Few-Shot Medical Image Segmentation via a Region-Enhanced Prototypical Transformer

Yazhou Zhu[1], Shidong Wang[2], Tong Xin[3], and Haofeng Zhang[1(✉)]

[1] School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China
`{zyz_nj,zhanghf}@njust.edu.cn`
[2] School of Engineering, Newcastle University, Newcastle upon Tyne NE17RU, UK
`shidong.wang@newcastle.ac.uk`
[3] School of Computing, Newcastle University, Newcastle upon Tyne NE17RU, UK
`tong.xin@newcastle.ac.uk`

Abstract. Automated segmentation of large volumes of medical images is often plagued by the limited availability of fully annotated data and the diversity of organ surface properties resulting from the use of different acquisition protocols for different patients. In this paper, we introduce a more promising few-shot learning-based method named Region-enhanced Prototypical Transformer (RPT) to mitigate the effects of large intra-class diversity/bias. First, a subdivision strategy is introduced to produce a collection of regional prototypes from the foreground of the support prototype. Second, a self-selection mechanism is proposed to incorporate into the Bias-alleviated Transformer (BaT) block to suppress or remove interferences present in the query prototype and regional support prototypes. By stacking BaT blocks, the proposed RPT can iteratively optimize the generated regional prototypes and finally produce rectified and more accurate global prototypes for Few-Shot Medical Image Segmentation (FSMS). Extensive experiments are conducted on three publicly available medical image datasets, and the obtained results show consistent improvements compared to state-of-the-art FSMS methods. The source code is available at: https://github.com/YazhouZhu19/RPT.

Keywords: Few-Shot Learning · Medical Image Segmentation · Bias Alleviation · Transformer

1 Introduction

Automatic medical image segmentation is the implementation of data-driven image segmentation concepts to identify a specific anatomical structure's surface or volume in a medical image ranging from X-ray and ultrasonography to CT and MRI scans. Deep learning algorithms are exquisitely suited for this task
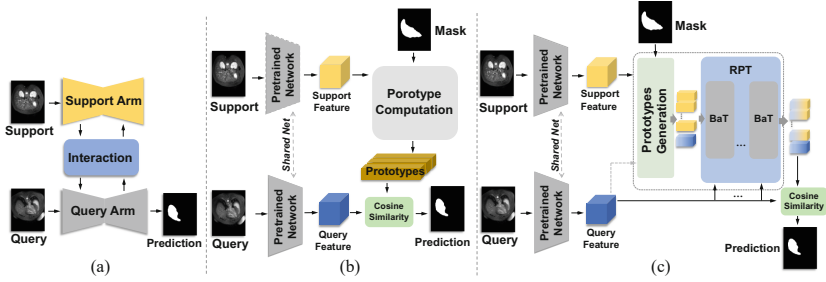
Fig. 1. Comparison between previous FSMS models and our model. (a) Interactive model. (b) Prototypical network based model. (c) Our proposed model.

because they can generate measurements and segmentations from medical images without the time-consuming manual work as in traditional methods. However, the performance of deep learning algorithms depends heavily on the availability of large-scale, high-quality, fully pixel-wise annotations, which are often expensive to acquire. To this end, few-shot learning is considered as a more promising approach and introduced into the medical image segmentation by [13].

Through revisiting existing FSMS algorithms [3–5,16,17,19], they can be grouped into two folders, including the interactive method originated from SENet [15] (shown in Fig. 1(a)) and the prototype networks [18,20] (demonstrated in Fig. 1(b)). For the interaction-based approach, the ideas of *attention* [19], and *contrastive learning* [22] are introduced to work interactively between parallel support and query arms. In contrast, prototype network-based approach almost dominates the FSMS research, such as SSL-ALPNet [13], ADNet [5] and SR&CL [21], whose core idea is to obtain semantic-level prototypes by compressing support features, and then make predictions by matching with query features. However, the problem of how to obtain an accurate and representative prototype remains.

The main reason affecting the representativeness of the prototype is the significant discrepancy between support and query. Specifically, in general, different protocols are taken for different patients, which results in a variety of superficial organ appearances, including the *size*, *shape*, and *contour* of features. In this case, the prototype generated from the support features may not accurately represent the key attributes of the target organ in the query image. In addition, it is also challenging to extract useful information (prototypes of novel classes) from the cluttered background due to the extremely heterogeneous texture between the target and its surroundings, which may contain information belonging to some novel classes or redundant information issue [19].

To mitigate the impact of intra-class diversity, it considers subdividing the foreground of the supporting prototypes to produce some regional prototypes, which are then rectified to suppress or exclude areas inconsistent with the query targets, as illustrated in Fig. 1(c). Concretely, in the prototype learning stage, multiple subdivided regional prototypes are enhanced with a more accurate class

center, which can be derived from the newly designed Regional Prototype Generation (RPG) and Query Prototype Generation (QPG) modules. Then, a designed Region-enhanced Prototypical Transformer (RPT) that is mainly composed of a number of stacked Bias-alleviated Transformer (BaT) blocks, each of which contains the core debiasing function-Search and Filter (S&F) modules, to filter out undesirable prototypes. As shown in Fig. 2, Our contributions are summarized as follows:

– A Region-enhanced Prototypical Transformer (RPT) consisting of stacked Bias-alleviated Transformer (BaT) blocks is proposed to mitigate the effects of large intra-class variations present in FSMS through Search and Filter (S&F) modules devised based on the self-selection mechanism.
– A subdivision strategy is proposed to perform in the foreground of the support prototype to generate multiple regional prototypes, which can be further iteratively optimized by the RPT to produce the optimal prototype.
– The proposed method can achieve state-of-the-art performance on three experimental datasets commonly used in medical image segmentation.

2   Methodology

2.1   Overall Architecture

Before introducing the overall architecture, it is necessary to briefly explain how data is processed. Specifically, the 3D supervoxel clustering method [5] is employed to generate pseudo-masks as supervision, which is learned in a self-supervised learning manner without any manual annotations. Meta-learning-based episodic tasks can then be constructed using the generated pseudo-masks. Notably, the pseudo-masks obtained by the 3D clustering method is more consistent with the volumetric properties of medical images than the 2D superpixel clustering method adopted in [13].

As depicted in Fig. 2, the overall architecture includes three main components: the Regional Prototype Generation (RPG) module, the Query Prototype Generation (QPG) module and the Region-enhanced Prototypical Transformer (RPT) consisting of three Bias-alleviated Transformer (BaT) blocks. The pipeline first extracts features from support and query images using a weight-shared ResNet-101 [6] as a backbone, which has been pretrained on the MS-COCO dataset [10]. We employ the ResNet101 pretrained on MS-COCO for optimal performance, and the comparison with ResNet50 pretrained on ImageNet dataset [2] is also included in the appendix. The extracted features are then taken as the input of the RPG and QPG modules to generate multiple region prototypes, which will be rectified by the following RPT to produce the optimal prototype.

2.2   Regional Prototype Generation

The core problem considered in this paper is what causes prototype bias. By examining the input data, it can be observed that images of healthy and diseased organs have a chance to be considered as support or query. This means that
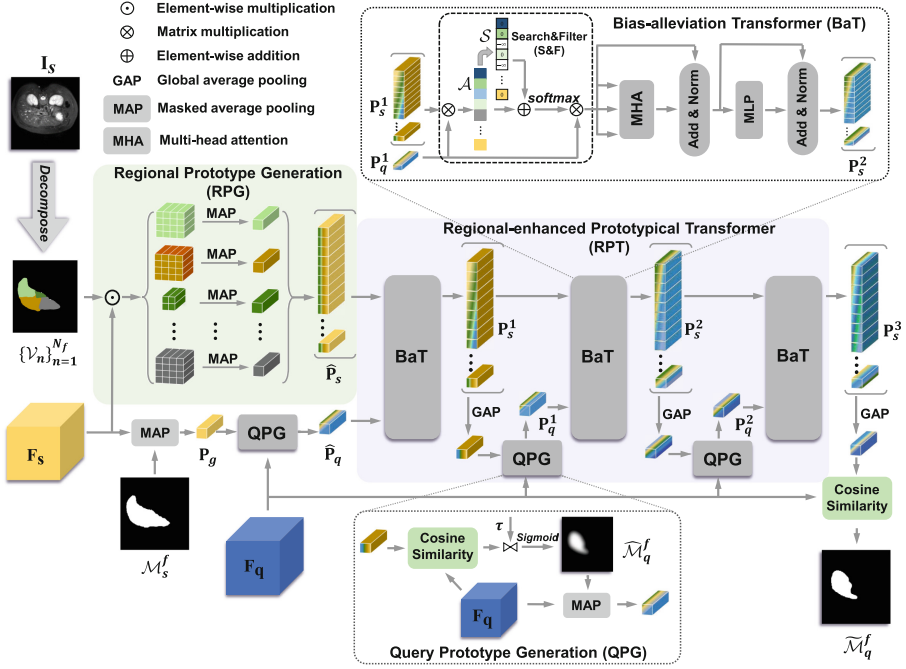
Fig. 2. Overview of the proposed Region-enhanced Prototypical Transformer.

if there are lesioned or edematous regions in some areas of the support images, they will be regarded as biased information which in reality cannot be accurately transferred for the query images containing only healthy organs. When these prototypes that contain the natural heterogeneity of the input images are processed by the Masked Average Pooling (MAP) operation, they inevitably lead to significant intra-class biases.

To cope with the above problems, we propose a Region Prototype Generation (RPG) module to generate multi-region prototypes by performing subdivisions in the foreground of the support images. Given an input support image $\mathbf{I}_s$ and the corresponding foreground mask $\mathcal{M}^f$, the foreground of this image can be obtained by calculating their product. The foreground image then can be partitioned into N_f regions, where N_f is set to 10 by default. By using the Voronoi-based partition method [1,23], a set of regional masks $\{\mathcal{V}_n\}_{n=1}^{N_f}$ can be derived for subsequent use of Masked Average Pooling (MAP) to generate a set of coarse regional prototypes $\hat{\mathcal{P}}_s = \{\hat{p}_n\}_{n=1}^{N_f}, \hat{p}_n \in \mathbb{R}^C$. Formally,

$$\hat{p}_n = \mathrm{MAP}(\mathbf{F}_s, \mathcal{V}_n) = \frac{1}{|\mathcal{V}_n|} \sum_{i=1}^{HW} \mathbf{F}_{s,i}\mathcal{V}_{n,i}, \tag{1}$$

where $\mathbf{F}_s \in \mathbb{R}^{C \times H \times W}$ is the feature extracted from the support images and $\mathcal{V}_n$ denotes the regional masks.

2.3 Query Prototype Generation

Once a set of coarse regional prototypes $\hat{\mathcal{P}}_s$ have been generated for the support images, we can employ the method introduced in [11] to learn the coarse query prototype $\hat{\mathbf{P}}_q \in \mathbb{R}^{1 \times C}$. Concretely, it first uses the MAP($\cdot$) operator as introduced in Eq. (1) to learn a global support prototype $\mathbf{P}_g = \text{MAP}(\mathbf{F}_s, \mathcal{M}_s)$ with $\mathbf{P}_g \in \mathbb{R}^{1 \times C}$, whose output can then be used to calculate the coarse query foreground mask $\hat{\mathcal{M}}_q^f$. Considering that the empirically designed threshold described in [11] may affect the quality of the $\hat{\mathcal{M}}_q^f$, we hereby introduce a learnable threshold τ. This process can be denoted as

$$\hat{\mathcal{M}}_q^f = 1 - \sigma(S(\mathbf{F}_q, \mathbf{P}_g) - \tau), \tag{2}$$

where $\mathbf{F}_q \in \mathbb{R}^{C \times H \times W}$ is feature extracted from query images, $S(a, b) = -\alpha cos(a, b)$ is the negative cosine similarity with a fixed scaling factor $\alpha = 20$, σ denotes the Sigmoid activation, and τ is obtained by applying one average-pooling and two fully-connected layers (FC) to the query feature, expressed as $\tau = \text{FC}(\mathbf{F}_q)$. After this, the coarse query foreground prototype can be achieved by using $\hat{\mathbf{P}}_q = \text{MAP}(\mathbf{F}_{q,i}, \hat{\mathcal{M}}_{q,i}^f)$.

2.4 Region-Enhanced Prototypical Transformer

The above received prototypes $\hat{\mathcal{P}}_s$ and $\hat{\mathbf{P}}_q$ are taken as input to the proposed Region-enhanced Prototypical Transformer (RPT) to rectify and regenerate the optimal global prototype $\mathbf{P}_s$. As shown in Fig. 2, our RPT mainly consists of L stacked Bias-alleviated Transformer (BaT) blocks each of which contains a Search and Filter (S&F) module, and QPG modules that maintain the query prototypes continuously updated. Taking the first BaT block as an example, it calculates an affinity map $\mathcal{A} = \hat{\mathbf{P}}_s \hat{\mathbf{P}}_q^\top \in \mathbb{R}^{N_f \times 1}$ to reveal the correspondence between the query and N_f support regional prototypes by taking an input containing the query prototype $\hat{\mathbf{P}}_q$ and the support prototype $\hat{\mathbf{P}}_s \in \mathbb{R}^{N_f \times C}$ obtained by concatenating all elements in $\hat{\mathcal{P}}_s$. Then, a selective map $\mathcal{S} \in \mathbb{R}^{N_f \times 1}$ can be derived from the proposed self-selection based S&F module by

$$\mathcal{S}_i(\mathcal{A}_i) = \begin{cases} 0 & \text{if } \mathcal{A}_i >= \xi \\ -\infty & otherwise \end{cases}, i \in \{0, 1, ..., N_f\}, \tag{3}$$

where ξ is the selection threshold achieved by $\xi = (min(\mathcal{A}) + mean(\mathcal{A}))/2$, $\mathcal{S}$ indicates the chosen regions from the support image that performs compatible with the query at the prototypical level. Then, the heterogeneous or disturbing regions of support foreground will be weeded out with softmax($\cdot$) function. The preliminary rectified prototypes $\hat{\mathbf{P}}_s^o \in \mathbb{R}^{N_f \times C}$ is aggregated as:

$$\hat{\mathbf{P}}_s^o = \text{softmax}(\hat{\mathbf{P}}_s \hat{\mathbf{P}}_q^\top + \mathcal{S}) \hat{\mathbf{P}}_q. \tag{4}$$

The refined $\hat{\mathbf{P}}_s^o$ will be fed into the following components designed based on the multi-head attention mechanism to produce the output $\mathbf{P}_s^1 \in \mathbb{R}^{N_f \times C}$. Formally,

$$\hat{\mathbf{P}}_s^{o+1} = \text{LN}(\text{MHA}(\hat{\mathbf{P}}_s^o) + \hat{\mathbf{P}}_s^o), \qquad \mathbf{P}_s^1 = \text{LN}(\text{MLP}(\hat{\mathbf{P}}_s^{o+1}) + \hat{\mathbf{P}}_s^{o+1}), \qquad (5)$$

where $\hat{\mathbf{P}}_s^{o+1} \in \mathbb{R}^{N_f \times C}$ is the intermediate generated prototype, $\text{LN}(\cdot)$ denotes the layer normalization, $\text{MHA}(\cdot)$ represents the standard multi-head attention module and $\text{MLP}(\cdot)$ is the multilayer perception.

By stacking multiple BaT blocks, our RPT can iteratively rectify and update all coarse support and the query prototype. Given the prototypes $\mathbf{P}_s^{l-1}$ and $\mathbf{P}_q^{l-1}$ from the previous BaT block, the updates for the current BaT block are computed by:

$$\mathbf{P}_s^l = \text{BaT}(\mathbf{P}_s^{l-1}, \mathbf{P}_q^{l-1}), \qquad \mathbf{P}_q^l = \text{QPG}(\text{GAP}(\mathbf{P}_s^l), \mathbf{F}_q), \qquad (6)$$

where $\mathbf{P}_s^l \in \mathbb{R}^{N_f \times C}$ and $\mathbf{P}_q^l \in \mathbb{R}^{1 \times C}$ ($l = 1, 2, ..., L$) are updated prototypes, $\text{GAP}(\cdot)$ denotes the global average pooling operation. The final output prototypes $\mathbf{P}_s$ optimized by the RPT can be used to predict the foreground of the query image by using Eq. (2: $\tilde{\mathcal{M}}_q^f = 1 - \sigma(S(\mathbf{F}_q, \text{GAP}(\mathbf{P}_s^3)) - \tau)$, while its background can be obtained by $\tilde{\mathcal{M}}_q^b = 1 - \tilde{\mathcal{M}}_q^f$ accordingly.

2.5   Objective Function

The binary cross-entropy loss $\mathcal{L}_{ce}$ is adopted to determine the error between the predict masks $\tilde{\mathcal{M}}_q$ and the given ground-truth $\mathcal{M}_q$. Formally,

$$\mathcal{L}_{ce} = -\frac{1}{HW} \sum_h^H \sum_w^W \mathcal{M}_q^f(x,y) log(\tilde{\mathcal{M}}_q^f(x,y)) + \mathcal{M}_q^b(x,y) log(\tilde{\mathcal{M}}_q^b(x,y)). \qquad (7)$$

Considering the prevalent class imbalance problem in medical image segmentation, the boundary loss [8] $\mathcal{L}_B$ is also adopted and it is written as

$$\mathcal{L}_B(\theta) = \int_\Omega \phi G(q) s_\theta(q) d_q, \qquad (8)$$

where θ denotes the network parameters, Ω denotes the spatial domain, $\phi G : \Omega \rightarrow \mathbb{R}$ denotes the *level set* representation of the ground-truth boundary, $\phi G(q) = -D_G(q)$ if $q \in G$ and $\phi G(q) = D_G(q)$ otherwise, D_G is distance map between the boundary of prediction and ground-truth, and $s_\theta(q) : \Omega \rightarrow [0,1]$ denotes softmax($\cdot$) function.

Overall, the loss used for training our RPT is defined as $\mathcal{L} = \mathcal{L}_{ce} + \eta\mathcal{L}_{dice} + (1-\eta)\mathcal{L}_B$, where $\mathcal{L}_{dice}$ is the Dice loss [12], η is initially set to 1 and decreased by 0.01 every epoch.

Table 1. Quantitative Comparison (in Dice score %) of different methods on abdominal datasets under *Setting 1* and *Setting 2*.

Setting	Method	Reference	Abd-MRI					Abd-CT				
			Lower		Upper		Mean	Lower		Upper		Mean
			LK	RK	Spleen	Liver		LK	RK	Spleen	Liver	
1	ADNet [5]	MIA'22	73.86	85.80	72.29	82.11	78.51	72.13	79.06	63.48	77.24	72.97
	AAS-DCL [22]	ECCV'22	80.37	86.11	76.24	72.33	78.76	74.58	73.19	72.30	78.04	74.52
	SR&CL [21]	MICCAI'22	79.34	87.42	76.01	80.23	80.77	73.45	71.22	**73.41**	76.06	73.53
	CRAPNet [3]	WACV'23	**81.95**	86.42	74.32	76.46	79.79	74.69	74.18	70.37	75.41	73.66
	Ours (RPT)	—	80.72	**89.82**	**76.37**	**82.86**	**82.44**	**77.05**	**79.13**	72.58	**82.57**	**77.83**
2	ADNet [5]	MIA'22	59.64	56.68	59.44	**77.03**	63.20	48.41	40.52	50.97	70.63	52.63
	AAS-DCL [22]	ECCV'22	76.90	83.75	74.86	69.94	76.36	64.71	**69.95**	66.36	71.61	68.16
	SR&CL [21]	MICCAI'22	77.07	84.24	73.73	75.55	77.65	67.39	63.37	67.36	73.63	67.94
	CRAPNet [3]	WACV'23	74.66	82.77	70.82	73.82	75.52	70.91	67.33	70.17	70.45	69.72
	Ours (RPT)	—	**78.33**	**86.01**	**75.46**	76.37	**79.04**	**72.99**	67.73	**70.80**	**75.24**	**71.69**

3 Experiments

Experimental Datasets: The proposed method is comprehensively evaluated on three publicly available datasets, including **Abd-MRI**, **Abd-CT** and **Card-MRI**. Concretely, **Abd-MRI** [7] is an abdominal MRI dataset used in the ISBI 2019 Combined Healthy Abdominal Organ Segmentation Challenge. **Abd-CT** [9] is an abdominal CT dataset from MICCAI 2015 Multi-Atlas Abdomen Labeling Challenge. **Card-MRI** [24] is a cardiac MRI dataset from MICCAI 2019 Multi-Sequence Cardiac MRI Segmentation Challenge. All 3D scans are reformatted into 2D axial and 2D short-axis slices. The abdominal datasets **Abd-MRI** and **Abd-CT** share the same categories of labels which includes the liver, spleen, left kidney (LK) and right kidney (RK). The labels for **Card-MRI** include left ventricular myocardium (LV-MYO), right ventricular myocardium (RV), and blood pool (LV-BP).

Experiment Setup: The model is trained for 30k iterations with batch size set to 1. During training, the initial learning rate is set to 1×10^{-3} with a step decay of 0.8 every 1000 iterations. The values of N_f and iterations L are set to 10 and 3, respectively. To simulate the scarcity of labeled data in medical scenarios, all experiments embrace a 1-way 1-shot setting, and 5-fold cross-validation is also carried out in the experiments, where we only record the mean value.

Evaluation: For a fair comparison, the metric used to evaluate the performance of 2D slices on 3D volumetric ground-truth is the Dice score used in [13]. Furthermore, two different supervision settings are used to evaluate the generalization ability of the proposed method: in Setting 1, the test classes may appear in the background of the training slices, while in Setting 2, the training slices containing the test classes are removed from the dataset to ensure that the test classes are unseen. Note that Setting 2 is impractical for Card-MRI scans, since all classes typically co-occur on one 2D slice, making label exclusion impossible. In addition, as in [13], abdominal organs are categorized into *upper* abdomen
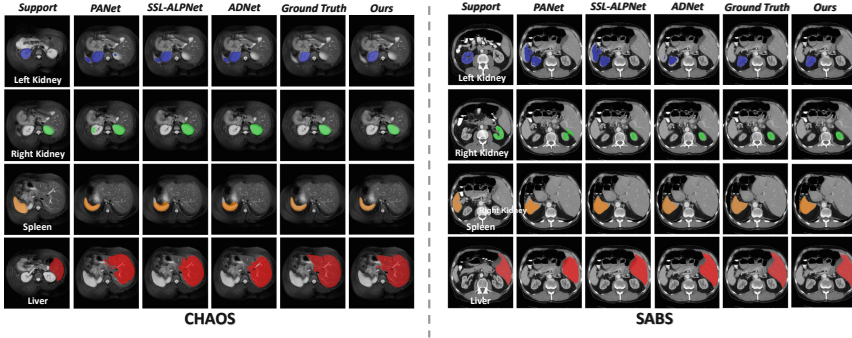
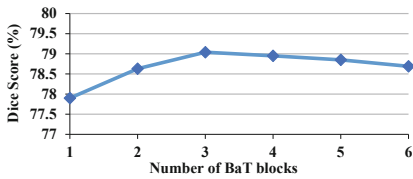**Fig. 3.** Qualitative results of our model on Abd-MRI and Abd-CT.

Fig. 4. Analysis of the number of BaT blocks.

Table 2. Ablation study of the three loss functions.

$\mathcal{L}_{ce}$	$\mathcal{L}_B$	$\mathcal{L}_{dice}$	Dice score
✓			78.43
✓	✓		78.81
✓	✓	✓	**79.04**

(liver, spleen) and *lower* abdomen (left, right kidney) to demonstrate whether the learned representations can encode spatial concepts.

3.1 Quantitative and Qualitative Results

Table 1 shows the performance comparison of the proposed method with state-of-the-art methods, including the vanilla PA-Net [20], SE-Net [15], ADNet [5], CRAPNet [3], SSL-ALPNet [13,14], AAS-DCL [22] and SR&CL [21] under two experimental settings. From Table 1, it can be seen that the proposed method outperforms all listed methods in terms of the Mean values obtained under two different settings. Especially, the Mean value on Abd-CT dataset under Setting 1 reaches 77.83, which is 3.31 higher than the best result achieved by AAS-DCL. Consistent improvements are also indicated for Card-MRI dataset and can be found in the Appendix. In addition to the quantitative comparisons, qualitative results of our model and the other model on Abd-MRI and Abd-CT are shown in Fig. 3 (See Appendix for CMR dataset). It is not difficult to see that our model shows considerable bound-preserving and generalization capabilities.

3.2 Ablation Studies

The ablation studies were conducted on Abd-MRI dataset under Setting 2. As can be seen from Fig. 4, the use of three stacked BaT blocks is suggested to

obtain the best Dice score. From Table 2, using a combination of boundary and dice loss gives a 0.61 increase in terms of the dice score compared to using only the cross-entropy loss. More ablation study results can be found in Appendix.

4   Conclusion

In this paper, we introduced a Region-enhanced Prototypical Transformer (RPT) to mitigate the impact of large intra-class variations present in medical image segmentation. The model is mainly beneficial from a subdivision-based strategy used for generating a set of regional support prototypes and a self-selection mechanism introduced to the Bias-alleviated Transformer (BaT) blocks. The proposed RPT can iteratively optimize the generated regional prototypes and output a more precise global prototype for predictions. The results of extensive experiments and ablation studies can demonstrate the advancement and effectiveness of the proposed method.

References

1. Aurenhammer, F.: Voronoi diagrams: a survey of a fundamental geometric data structure. ACM Comput. Surv. (CSUR) **23**(3), 345–405 (1991)
2. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: ImageNet: a large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255. IEEE (2009)
3. Ding, H., Sun, C., Tang, H., Cai, D., Yan, Y.: Few-shot medical image segmentation with cycle-resemblance attention. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 2488–2497 (2023)
4. Feng, R., et al.: Interactive few-shot learning: limited supervision, better medical image segmentation. IEEE Trans. Med. Imaging **40**(10), 2575–2588 (2021)
5. Hansen, S., Gautam, S., Jenssen, R., Kampffmeyer, M.: Anomaly detection-inspired few-shot medical image segmentation through self-supervision with supervoxels. Med. Image Anal. **78**, 102385 (2022)
6. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
7. Kavur, A.E., et al.: Chaos challenge-combined (CT-MR) healthy abdominal organ segmentation. Med. Image Anal. **69**, 101950 (2021)
8. Kervadec, H., Bouchtiba, J., Desrosiers, C., Granger, E., Dolz, J., Ayed, I.B.: Boundary loss for highly unbalanced segmentation. In: International Conference on Medical Imaging with Deep Learning, pp. 285–296 (2019)
9. Landman, B., Xu, Z., Igelsias, J., Styner, M., Langerak, T., Klein, A.: MICCAI multi-atlas labeling beyond the cranial vault-workshop and challenge. In: Proceedings of MICCAI Multi-Atlas Labeling Beyond Cranial Vault-Workshop Challenge, vol. 5, p. 12 (2015)
10. Lin, T.-Y., et al.: Microsoft COCO: common objects in context. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8693, pp. 740–755. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10602-1_48
11. Liu, J., Qin, Y.: Prototype refinement network for few-shot segmentation. arXiv preprint arXiv:2002.03579 (2020)

12. Ma, J., et al.: Loss odyssey in medical image segmentation. Med. Image Anal. **71**, 102035 (2021)
13. Ouyang, C., Biffi, C., Chen, C., Kart, T., Qiu, H., Rueckert, D.: Self-supervision with superpixels: training few-shot medical image segmentation without annotation. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) ECCV 2020. LNCS, vol. 12374, pp. 762–780. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58526-6_45
14. Ouyang, C., Biffi, C., Chen, C., Kart, T., Qiu, H., Rueckert, D.: Self-supervised learning for few-shot medical image segmentation. IEEE Trans. Med. Imaging **41**(7), 1837–1848 (2022)
15. Roy, A.G., Siddiqui, S., Pölsterl, S., Navab, N., Wachinger, C.: 'squeeze & excite' guided few-shot segmentation of volumetric images. Med. Image Anal. **59**, 101587 (2020)
16. Shen, Q., Li, Y., Jin, J., Liu, B.: Q-Net: query-informed few-shot medical image segmentation. arXiv preprint arXiv:2208.11451 (2022)
17. Shen, X., Zhang, G., Lai, H., Luo, J., Lu, J., Luo, Y.: PoissonSeg: semi-supervised few-shot medical image segmentation via poisson learning. In: IEEE International Conference on Bioinformatics and Biomedicine, pp. 1513–1518 (2021)
18. Snell, J., Swersky, K., Zemel, R.: Prototypical networks for few-shot learning. In: Advances in Neural Information Processing Systems, vol. 30 (2017)
19. Sun, L., et al.: Few-shot medical image segmentation using a global correlation network with discriminative embedding. Comput. Biol. Med. **140**, 105067 (2022)
20. Wang, K., Liew, J.H., Zou, Y., Zhou, D., Feng, J.: PANet: few-shot image semantic segmentation with prototype alignment. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 9197–9206 (2019)
21. Wang, R., Zhou, Q., Zheng, G.: Few-shot medical image segmentation regularized with self-reference and contrastive learning. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 514–523 (2022)
22. Wu, H., Xiao, F., Liang, C.: Dual contrastive learning with anatomical auxiliary supervision for few-shot medical image segmentation. In: Avidan, S., Brostow, G., Cissé, M., Farinella, G.M., Hassner, T. (eds.) European Conference on Computer Vision, pp. 417–434. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-20044-1_24
23. Zhang, J.W., Sun, Y., Yang, Y., Chen, W.: Feature-proxy transformer for few-shot segmentation. In: Advance in Neural Information Processing Systems (2022)
24. Zhuang, X.: Multivariate mixture model for myocardial segmentation combining multi-source images. IEEE Trans. Pattern Anal. Mach. Intell. **41**(12), 2933–2946 (2018)