



Multi-shot Prototype Contrastive Learning and Semantic Reasoning for Medical Image Segmentation

Yuhui Song¹, Xiuquan Du¹, Yanping Zhang¹, and Chenchu Xu^{1,2}(✉)

¹ Artificial Intelligence Institute, Anhui University, Hefei, China
cxu332@gmail.com

² Institute of Artificial Intelligence, Hefei Comprehensive National Science Center, Hefei, China

Abstract. Despite the remarkable achievements made by deep convolutional neural networks in medical image segmentation, the limitation that they rely heavily on high-precision and intensively annotated samples makes it difficult to adapt to novel classes that have not been seen before. Few-shot learning is introduced to solve these challenges by learning the generalized representation of a semantic class from very few annotated support samples that can be used as a reference for unannotated query samples. In this paper, instead of averaging multiple support prototypes, we propose a multi-shot prototype contrastive learning and semantic reasoning network (MPSNet) for medical image segmentation. The multi-shot learning network exists independently within the support set, obtains effective semantic features for support images and gives priority to training the core segmentation model of prototype contrastive learning. We also propose a semantic reasoning network that takes the prior semantic features and prior segmentation model learned from the support set as the immediate and necessary conditions for the query image to deduce its segmentation mask. The proposed method is verified to be superior to the state-of-the-art methods on three public datasets, revealing its powerful segmentation and generalization abilities. Code: <https://github.com/H51705/FSS-MPSNet>.

Keywords: Medical image segmentation · Multi-shot learning · Prototype contrastive learning · Semantic reasoning

1 Introduction

The successes of deep convolutional neural networks (CNNs) [9] in image tasks [12, 20–22] rely heavily on a large number of intensively annotated training samples, and obtaining these annotated images is time-consuming, laborious and costly. In addition to the extreme lack of training data, these segmentation models also have low generalization ability due to the specificity of training classes.

This limitation is more significant in medical images. Therefore, few-shot learning (FSL) [11, 14, 17] emerged to deal with these challenges. By learning the generalized representation of a certain semantic class from a few labeled samples (i.e., support images) to guide the segmentation of that class in a large number of unlabeled images (i.e., query images).

The development of FSL has derived many effective medical image segmentation methods, including pure CNNs [2, 10] and CNNs with a nonparametric learning module embedded [1], among which the prototype network is the most widely used [8, 15, 23]. For example, Yu et al. [23] proposed a location-sensitive local prototype network. Tang et al. [15] proposed a recurrent mask refinement algorithm based on a prototype network. In addition, Ouyang et al. [8] introduced a self-supervised super-pixel few-shot segmentation (FSS) method to get rid of the demand for labeled data in the training stage. Although FSS methods based on prototype networks have the ability to achieve image segmentation, when the prototype learned from the support image is used to judge the pixel category of the query image, it still cannot avoid the problem of insufficient prototype discriminability caused by the differences between the images as well as the complexity of the target. Furthermore, it is easy to lose important information when the target features are taken as a prototype by GAP (global average pooling). In this regard, Wang et al. [18] introduced a prototype alignment regularization between the support image and the query image to better utilize the information of the support set. Meanwhile, compared with one-shot mode, the multi-shot prototype strategy is widely used to improve segmentation performance [6, 8, 16]. However, the existing multi-shot strategies usually take the average of multiple support prototypes as the final prototype, which not only weakens the independent contribution of multiple support images but also destroys the semantic information of support prototypes. Especially for medical images with lower image intensity and contrast, prototype alignment regularization and the mean prototype strategy do not significantly increase the discriminant ability and prediction accuracy of prototypes. Especially for medical images with lower image intensity and contrast, prototype alignment regularization and mean prototype strategy may aggravate the confusion of target pixel categories. For pure prototype technology, Zhou et al. [25] explored a better non-parametric segmentation model based on non-learnable prototypes in their latest work and successfully applied it to the semantic segmentation of natural images, but this method has not been verified in medical images.

In this paper, we propose a prototype contrastive learning and semantic reasoning network based on multi-shot strategy (MPSNet). To our knowledge, this is a great improvement in medical image analysis. First, we design a multi-shot learning network (MLN) within the support set to generate prior semantic features and a prior segmentation model for the query image. Cross-validation mode is used to construct the support-query image pairs within the support set, and a complete FSS model is formed by prototype contrastive learning and supervised training. Second, we propose a prototype contrastive learning module (PCLM) to ascertain the positive contribution of multiple support images to the query image in segmentation guidance, leading to better segmentation performance of

the query image. Third, we design a semantic reasoning network (SRN) that is convenient to directly transfer the prior semantic features and prior segmentation model to the query image to deduce its segmentation mask quickly. Our contributions can be summarized as follows:

- 1) A novel FSS method based on a multi-shot prototype strategy is proposed for the first time to replace the commonly used mean prototype method to improve the guidance ability of the support images to the query images in segmentation.
- 2) A multi-shot prototype contrastive learning network within the support set is constructed with supervised training to generate prior semantic features and a prior segmentation model, and transfer them to the query image to deduce its segmentation mask.
- 3) The proposed method achieves the latest performance on three public datasets that is superior to the state-of-the-art (SOTA) methods.

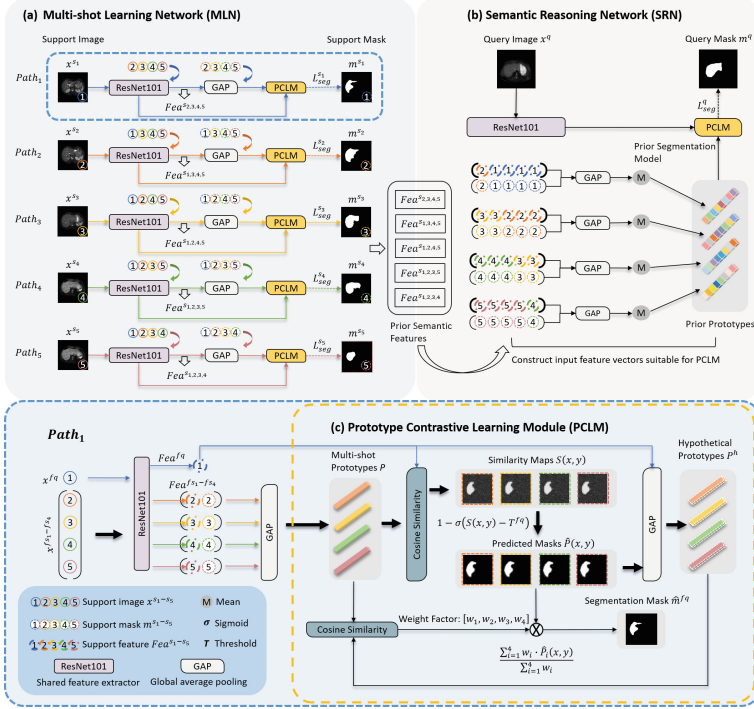


Fig. 1. The proposed MPSNet consists of three parts: (a) the multi-shot learning network (MLN), (b) the semantic reasoning network (SRN) and (c) the prototype contrastive learning module (PCLM). PCLM is a core segmentation module that is embedded in both MLN and SRN. MPSNet follows a 1-way 5-shot learning mode. For each episode, the input is 5 support image-mask pairs and a query image, and the output is the segmentation mask of the query image.

2 Method

2.1 Problem Setting

In FSS problem, a segmentation model is trained on a training set D_{train} , then the model is used to evaluate a test set D_{test} . Assumes that the semantic classes in D_{train} is C_{train} and in D_{test} is C_{test} , the particular definition of FSL is $C_{train} \cap C_{test} = \phi$. The training set and test set have a common intrinsic pattern that both contain a support set S and a query set Q . At training stage, we use episode mode to set N -way K -shot segmentation task, where N represents the number of semantic classes contained in each episode and K represents the number of images contained in each class. For the task of few-shot medical image segmentation, N is usually set to 1, that is, each episode only sets the support set and the query set for a specific class c . For one episode, suppose that the support set and the query set are $S_{train} = \{(I_k^s, M_k^s)\}_{k=1}^{K^s}$ and $Q_{train} = \{(I_n^q, M_n^q)\}_{n=1}^{N^q}$. The support images, masks and the query image constitute the model input $Input_{model} = \{I^s, M^s, I^q\}$, query mask M^q as the supervisory information. At testing stage, we set episodes with the same mode, that is, we use the trained FSS model to segment the query set under the function of the support set. Since our research is carried out based on multi-shot strategy, we mainly take 5-shot as an example to verify the effectiveness of the proposed method.

2.2 Multi-shot Learning Network

The framework of MLN is shown in Fig. 1(a), it is a supervised learning network that exists independently within the support set, mainly including 5 same structures but different input and output of 4-shot learning branch $\{Path_l\}_{l=1}^5$. When one support image is used as the fake query image, the remaining 4 support images are used as the fake support images. Therefore, in the support set, each support image can be used as a fake query image to form a 4-shot learning network. The specific implementation process can be expressed as follows:

$$\hat{m}^{s_i} / \hat{m}^{f_q} = PCLM(GAP(RN(\{x^{fs_i}\}_{i=1}^4), \{m^{fs_i}\}_{i=1}^4), RN(x^{fq})), \quad (1)$$

where $RN(\cdot)$ refers to the shared feature extractor ResNet101, $GAP(\cdot)$ refers to global average pooling, $PCLM(\cdot)$ refers to the proposed prototype contrastive learning module. x^{fs_i} , m^{fs_i} , x^{fq} represent the fake support image, mask and query image. We use cross-entropy (CE) loss to supervise the training process:

$$L_{seg}^s = \sum_{i=1}^5 L_{seg}^{s_i} = \sum_{i=1}^5 CE(\hat{m}^{s_i}, m^{s_i}). \quad (2)$$

As shown in the lower part of Fig. 1, we take $Path_1$ branch as an example in detail the specific structures. Fake support images $\{x^{fs_i}\}_{i=1}^4 \in \mathbb{R}^{3 \times H \times W}$ and fake query image $x^{fq} \in \mathbb{R}^{3 \times H \times W}$ are fed into ResNet101 for support features $\{Fea^{fs_i}\}_{i=1}^4 \in \mathbb{R}^{C \times h \times w}$ and query features $Fea^{fq} \in \mathbb{R}^{C \times h \times w}$. Under

the function of support masks $\{m^{fs_i}\}_{i=1}^4 \in \mathbb{R}^{1 \times H \times W}$, the support prototypes $P = \{P_i\}_{i=1}^4$ have been extracted from the support features, denoted as $P_i \in \mathbb{R}^{C \times 1 \times 1}$, this process is implemented by GAP:

$$P_i = \frac{\sum Fea^{fs_i} \odot m^{fs_i}}{\sum m^{fs_i}}, \quad (3)$$

where support features Fea^{fs_i} are resized to the mask size (H, W) , (h, w) denotes the shape of the feature maps, C denotes the channel number of the feature maps, \odot denotes the Hadamard product. The obtained multi-shot prototypes P and query features Fea^{fq} are sent into PCLM to get the final segmentation of the query image.

2.3 Prototype Contrastive Learning Module

As shown in Fig. 1(c), PCLM is embedded in MLN and SRN as a core segmentation module, assumes the task of predicting the pixel-level categories of the query image by taking the multi-shot prototypes as evaluation criteria. Following Hansen et al. [3], we consider using a single foreground prototype as the feature compression of the target region, and adopt the anomaly score thresholding method to segment the query image. This method not only avoids the construction of the prototype for the background with more impurities, but also saves the calculation cost compared with the decoder. First of all, the similarity maps $\{S^i(x, y)\}_{i=1}^4$ are obtained by doing cosine similarity between the multi-shot prototypes P and the query features Fea^{fq} :

$$S^i(x, y) = -\alpha \frac{Fea^{fq}(x, y) \cdot P_i}{\|Fea^{fq}(x, y)\| \|P_i\|}. \quad (4)$$

The foreground prediction mask is then obtained:

$$\hat{m}_{fore_i}^{fq} = 1 - \delta(S^i(x, y) - T^{fq}) = 1 - \delta(S^i(x, y) - g_\theta(Fea^{fq})), \quad (5)$$

where (x, y) refers to the spatial location of the query image, and the prototype generates a similarity coefficient for each pixel. $\delta(\cdot)$ denotes the Sigmoid function with a steepness parameter $k = 0.5$. T^{fq} is the correlation threshold obtained by the query features Fea^{fq} through the full connection layers $g_\theta(\cdot)$. Then, we assume that the prediction mask is no different from the real mask, and GAP it with Fea^{fq} to obtain the hypothetical prototypes $P^h = \{P_i^h\}_{i=1}^4$:

$$P_i^h = \frac{\sum Fea^{fq} \odot \hat{m}_{fore_i}^{fq}}{\sum \hat{m}_{fore_i}^{fq}}. \quad (6)$$

Theoretically, under the assumption that there is no pixel intensity and contrast difference between the query image and the support image, the hypothetical

prototypes P^h should be similar to the multi-shot prototypes P . But in reality, it is difficult to make the query image and the support images have a consistent feature hierarchy. That is to say, the prior knowledge provided by multiple support images for the query image is not equally important. We need to assign a weight factor w_i to each support image to represent its positive contribution to query image segmentation. To this end, we determine the weight of each foreground prediction mask by comparing the hypothetical prototypes P^h with real support prototypes P :

$$w_i = \frac{P_i^h \cdot P_i}{\|P_i^h\| \|P_i\|}. \quad (7)$$

At this time, the foreground and background segmentation mask can be modified as:

$$\hat{m}_{fore}^{fq} = \frac{\sum_{i=1}^4 w_i \cdot \hat{m}_{fore_i}^{fq}}{\sum_{i=1}^4 w_i}, \quad (8)$$

$$\hat{m}_{back}^{fq} = 1 - \hat{m}_{fore}^{fq}, \quad (9)$$

$$\hat{m}^{fq} = \text{concatenate}(\hat{m}_{fore}^{fq}, \hat{m}_{back}^{fq}), \quad (10)$$

we take \hat{m}^{fq} as the output of PCLM that is also the final segmentation mask.

2.4 Semantic Reasoning Network

In our 5-shot segmentation task, the query image needs to be accurately segmented under the guidance of the given 5 support images. Such an FSS task can be realized through the migration of the prior semantic features and prior segmentation model from MLN. Specifically, as shown in Fig. 1(b), we directly prototype the support features obtained from MLN with the corresponding support masks. It is worth noting that the support images in each supervised training path are gathered together to form the support feature-mask cluster to obtain the prototype cluster. The purpose of this is to make the FSS network designed for query images conform to the laws of prior semantic features and the prior segmentation model. At the same time, prior prototypes are obtained by internal averaging of the prototype cluster. On this basis, the implementation of the same PCLM module as MLN can output the segmentation mask \hat{m}^q of the query image, and use the query mask m^q to conduct supervised training:

$$\hat{m}^q = PCLM(GAP(\{Fea^{s_i}\}_{i=1}^5), \{m^{s_i}\}_{i=1}^5), RN(x^q), \quad (11)$$

$$L_{seg}^q = CE(\hat{m}^q, m^q), \quad (12)$$

$$L_{seg} = L_{seg}^s + L_{seg}^q, \quad (13)$$

so far, the overall loss of the proposed segmentation network MPSNet is L_{seg} .

3 Experiments

3.1 Experimental Setup

Datasets: We evaluate the proposed MPSNet on three public datasets: the Combined (CT-MR) Healthy Abdominal Organ Segmentation (CHAOS) dataset [4] (Abd-MRI) with 20 3D T2-SPiR MRI scan images, the MICCAI 2015 Multi-Atlas Abdomen Labeling Challenge [5] (Abd-CT) with 30 3D CT scan images and the MICCAI 2019 Multi-sequence Cardiac MRI Segmentation Challenge (bSSFP fold) [26] (Cardiac-MRI) with 35 3D cardiac MRI scans. Our task is to segment the liver, right kidney (RK), left kidney (LK), and spleen from Abd-MRI and Abd-CT, and the left ventricle (LV), myocardium (Myo), and right ventricle (RV) from Cardiac MRI. In the preprocessing of these datasets, we follow Ouyang et al.’s scheme [8] to make a fair comparison with the SOTA methods. Similarly, we also use the super-pixels technology to generate the pseudo-labels for the preprocessed datasets.

Implementation Details: We use 5-fold cross validation to conduct training and testing. We chose the same allocation strategy for the support slices and the query slices as Roy et al. [10], and we randomly selected 1000 support-query image pairs to form the training set. All codes are based on PyTorch. We use the stochastic gradient descent (SGD) optimizer with a batch size of 1 and set the initial learning rate to 0.001, momentum to 0.9 and weight decay to 0.0005. The total number of iterations of the training is 30,000, and it takes an average of 9h to complete end-to-end training on an Nvidia RTX 3090 graphics card. To facilitate direct comparison with the SOTA methods, we follow the practice of most papers, using the pretrained ResNet101 framework as the feature encoder and the mean Sørensen-Dice coefficient (DSC) as the evaluation metric.

3.2 Superior Performance over the SOTA Methods

We compare the quantitative results of our method with the SOTA methods, including SE-Net [10], RP-Net [15], GCN-DE [13], ADNet [3], ALPNet [8] and RSCNet [19] are FSS methods specially proposed for medical images that have been verified on the datasets used by us. The remaining methods are classical FSS methods that are used in natural image. In particular, PANet [18] contains a prototype alignment strategy that is also applicable to medical images. As shown in Table 1, our method has obtained the best mean DSC of 82.01%, 74.72% and 77.90% for Abd-MRI, Abd-CT and Cardiac MRI respectively. Compared with the second-best method ALPNet [8], the mean DSC achieved by our method increased by 3.17%, 1.37% and 1%.

3.3 Ablation Studies

The specific implementation scheme of the ablation studies can be divided into three aspects: 1) MLN is removed from MPSNet to prove that the prior semantic

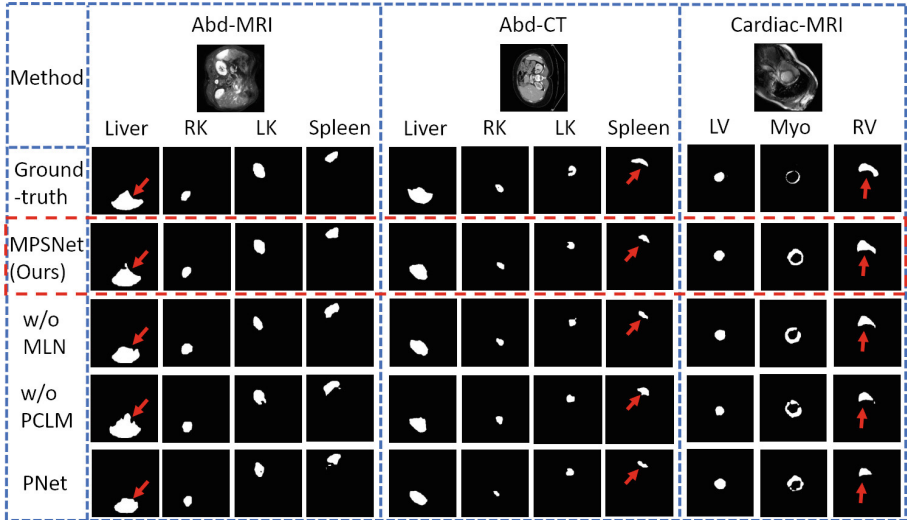
Table 1. Quantitative comparison between MPSNet and the SOTA methods.

Dataset	Abd-MRI					Abd-CT					Cardiac-MRI			
Methods	Liver	RK	LK	Spleen	Mean	Liver	RK	LK	Spleen	Mean	LV	Myo	RV	Mean
SE-Net [10]	27.43	61.32	62.11	51.80	50.66	38.20	23.60	32.70	32.53	31.76	58.04	25.18	12.86	32.03
RP-Net [15]	73.51	85.78	81.40	76.35	79.26	79.62	70.00	70.48	69.85	72.48	—	—	—	—
GCN-DE [13]	49.47	83.03	76.07	60.63	67.30	46.77	75.50	68.13	56.53	61.73	—	—	—	—
PANet [18]	66.59	42.18	44.32	54.59	51.90	69.16	36.42	38.83	42.40	46.70	70.62	46.03	67.16	61.27
PPNet [7]	73.12	71.78	62.13	66.57	68.40	—	—	—	—	—	67.78	42.61	60.80	57.06
CANet [24]	72.88	77.15	69.53	67.05	71.65	—	—	—	—	—	78.99	43.61	61.10	61.07
ADNet [3]	80.81	83.28	75.28	75.92	78.82	—	—	—	—	—	87.53	62.43	77.31	75.76
ALPNet [8]	76.10	85.18	81.92	72.18	78.84	78.29	71.81	72.36	70.96	73.35	83.99	66.74	79.96	76.90
RSCNet [19]	75.55	84.24	77.07	73.73	77.65	73.63	63.37	67.39	67.36	67.94	—	—	—	—
MPSNet (Ours)	81.88	87.26	82.10	76.80	82.01	81.13	69.56	70.26	77.96	74.72	87.98	65.08	80.63	77.90

features and the prior segmentation model are more effective than the independent segmentation based on the prototype network (w/o MLN). 2) We use PCLM to replace the mean prototype strategy to verify its improvement in segmentation performance (w/o PCLM). 3) The initial segmentation model based on the

Table 2. Quantitative results of the ablation studies.

Dataset	Abd-MRI					Abd-CT					Cardiac-MRI			
Methods	Liver	RK	LK	Spleen	Mean	Liver	RK	LK	Spleen	Mean	LV	Myo	RV	Mean
MPSNet	81.88	87.26	82.10	76.80	82.01	81.13	69.56	70.26	77.96	74.72	87.98	65.08	80.63	77.90
w/o MLN	79.18	78.88	70.62	66.90	73.90	82.25	57.21	58.48	66.28	66.05	84.04	51.83	68.47	68.12
w/o PCLM	80.52	86.51	81.26	74.72	80.75	80.89	69.29	68.29	75.37	73.46	86.01	62.51	77.19	75.24
PNet	81.13	78.63	74.35	68.81	75.73	82.85	62.48	64.20	67.43	69.24	83.42	50.08	66.76	66.75

**Fig. 2.** Visualization of query image segmentation in test set.

prototype network is reverted, that is, MLN and PCLM are removed, and the multi-shot support prototypes are directly averaged for similarity measurement with the query features (PNet). Note that SRN cannot be implemented when MLN and PCLM are absent.

Table 2 shows the quantitative results of the ablation studies. In the absence of MLN, the average DSC is 8.11%, 8.67% and 9.78% lower than MPSNet. In the absence of PCLM, the average DSC is 1.26%, 1.26% and 2.66% lower than MPSNet. More significantly, our method has improved average DSC by 6.28%, 5.48% and 11.15% over PNet due to the inclusion of MLN and PCLM. Ablation studies fully demonstrate the effectiveness of the proposed method and the necessity of the internal parts. As shown in Fig. 2, one subject is randomly selected for each dataset to present the visualization results. In Fig. 3, we plot the training loss of the support images and the query image. It's obvious that the query loss decreases faster and becomes more stable than the support loss, indicating that our proposed MLN has a positive effect.

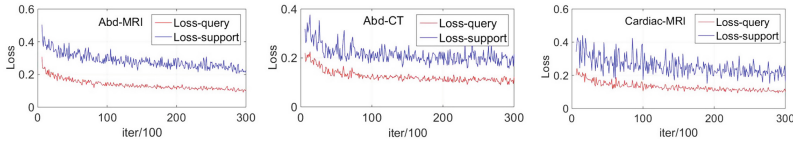


Fig. 3. Comparison of training loss between the support images and the query image. The value of the abscissa represents the loss measured every 100 iterations.

4 Conclusion

In this paper, we propose a novel FSS method MPSNet for medical image segmentation to replace the approach of averaging multiple support prototypes in the existing FSS methods. This is the first time that a multi-shot learning pattern is built within the support set and applied to the query image to significantly improve segmentation performance. Compared with the SOTA methods, our method is evaluated on three public datasets and achieves improvements in DSC of 3.17%, 1.37% and 1%, respectively. Relevant ablation studies also demonstrate the necessity and validity of our method.

Acknowledgment. This work was supported in part by The National Natural Science Foundation of China (62106001, U1908211), The University Synergy Innovation Program of Anhui Province (GXXT-2021-007, GXXT-2022-052), and The Anhui Provincial Natural Science Foundation (2208085Y19).

References

1. Feng, R., Zheng, X., Gao, T., Chen, J., Wu, J.: Interactive few-shot learning: Limited supervision, better medical image segmentation. *IEEE Trans. Med. Imaging* **40**, 2575–2588 (2021)
2. Guo, S., Xu, L., Feng, C., Xiong, H., Zhang, H.: Multi-level semantic adaptation for few-shot segmentation on cardiac image sequences. *Med. Image Anal.* **1**, 102170 (2021)
3. 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)
4. Kavur, A., Gezer, N., Bar, M., Aslan, S., Conze, P.H.: Chaos challenge-combined (ct-mr) healthy abdominal organ segmentation. *Med. Image Anal.* **69**, 101950 (2021)
5. Landman, B., Xu, Z., Igelsias, J., Styner, M., Langerak, T., Klein, A.: Miccai multi atlas labeling beyond the cranial vault-workshop and challenge. In: *MICCAI Multi-Atlas Labeling Beyond Cranial Vault-Workshop Challenge*, vol. 5, no. 12 (2015)
6. Li, G., Jampani, V., Sevilla-Lara, L., Sun, D., Kim, J.: Adaptive prototype learning and allocation for few-shot segmentation. In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8334–8343 (2021)
7. Liu, Y., Zhang, X., Zhang, S., He, X.: Part-aware prototype network for few-shot semantic segmentation. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) *ECCV 2020. LNCS*, vol. 12354, pp. 142–158. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58545-7_9
8. Ouyang, C., Biffi, C., Chen, C., Kart, T., Qiu, H., Rueckert, D.: Self-supervision with superpixels: training few-shot medical image segmentation without annotation, pp. 762–780 (2020)
9. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) *MICCAI 2015. LNCS*, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
10. Roy, A., Siddiqui, S., Polsterl, S., Navab, N., Wachinger, C.: ‘squeeze & excite’ guided few-shot segmentation of volumetric images. *Med. Image Anal.* **59**, 101587 (2020)
11. Snell, J., Swersky, K., Zemel, R.: Prototypical networks for few-shot learning. In: *Neural Information Processing Systems Foundation*, p. 4077–4087 (2017)
12. Song, Y., Du, X., Zhang, Y., Li, S.: Two-stage segmentation network with feature aggregation and multi-level attention mechanism for multi-modality heart images. *Comput. Med. Imaging Graph.* **97**, 102054 (2022)
13. Sun, L., Ding, X., Huang, Y., Wang, G., Yu, Y.: Few-shot medical image segmentation using a global correlation network with discriminative embedding. *Comput. Biol. Med.* **140**, 105067 (2022)
14. Sung, F., Yang, Y., Zhang, L., Xiang, T., Torr, P., Hospedales, T.: Learning to compare: relation network for few-shot learning. In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1199–1208 (2018)
15. Tang, H., Liu, X., Sun, S., Yan, X., Xie, X.: Recurrent mask refinement for few-shot medical image segmentation. In: *IEEE/CVF International Conference on Computer Vision*. p. 3918–3928 (2020)

16. Tian, Z., Zhao, H., Shu, M., Yang, Z., Jia, J.: Prior guided feature enrichment network for few-shot segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **99**, 1–1 (2020)
17. Vinyals, O., Blundell, C., Lillicrap, T., Kavukcuoglu, K., Wierstra, D.: Matching networks for one shot learning. In: *Neural Information Processing Systems Foundation*, pp. 3630–3638 (2016)
18. Wang, K., Liew, J., Zou, Y., Zhou, D., Feng, J.: Panet: few-shot image semantic segmentation with prototype alignment. In: *IEEE/CVF International Conference on Computer Vision*, pp. 9197–9206 (2019)
19. Wang, R., Zhou, Q., Zheng, G.: Few-shot medical image segmentation regularized with self-reference and contrastive learning. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) *MICCAI 202. LNCS*, vol. 13434, pp. 514–523. Springer, Heidelberg (2022). https://doi.org/10.1007/978-3-031-16440-8_49
20. Xu, C., Gao, Z., Zhang, H., Li, S., Albuquerque, V.: Video salient object detection using dual-stream spatiotemporal attention. *Appl. Soft Comput.* **108**, 107433 (2021)
21. Xu, C., et al.: Bmanet: boundary mining with adversarial learning for semi-supervised 2d myocardial infarction segmentation. *IEEE J. Biomed. Health Inf.* **27**(1), 87–96 (2023)
22. Xu, C., Zhang, D., Chong, J., Chen, B., Li, S.: Synthesis of gadolinium-enhanced liver tumors on nonenhanced liver mr images using pixel-level graph reinforcement learning. *Med. Image Anal.* **69**, 101976 (2021)
23. Yu, Q., Dang, K., Tajbakhsh, N., Terzopoulos, D., Ding, X.: A location-sensitive local prototype network for few-shot medical image segmentation. In: *IEEE International Symposium on Biomedical Imaging*, p. 262–266 (2021)
24. Zhang, C., Lin, G., Liu, F., Yao, R., Shen, C.: Canet: class-agnostic segmentation networks with iterative refinement and attentive few-shot learning. In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5217–5226 (2019)
25. Zhou, T., Wang, W., Konukoglu, E., Gool, L.: Rethinking semantic segmentation: a prototype view, pp. 2582–2593 (2022)
26. Zhuang, X.: Multivariate mixture model for myocardial segmentation combining multi-source images. *IEEE Trans. Pattern Anal. Mach. Intell.* **41**(12), 2933–2946 (2018)