation show that our method outperforms the state-of-the-art SF-UDA approaches.

Acknowledgement. This work was partially supported by the Hong Kong Innovation and Technology Fund under Project ITS/030/21, as well as by the HKUST-BICI Exploratory Fund (HCIC-004) and Foshan HKUST Projects under Grants FSUST21-HKUST10E and FSUST21-HKUST11E.

References

1. Ahn, J., Kwak, S.: Learning pixel-level semantic affinity with image-level supervision for weakly supervised semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4981–4990 (2018)
2. Bateson, M., Kervadec, H., Dolz, J., Lombaert, H., Ben Ayed, I.: Source-relaxed domain adaptation for image segmentation. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12261, pp. 490–499. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59710-8_48
3. Chen, C., Liu, Q., Jin, Y., Dou, Q., Heng, P.-A.: Source-free domain adaptive fundus image segmentation with denoised pseudo-labeling. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12905, pp. 225–235. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87240-3_22
4. Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H.: Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) ECCV 2018. LNCS, vol. 11211, pp. 833–851. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01234-2_49
5. Ding, N., Xu, Y., Tang, Y., Xu, C., Wang, Y., Tao, D.: Source-free domain adaptation via distribution estimation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 7212–7222 (2022)
6. Fu, H., Cheng, J., Xu, Y., Wong, D.W.K., Liu, J., Cao, X.: Joint optic disc and cup segmentation based on multi-label deep network and polar transformation. *IEEE Trans. Med. Imaging* **37**(7), 1597–1605 (2018)
7. Fumero, F., Alayón, S., Sanchez, J.L., Sigut, J., Gonzalez-Hernandez, M.: Rim-one: 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)

8. 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)
9. Hu, S., Liao, Z., Xia, Y.: ProsFDA: prompt learning based source-free domain adaptation for medical image segmentation. arXiv preprint [arXiv:2211.11514](https://arxiv.org/abs/2211.11514) (2022)
10. Jing, M., Zhen, X., Li, J., Snoek, C.G.M.: Variational model perturbation for source-free domain adaptation. In: Oh, A.H., Agarwal, A., Belgrave, D., Cho, K. (eds.) Advances in Neural Information Processing Systems (2022). https://openreview.net/forum?id=yTJze_xm-u6
11. Kamnitsas, K., et al.: Unsupervised domain adaptation in brain lesion segmentation with adversarial networks. In: Niethammer, M., et al. (eds.) IPMI 2017. LNCS, vol. 10265, pp. 597–609. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-59050-9_47
12. Lee, D.H., et al.: Pseudo-label: the simple and efficient semi-supervised learning method for deep neural networks. In: Workshop on Challenges in Representation Learning, ICML, vol. 3, p. 896. Atlanta (2013)
13. Lee, J., Jung, D., Yim, J., Yoon, S.: Confidence score for source-free unsupervised domain adaptation. In: International Conference on Machine Learning, pp. 12365–12377. PMLR (2022)
14. Li, R., Jiao, Q., Cao, W., Wong, H.S., Wu, S.: Model adaptation: unsupervised domain adaptation without source data. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 9641–9650 (2020)
15. 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)
16. Liu, X., Xing, F., El Fakhri, G., Woo, J.: Memory consistent unsupervised off-the-shelf model adaptation for source-relaxed medical image segmentation. *Med. Image Anal.* **83**, 102641 (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: de Bruijne, Marleen (ed.) MIC-CAI 2021. LNCS, vol. 12902, pp. 549–559. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87196-3_51
18. Orlando, J.I., et al.: Refuge challenge: a unified framework for evaluating automated methods for glaucoma assessment from fundus photographs. *Med. Image Anal.* **59**, 101570 (2020)
19. Roy, S., et al.: Uncertainty-guided source-free domain adaptation. In: Avidan, S., Brostow, G., Cissé, M., Farinella, G.M., Hassner, T. (eds.) ECCV 2022. LNCS, vol. 13685, pp. 537–555. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-19806-9_31
20. Teja, S.P., Fleuret, F.: Uncertainty reduction for model adaptation in semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 9613–9623, June 2021
21. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.C.: Mobilenetv 2: inverted residuals and linear bottlenecks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4510–4520 (2018)
22. Sivaswamy, J., Krishnadass, 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 Biomed. Imaging Data Papers* **2**(1), 1004 (2015)
23. Wang, D., Shelhamer, E., Liu, S., Olshausen, B., Darrell, T.: Tent: fully test-time adaptation by entropy minimization. In: International Conference on Learning Representations (2021). <https://openreview.net/forum?id=uXI3bZLkr3c>

24. 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: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11764, pp. 102–110. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32239-7_12
25. Xu, Z., et al.: Denoising for relaxing: unsupervised domain adaptive fundus image segmentation without source data. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) MICCAI 2022. LNCS, vol. 13435, pp. 214–224. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16443-9_21
26. Yang, C., Guo, X., Chen, Z., Yuan, Y.: Source free domain adaptation for medical image segmentation with Fourier style mining. *Med. Image Anal.* **79**, 102457 (2022)
27. Yang, S., Wang, Y., Wang, K., Jui, S., van de weijer, J.: Attracting and dispersing: a simple approach for source-free domain adaptation. In: Oh, A.H., Agarwal, A., Belgrave, D., Cho, K. (eds.) *Advances in Neural Information Processing Systems* (2022). <https://openreview.net/forum?id=ZlCpRiZN7n>
28. Yang, S., van de Weijer, J., Herranz, L., Jui, S., et al.: Exploiting the intrinsic neighborhood structure for source-free domain adaptation. In: *Advances in Neural Information Processing Systems*, vol. 34, pp. 29393–29405 (2021)
29. Yao, H., Hu, X., Li, X.: Enhancing pseudo label quality for semi-supervised domain-generalized medical image segmentation. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, pp. 3099–3107 (2022)