Unsupervised Adaptation of Polyp Segmentation Models via Coarse-to-Fine Self-Supervision

Jiexiang Wang[1] and Chaoqi Chen[2(✉)]

[1] ByteDance, Beijing, China
wangjiexiang@bytedance.com
[2] The University of Hong Kong, Hong Kong, China
cqchen@gmail.com

Abstract. Unsupervised Domain Adaptation (UDA) has attracted a surge of interest over the past decade but is difficult to be used in real-world applications. Considering the privacy-preservation issues and security concerns, in this work, we study a practical problem of Source-Free Domain Adaptation (SFDA), which eliminates the reliance on annotated source data. Current SFDA methods focus on extracting domain knowledge from the source-trained model but neglects the intrinsic structure of the target domain. Moreover, they typically utilize pseudo labels for self-training in the target domain, but suffer from the notorious error accumulation problem. To address these issues, we propose a new SFDA framework, called Region-to-Pixel Adaptation Network (RPANet), which learns the region-level and pixel-level discriminative representations through coarse-to-fine self-supervision. The proposed RPANet consists of two modules, Foreground-aware Contrastive Learning (FCL) and Confidence-Calibrated Pseudo-Labeling (CCPL), which explicitly address the key challenges of "how to distinguish" and "how to refine". To be specific, FCL introduces a supervised contrastive learning paradigm in the region level to contrast different region centroids across different target images, which efficiently involves all pseudo labels while robust to noisy samples. CCPL designs a novel fusion strategy to reduce the overconfidence problem of pseudo labels by fusing two different target predictions without introducing any additional network modules. Extensive experiments on three cross-domain polyp segmentation tasks reveal that RPANet significantly outperforms state-of-the-art SFDA and UDA methods without access to source data, revealing the potential of SFDA in medical applications.

Keywords: Polyp Segmentation · Source-Free Domain Adaptation · Coarse-to-Fine Self-Supervision

1 Introduction

In clinical applications, deep learning models are typically trained on data collected from a small number of hospitals, but with the objective of being deployed

J. Wang and C. Chen—Contributed equally to this work.

A. Frangi et al. (Eds.): IPMI 2023, LNCS 13939, pp. 250–262, 2023.
https://doi.org/10.1007/978-3-031-34048-2_20

across other hospitals to meet a broader range of needs. However, domain shift [30], such as the variations of patient population and imaging conditions, hinders the deployment of well-trained models to a new hospital. This problem has inspired a body of research on UDA [13,29] by explicitly mitigating the distributional shift between a labeled source domain and an unlabeled target domain [3,5,6,8,12,24,27,28,32,33,36,38]. Despite their general efficacy for various tasks, conventional UDA techniques still have two fundamental shortcomings. First, when encountering a new domain, UDA requires joint training of source and target data, which would be cumbersome if the number of source data is large (*e.g.,* high-quality annotated data from well-known medical institutions). Second, given the privacy-preservation issues and security concerns in medical scenarios, medical data from different clinical centers usually need to be kept locally, *i.e.,* the patient data from a local hospital may not be shared to other hospitals. In addition, in terms of memory overhead, source data mostly have larger size than source-trained models [16], which further imposes a great challenge for real-world applications. This motivates us to investigate a highly realistic and challenging setting called *Source-Free Domain Adaptation* (SFDA) [9,21,23], a.k.a. model adaptation, where only a trained source model and an unlabeled target dataset are available during adaptation.

Compared to mainstream UDA methods, which enable knowledge transfer with concurrent access to both source and target samples, SFDA needs to distill the domain knowledge from the fixed source model and adapt it to the target domain. In this case, it is infeasible to leverage the prevailing feature alignment strategies in UDA, such as adversarial learning [12,34], moment matching [27], and relation-based alignment [7], to achieve adaptation. More importantly, we cannot explicitly measure the domain discrepancy between source and target distributions, making the adaptation process brittle to sophisticated adaptation scenarios with significant distributional shifts. To solve this problem, most of the SFDA methods [1,9,19,21,23,25,26] is comprised of two stages, including source pre-training and target adaptation. For example, Bateson *et al.* [1] develop an entropy minimization term incorporated with a class-ratio prior for preventing trivial solution. Chen *et al.* [9] propose a denoised pseudo-labeling strategy to improve the performance of self-training by collaboratively using uncertainty estimation and prototype estimation to select reliable pseudo labels. Liu *et al.* [25] introduce an adaptive batch-wise normalization statistics adaptation framework, which progressively learns the target domain-specific mean and variance and enforces high-order statistics consistency with an adaptive weighting strategy.

In spite of the fruitful progress and promising results, existing SFDA methods suffer from two key challenges. (1) *How to distinguish:* prior efforts typically learn the discriminative power from the source-trained model but ignores the self-supervised ability within the target domain for distinguishing foregrounds and backgrounds, which is crucial for exploring the intra-domain contextual and semantic structures. (1) *How to refine:* the prevailing pseudo-labeling-based approaches [9,23,26] may be confined by false prediction, error accumulation and

even trivial solution as the pseudo-labels are initialized by source-trained model and gradually refined under the absence of source supervision. In a nutshell, how to simultaneously utilize domain knowledge from the source domain and mine meaningful self-supervision signals from the target domain are crucial to the success of SFDA, but remains out-of-reach for current methods.

Remedying these issues, we propose a new SFDA framework, called Region-to-Pixel Adaptation Network (RPANet), which unifies region-level and pixel-level representation learning in a *coarse-to-fine* manner. The basic idea is to progressively endow the target segmentation models with the capability of distinguishing foregrounds and backgrounds via self-supervision. Specifically, the proposed RPANet consists of two key modules, Foreground-aware Contrastive Learning (FCL) and Confidence-Calibrated Pseudo-Labeling (CCPL), which respectively address the challenges of "how to distinguish" and "how to refine". FCL develops a supervised contrastive learning paradigm to learn region-level discriminative representations by contrasting different region centroids (computed by pseudo labels) across different target images, which is memory-efficient and robust to noisy pseudo labels. CCPL designs a novel fusion strategy to reduce the overconfidence problem of pseudo labels by fusing two different target predictions without introducing any cumbersome learning modules. Extensive experiments on three cross-domain polyp segmentation tasks demonstrate that our RPANet significantly outperforms state-of-the-art SFDA and UDA methods under the absence of source data.

2 Region-to-Pixel Adaptation Network

In this paper, we investigate the problem of SFDA, where we only have access to a pre-trained source model $\boldsymbol{f}_s$ and an unlabeled target dataset $\mathcal{D}_t = \{x_t | x_t \in \mathbb{R}^{H \times W \times 3}\}$. $\boldsymbol{f}_s$ was trained on labeled source data $\mathcal{D}_s$. The source and target data do not follow the IID assumption. The objective of SFDA is to learn a segmentation model that performs well on the target domain. Figure 1 illustrates an overview of RPANet, which consists of two components, namely, Foreground-aware Contrastive Learning (FCL) and Confidence-Calibrated Pseudo-Labeling (CCPL). In particular, FCL and CCPL are complementary to each other, *i.e.,* FCL enhances the robustness of CCPL by providing region-level discriminative representations as an initialization, while CCPL refines the target pseudo labels to mitigate the bias introduced by FCL.

2.1 Foreground-Aware Contrastive Learning

Contrastive learning [14,18] has achieved compelling results in self-supervised representation learning by making the representations of the positive pair close while keep negative pairs apart. In view of the dense-prediction property of segmentation tasks, some of the prior works [4,35] utilize contrastive learning as a pre-training step. They typically conduct this learning process in the image level (*i.e.,* the augmentations of original image are regarded as positive samples and

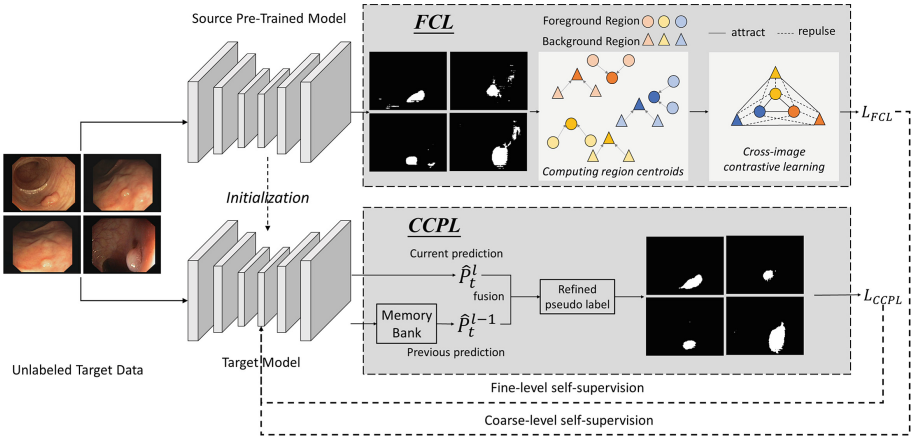


Fig. 1. Overview of our Region-to-Pixel Adaptation Network (RPANet), which includes a pre-trained source model, a target model, and two model adaptation modules (*i.e.,* FCL and CCPL). These two model adaptation modules work in a coarse-to-fine manner to learn discriminative representations in the two levels.

other images from the dataset are regarded as negative samples) while neglecting the holistic context of the entire dataset, *i.e.,* the relations among different real images. The rationale is that these approaches rely on image augmentation techniques to make contrast under the absence of ground-truth annotations. In SFDA, we also have the same dilemma. Instead of following the conventional wisdom, we make use of pseudo labels [20] predicted from the source pre-trained model to perform contrastive learning in a *supervised* way.

Then, an intuitive solution is to utilize pseudo labels to perform pixel-level contrastive learning. However, 1) the pseudo labels may be noisy, and 2) the pixel-level annotations would bring in huge memory overhead for constructing the memory bank. To solve these issues, we propose the FCL module to contrastively learn region-level discriminative representations in a fully supervised way. Here, we use the region centroid to stand for the very region to reduce computational cost. In addition, when computing the centroid, we consider the prediction confidence of different pixels to dynamically assign weights to them.

Formally, the source-trained model $\boldsymbol{f}_s$ takes a target image x_t as input and produces the prediction map $P_t = \boldsymbol{f}_s(x_t)$, and its corresponding feature map is denoted as $F_t \in \mathbb{R}^{H \times W \times K}$. The entropy map I_t is defined as,

$$I_t^{(h,w)} = H(P_t), \tag{1}$$

where $H(\cdot)$ is the entropy function. Then, the region centroid of class k is formulated as,

$$m_k = \frac{\sum_{h,w} F_t \mathbb{1}[P_t^{(h,w)} = k] \cdot (1 - I_t^{(h,w)})}{\sum_{h,w} \mathbb{1}[P_t^{(h,w)} = k]} \tag{2}$$

where $\mathbb{1}(\cdot)$ is an indicator function, and each pixel is re-weighted by its prediction confidence when computing the centroid. After that, the region-level contrastive loss regarding category k can be formulated as,

$$\mathcal{L}_k^{\mathrm{FCL}} = -\frac{1}{|\mathcal{M}_k|} \sum_{m^+ \in M_k} \log \frac{\exp(\boldsymbol{m} \cdot \boldsymbol{m}^+ / \tau)}{\exp(\boldsymbol{m} \cdot \boldsymbol{m}^+ / \tau) + \sum_{\boldsymbol{m}^- \in \mathcal{N}_k} \exp(\boldsymbol{m} \cdot \boldsymbol{m}^- / \tau)} \tag{3}$$

where $\mathcal{M}_k$ and $\mathcal{N}_k$ stand for the collections of the positive and negative samples, for $\boldsymbol{m}$. m^+ and m^- stand for the certain positive and negative samples, respectively. As its core, FCL leverage the "learning to compare" ability of contrastive learning to make a clear distinction between foreground and background regions as well as model the cross-image foreground/background relations. By doing so, regions that are more likely to be the foregrounds are coarsely highlighted.

2.2 Confidence-Calibrated Pseudo-labeling

Pseudo-labeling [20] has been proved to be a simple yet effective approach for SFDA [9,23,26]. Most of state-of-the-art pseudo-labeling methods prioritize their focus on designing sampling strategies to select pseudo labels with high prediction confidence and fit task-specific properties. However, they ignore the structured dependencies of different pixels inside a single image, which is prone to result in overconfident predictions regarding some local regions. More importantly, in some medical applications, where the foreground and background are highly entangled, these methods may be prohibitively difficult to select trustworthy pseudo labels. For example, in colonoscopy images, polyps and normal tissues are visually similar and have low contrast, greatly impeding the correct assignment of pseudo labels. To address this problem, the proposed CCPL calibrates the outputs soft pseudo labels in the pixel level, which integrates the output of previous and current predictions via a simple yet effective fusion strategy.

Given the output of the target segmentation network $\hat{P}_t = \boldsymbol{f}_t(x_t)$ (pre-softmax logits), the target pseudo-label is updated as follows,

$$\hat{P}_t^l = \alpha \cdot \mathrm{softmax}\left(\frac{\hat{P}_t^l}{\phi(\hat{P}_t^l, \hat{P}_t^{l-1})}\right) + (1-\alpha) \cdot \mathrm{softmax}\left(\frac{\hat{P}_t^{l-1}}{\phi(\hat{P}_t^l, \hat{P}_t^{l-1})}\right) \tag{4}$$

where $\hat{P}_t^{l-1}$ is the pre-softmax logits of $(l-1)$-th times, l denotes the iteration times, and α (ranged from 0 to 1) is a modulated factor. In practice, we set $\alpha = 0.5$ in all experiments. The function $\phi(a,b) = \sqrt{\sum_i^{|a|}(a_i^2 + b_i^2)}$ is a normalization term, which could smooth the pseudo label probability distribution to avoid overconfident predictions (dominate pixels).

After obtaining reliable pseudo labels, the segmentation network is optimized in a supervised way by minimizing the cross entropy loss,

$$\mathcal{L}_{\mathrm{CCPL}} = -\sum_{h,w} \sum_{c} \hat{Y}_t^{(h,w,c)} \log(P_t^{(h,w,c)}) \tag{5}$$

where $\hat{Y}_t^{(h,w,c)}$ is the soft pseudo labels.

2.3 Objective Function

First of all, we utilize the weights of source pre-trained model $\boldsymbol{f}_s$ for the initialization of target model. The overall training objective of RPANet, which includes FCL and CCPL, can be formulated as follow,

$$\mathcal{L}_{\text{RPANet}} = \beta \sum_{k} \mathcal{L}_{k}^{\text{FCL}} + \gamma \mathcal{L}_{\text{CCPL}} \tag{6}$$

where β and γ are two trade-off parameters.

3 Experiments

3.1 Dataset

We extensively evaluate the proposed method RPANet on the polyp segmentation tasks with three public datasets and an in-house dataset. (1) ClinicDB [2] contains 612 Standard Definition (SD) frames from 31 sequences. (2) ETIS-LARIB [31] contains 196 High Definition (HD) frames from 34 sequences. (3) Kvasir-SEG [17] contains 1000 polyp frames with various resolutions. (4) In-house dataset is collected from a local hospital and contains 5175 frames with polyp. The annotations are sketched by two experienced gastroenterologists. In experiments, considering the number of annotated data in different datasets, we use in-house dataset as the source domain, and the other three public datasets as the target domain respectively.

3.2 Implementation Details and Evaluation Metrics

We adopt DeepLab-v2 [10] as the segmentation model and ResNet101 [15] pre-trained on ImageNet as the backbone network. In the training phase, we use Stochastic Gradient Descent (SGD) with momentum 0.9 as the optimizer. The learning rate is set as 2.5×10^{-4}, which is decayed by a polynomial annealing policy [10]. The batch size is 4 and we train the models for 20 epochs. Following prior works, standard data augmentation techniques are utilized in experiments. For Eq. (3), we set the temperature τ as 0.1. For Eq. (6), we set $\beta = 1$ and $\gamma = 1$ in all experiments. All experiments are conducted on 4 NVIDIA Tesla V100 GPUs with PyTorch deep learning framework. Following [11], we use 6 metrics to quantitatively evaluate the superiority of our method, including mean Dice, mean IoU, weighted Dice metric F_{β}^{w}, structure similarity measure S_{α}, enhanced-alignment metric E_{ϕ}^{max}, and MAE metric.

3.3 Comparisons with State-of-the-Arts

State-of-the-Arts. We extensively compare the proposed RPANet with state-of-the-art domain adaptive semantic segmentation methods, including Bidirectional Learning **(BDL)** [22], Fourier Domain Adaptation **(FDA)** [37], Historical

Table 1. Quantitative results of different adaptation methods on ClinicDB [2], ETIS-LARIB [31], and Kvasir-SEG [17] datasets.

	Methods	mean Dice	mean IoU	F_β^w	S_α	E_ϕ^{max}	MAE
ClinicDB	w/o adaptation	67.3	55.5	63.9	76.7	82.8	0.047
	BDL [22]	76.8	67.3	73.8	83.3	87.1	0.037
	FDA [37]	78.6	71.1	78.7	85.6	87.6	0.027
	HCL [16]	74.1	64.5	74.5	82.4	85.2	0.034
	DPL [9]	76.3	66.9	75.3	83.7	87.0	0.034
	RPANet w/o FCL	75.7	66.1	74.9	83.2	86.7	0.035
	RPANet w/o CCPL	73.2	62.8	70.6	80.7	86.4	0.041
	RPANet	**80.0**	**71.9**	**80.2**	**86.4**	**89.2**	**0.029**
ETIS-LARIB	w/o adaptation	49.6	41.4	47.2	70.9	70.6	0.026
	BDL [22]	59.3	51.8	57.1	76.3	74.8	0.019
	FDA [37]	61.5	53.5	59.7	77.3	77.0	0.017
	HCL [16]	57.3	49.5	55.4	75.1	73.8	0.024
	DPL [9]	58.8	51.4	56.6	76.1	74.5	0.020
	RPANet w/o FCL	59.1	51.6	56.9	76.3	74.6	0.019
	RPANet w/o CCPL	59.9	52.2	58.1	76.7	75.5	0.017
	RPANet	**63.2**	**55.2**	**61.3**	**78.3**	**77.9**	**0.016**
Kvasir-SEG	w/o adaptation	69.0	57.5	62.4	74.2	79.6	0.109
	BDL [22]	84.2	76.0	82.3	86.7	91.1	0.044
	FDA [37]	84.6	76.5	83.0	87.0	91.4	0.042
	HCL [16]	84.0	75.8	82.4	86.6	91.0	0.043
	DPL [9]	80.4	71.0	79.0	84.1	88.5	0.050
	RPANet w/o FCL	80.5	71.7	76.6	83.5	88.0	0.062
	RPANet w/o CCPL	79.2	69.7	75.3	82.5	87.4	0.063
	RPANet	**85.8**	**78.2**	**84.4**	**87.9**	**92.1**	**0.039**

Contrastive Learning **(HCL)** [16], and Denoised Pseudo-Labeling **(DPL)** [9]. In our polyp segmentation experiments, we use the provided source codes to implement these baseline methods. Note that BDL and FDA are conventional UDA methods, while HCL and DPL are SFDA methods. "W/o adaptation" denotes that the model is trained on the source data and directly applied to the target data without any adaptation.

Table 1 reports the performance comparisons of different adaptation methods on three target domains. The proposed RPANet consistently outperforms all compared methods on all domain adaptation tasks. RPANet significantly promotes the w/o adaptation result from 62.0% to 76.3% on average in terms of mean Dice, which brings about 14.3% improvement. In particular, the performance of HCL and DPL cannot be on par with the conventional UDA methods

Table 2. Further empirical analysis regarding the different design choices of RPANet. Mean Dice is reported on the target domains.

Method	ClinicDB	ETIS-LARIB	Kvasir-SEG
RPANet	80.0	63.2	85.8
Integration with modern SFDA methods			
FCL + HCL [16]	79.3	59.1	85.6
FCL + DPL [9]	78.5	60.1	84.1
CCPL + HCL [16]	78.6	58.8	84.9
CCPL + DPL [9]	77.4	60.3	83.6
Ablation of FCL and CCPL			
FCL w/o entropy regularization	78.9	62.6	85.2
CCPL w/o confidence calibration	78.3	61.8	84.1

(BDL and FDA), revealing the importance and difficulty of adapting polyps segmentation models without access to source data. By contrast, RPANet is almost systematically better than the state-of-the-art UDA methods, highlighting the contributions of our coarse-to-fine self-supervision scheme. To qualitatively illustrate the superiority of our method, we show the examples of polyps segmentation results in Fig. 2 and Fig. 3. We can observe that our RPANet is capable of precisely segmenting polyps, reducing ambiguous predictions, and refining the boundaries between lesions and normal tissues. The justification is that RPANet structurally regularizes the outputs via FCL and CCPL, thereby solving the problem of incomplete and noisy predictions.

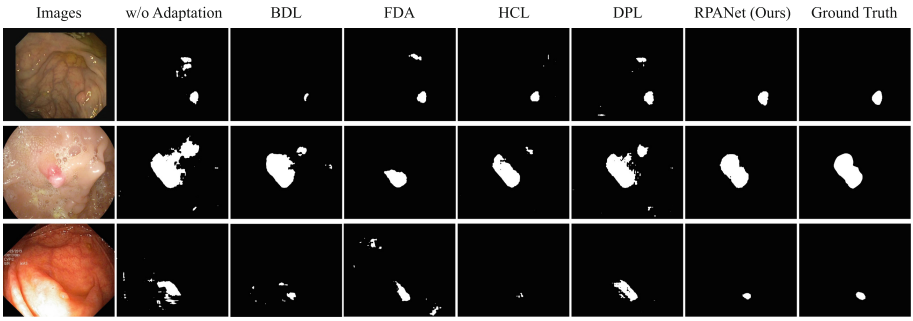


Fig. 2. Qualitative evaluation of different adaptation methods on target domains. From top to bottom: ClinicDB, ETIS-LARIB, Kvasir-SEG.

3.4 Further Empirical Analysis

Ablation Study. To investigate the individual effect of each component (*i.e.,* FCL and CCPL) in our proposed RPANet, we conduct ablation experiments on

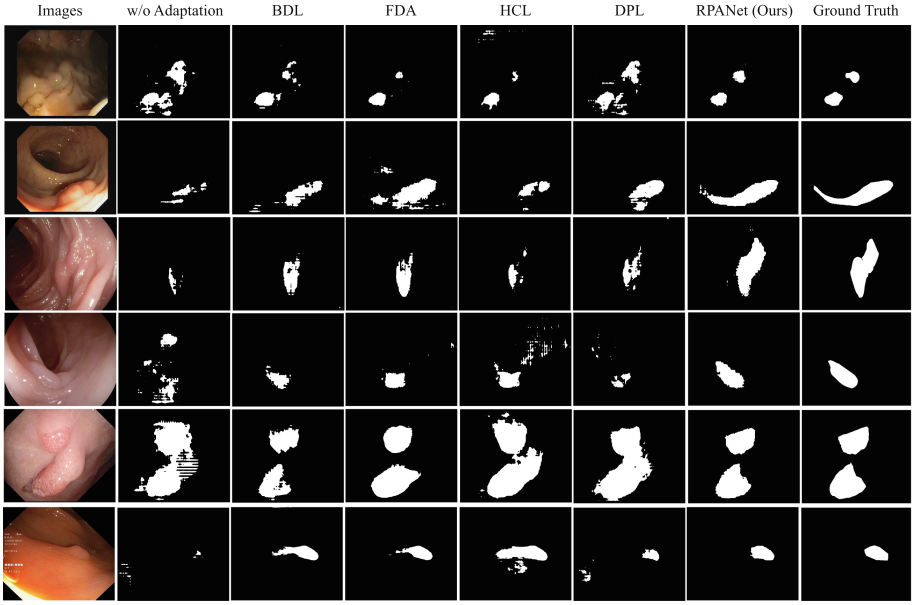


Fig. 3. Qualitative evaluation of different adaptation methods on target domains. From top to bottom: ClinicDB, ETIS-LARIB, Kvasir-SEG.

three domain adaptive polyp segmentation benchmarks. The results are also presented in Table 1. From the table, we can observe that both FCL and CCPL are reasonably designed, and when one of them are removed, the final performance would drop accordingly.

Design Choice. We further explore the alternative design choices for each component to better understand the essence of RPANet. (1) Integration with modern SFDA methods. (2) Ablation of FCL and CCPL. The results are demonstrated in Table 2. We can see that both FCL and CCPL substantially improve state-of-the-art SFDA methods by providing more accurate self-supervision signals. In addition, when the elaborately devised regularization and calibration terms are removed, the performance will be affected to some extent.

Visualization. We visualize the heat map and entropy map on some challenging target images (*i.e.,* the boundaries between the polyp and its surrounding tissues are very ambiguous) as training proceeds in Fig. 4. From the figure, we found that the shapes and boundaries are progressively refined as the training epoch increases, which clearly demonstrates the effectiveness of our coarse-to-fine refinement paradigms. In particular, we can observe that our RPANet achieves competitive results even with very limited training times, such as 8 epochs.

Parameter Sensitivity. We provide the sensitivity analysis regarding hyper-parameters β and γ in Fig. 5. As can be seen, the performance of our RPANet is insensitive to the variations of the value of β and γ on different benchmark datasets, revealing the robustness of the proposed modules. For example, changing β in the range [0.5, 1.5] only incurs small performance variations *i.e.,* 1.6%, 1.1%, and 1.6% in ClinicDB, ETIS-LARIB, and Kvasir-SEG respectively.

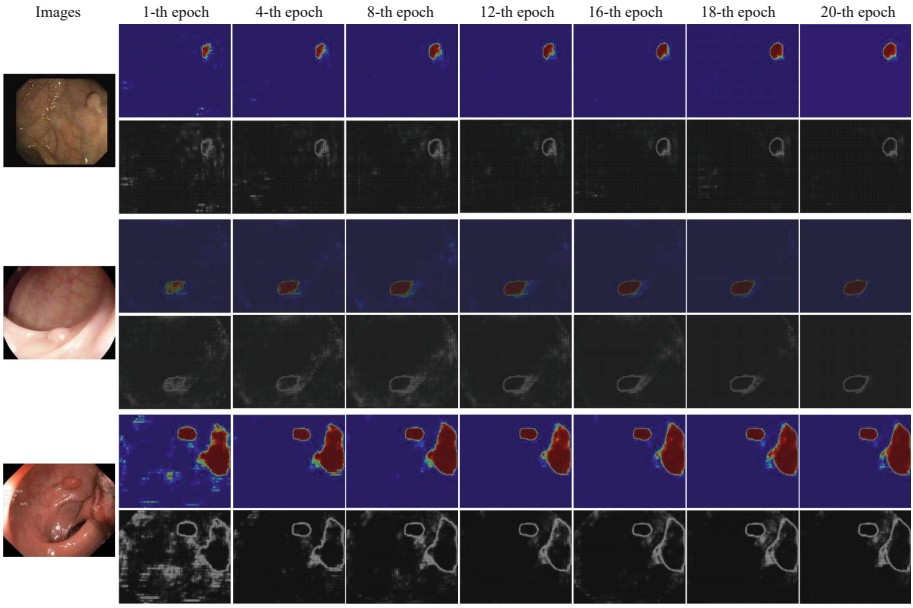
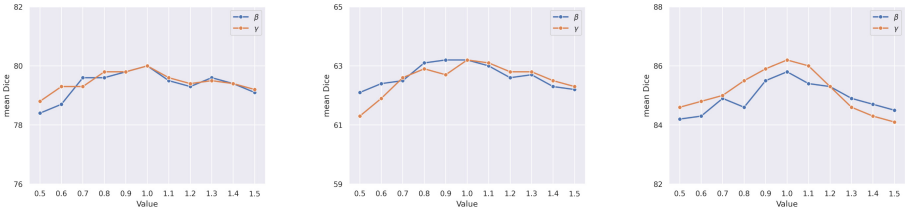


Fig. 4. Visualization results on target images as training proceeds. From top to bottom: ClinicDB, ETIS-LARIB, Kvasir-SEG. In every two rows, **upper:** heat map, **lower:** entropy map. Noting that the brighter the color is, the larger the corresponding value is. (Color figure online)



(a) ClinicDB (b) ETIS-LARIB (c) Kvasir-SEG

Fig. 5. Sensitivity analysis regarding hyper-parameters β and γ.

4 Conclusion

In this paper, we proposed the RPANet to solve the unsupervised adaptation of polyp segmentation models in the absence of source data. The key idea of our method is to endow the target models with the capability of distinguishing foregrounds and backgrounds via self-supervision in a coarse-to-fine manner. The proposed RPANet instantiates this objective with the incorporation of two elaborate modules, *i.e.,* FCL and CCPL. FCL and CCPL learn region-level and pixel-level discriminative representations via supervised contrastive learning and confidence-calibrated pseudo-label refinement, respectively. Experiments on three cross-domain poly segmentation tasks verified the effectiveness of our method, revealing the possibility of SFDA for real-world medical applications.

References

1. 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
2. Bernal, J., Sánchez, F.J., Fernández-Esparrach, G., Gil, D., Rodríguez, C., Vilariño, F.: WM-DOVA maps for accurate polyp highlighting in colonoscopy: validation vs. saliency maps from physicians. Comput. Med. Imaging Graph. **43**, 99–111 (2015)
3. Bian, C., et al.: Uncertainty-aware domain alignment for anatomical structure segmentation. Med. Image Anal. **64**, 101732 (2020)
4. Chaitanya, K., Erdil, E., Karani, N., Konukoglu, E.: Contrastive learning of global and local features for medical image segmentation with limited annotations. In: NeurIPS (2020)
5. Chen, C., Li, J., Han, X., Liu, X., Yu, Y.: Compound domain generalization via meta-knowledge encoding. In: CVPR (2022)
6. Chen, C., et al.: Relation matters: foreground-aware graph-based relational reasoning for domain adaptive object detection. IEEE TPAMI (2022)
7. Chen, C., Wang, J., Pan, J., Bian, C., Zhang, Z.: GraphSKT: graph-guided structured knowledge transfer for domain adaptive lesion detection. IEEE TMI (2022)
8. Chen, C., et al.: Progressive feature alignment for unsupervised domain adaptation. In: CVPR (2019)
9. 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
10. Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L.: DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFS. IEEE TPAMI **40** (2017)
11. Fan, D.-P., et al.: PraNet: parallel reverse attention network for polyp segmentation. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12266, pp. 263–273. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59725-2_26
12. Ganin, Y., et al.: Domain-adversarial training of neural networks. JMLR **17**(1), 1–35 (2016)

13. Guan, H., Liu, M.: Domain adaptation for medical image analysis: a survey. IEEE. Trans. Biomed. Eng. (2021)
14. He, K., Fan, H., Wu, Y., Xie, S., Girshick, R.: Momentum contrast for unsupervised visual representation learning. In: CVPR (2020)
15. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: CVPR (2016)
16. Huang, J., Guan, D., Xiao, A., Lu, S.: Model adaptation: historical contrastive learning for unsupervised domain adaptation without source data. In: NeurIPS (2021)
17. Jha, D., et al.: Kvasir-SEG: a segmented polyp dataset. In: Ro, Y.M., et al. (eds.) MMM 2020. LNCS, vol. 11962, pp. 451–462. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-37734-2_37
18. Khosla, P., et al.: Supervised contrastive learning. In: NeurIPS (2020)
19. Kundu, J.N., Venkat, N., Babu, R.V., et al.: Universal source-free domain adaptation. In: CVPR (2020)
20. 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. In: ICML, vol. 3, p. 896 (2013)
21. Li, R., Jiao, Q., Cao, W., Wong, H.S., Wu, S.: Model adaptation: unsupervised domain adaptation without source data. In: CVPR (2020)
22. Li, Y., Yuan, L., Vasconcelos, N.: Bidirectional learning for domain adaptation of semantic segmentation. In: CVPR, pp. 6936–6945 (2019)
23. Liang, J., Hu, D., Feng, J.: Do we really need to access the source data? Source hypothesis transfer for unsupervised domain adaptation. In: ICML (2020)
24. Liu, J., Guo, X., Yuan, Y.: Prototypical interaction graph for unsupervised domain adaptation in surgical instrument segmentation. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12903, pp. 272–281. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87199-4_26
25. 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, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12902, pp. 549–559. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87196-3_51
26. Liu, Y., Zhang, W., Wang, J.: Source-free domain adaptation for semantic segmentation. In: CVPR (2021)
27. Long, M., Cao, Y., Wang, J., Jordan, M.: Learning transferable features with deep adaptation networks. In: ICML, pp. 97–105 (2015)
28. Ouyang, C., Kamnitsas, K., Biffi, C., Duan, J., Rueckert, D.: Data efficient unsupervised domain adaptation for cross-modality image segmentation. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11765, pp. 669–677. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32245-8_74
29. Pan, S.J., Yang, Q.: A survey on transfer learning. IEEE TKDE **22** (2009)
30. Quiñonero-Candela, J., Sugiyama, M., Schwaighofer, A., Lawrence, N.D.: Dataset Shift in Machine Learning. MIT Press, Cambridge (2008)
31. Silva, J., Histace, A., Romain, O., Dray, X., Granado, B.: Toward embedded detection of polyps in WCE images for early diagnosis of colorectal cancer. Int. J. Comput. Assist. Radiol. Surg. **9**(2), 283–293 (2014)
32. Sun, L., Wang, J., Huang, Y., Ding, X., Greenspan, H., Paisley, J.: An adversarial learning approach to medical image synthesis for lesion detection. IEEE JBHI **24** (2020)
33. Tzeng, E., Hoffman, J., Saenko, K., Darrell, T.: Adversarial discriminative domain adaptation. In: CVPR (2017)

34. Wang, J., Huang, H., Chen, C., Ma, W., Huang, Y., Ding, X.: Multi-sequence cardiac MR segmentation with adversarial domain adaptation network. In: Pop, M., et al. (eds.) STACOM 2019. LNCS, vol. 12009, pp. 254–262. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-39074-7_27
35. Wang, X., Zhang, R., Shen, C., Kong, T., Li, L.: Dense contrastive learning for self-supervised visual pre-training. In: CVPR (2021)
36. Xia, Y., et al.: Uncertainty-aware multi-view co-training for semi-supervised medical image segmentation and domain adaptation. Med. Image Anal. **65**, 101766 (2020)
37. Yang, Y., Soatto, S.: FDA: Fourier domain adaptation for semantic segmentation. In: CVPR (2020)
38. Yu, S., et al.: Cross-domain depth estimation network for 3D vessel reconstruction in OCT angiography. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12908, pp. 13–23. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87237-3_2