UPCoL: Uncertainty-Informed Prototype Consistency Learning for Semi-supervised Medical Image Segmentation

Wenjing Lu[1], Jiahao Lei[2], Peng Qiu[2], Rui Sheng[3], Jinhua Zhou[4], Xinwu Lu[2,5], and Yang Yang[1(✉)]

[1] Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China
yangyang@cs.sjtu.edu.cn
[2] Department of Vascular Surgery, Shanghai Ninth People's Hospital Affiliated to Shanghai Jiao Tong University, Shanghai, China
[3] Chaohu Clinical Medcial College, Anhui Medical University, Hefei, China
[4] School of Biomedical Engineering, Anhui Medical University, Hefei, China
[5] Shanghai Key Laboratory of Tissue Engineering, Shanghai Ninth People's Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

Abstract. Semi-supervised learning (SSL) has emerged as a promising approach for medical image segmentation, while its capacity has still been limited by the difficulty in quantifying the reliability of unlabeled data and the lack of effective strategies for exploiting unlabeled regions with ambiguous predictions. To address these issues, we propose an Uncertainty-informed Prototype Consistency Learning (UPCoL) framework, which learns fused prototype representations from labeled and unlabeled data judiciously by incorporating an entropy-based uncertainty mask. The consistency constraint enforced on prototypes leads to a more discriminative and compact prototype representation for each class, thus optimizing the distribution of hidden embeddings. We experiment with two benchmark datasets of two-class semi-supervised segmentation, left atrium and pancreas, as well as a three-class multi-center dataset of type B aortic dissection. For all three datasets, UPCoL outperforms the state-of-the-art SSL methods, demonstrating the efficacy of the uncertainty-informed prototype learning strategy (Code is available at https://github.com/VivienLu/UPCoL).

Keywords: Semi-supervised learning · Uncertainty assessment · Prototype learning · Medical image segmentation

1 Introduction

The field of medical image segmentation has been increasingly drawn to semi-supervised learning (SSL) due to the great difficulty and cost of data labeling.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-43901-8_63.

H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14223, pp. 662–672, 2023.
https://doi.org/10.1007/978-3-031-43901-8_63

By utilizing both labeled and unlabeled data, SSL can significantly reduce the need for labeled training data and address inter-observer variability [1,8,14].

Typical SSL approaches involve techniques such as self-training, uncertainty estimation, and consistency regularization. Self-training aims to expand labeled training set by selecting the most confident predictions from the unlabeled data to augment the labeled data [23]. To obtain high-quality pseudo-labels, uncertainty estimation is often employed in self-training models. Various uncertainty estimation methods have been proposed to reduce the influence of ambiguous unlabeled data, e.g., Monte Carlo dropout [27] and ensemble-based methods [15,22]. Also, some metrics have been defined to quantify the degree of uncertainty. The most widely-used one is information entropy [18,27], where a threshold or a percentage is set to determine whether an unlabeled sample is reliable, i.e., its predicted label can be used as pseudo-label during the training phase. Besides pseudo-labeling, it is common to add a consistency regularization in loss function [5,11,28]. For instance, UA-MT [27] and CoraNet [15] impose consistency constraints on the teacher-student model for specific regions (certain/uncertain area or both). URPC [12] uses multilevel extraction of multi-scale uncertainty-corrected features to moderate the anomalous pixel of consistency loss.

Despite current progress, the performance of pseudo-labeling and consistency constraints has been limited for two reasons. First, defining an appropriate quantification criterion for reliability across various tasks can be challenging due to the inherent complexity of uncertainty. Second, most of the consistency constraints are imposed at decision space with the assumption that the decision boundary must be located at the low-density area, while the latent feature space of unlabeled data has not been fully exploited, and the low-density assumption may be incapable to guide model learning in the correct way.

Recently, prototype alignment has been introduced into SSL. Prototype-based methods have the potential of capturing underlying data structure including unlabeled information, and optimizing the distribution of feature embeddings across various categories [3,17]. Existing semi-supervised segmentation methods based on prototype learning aim to learn each class prototype from sample averaging and leverage consistency constraints to train the segmentation network. U^2PL [18] distinguishes reliable samples among unlabeled data by uncertainty estimation, and constructs prototypes for the whole dataset by class averaging its features with labeled sample and reliable unlabeled features. CISC-R [19] queries a guiding labeled image that shares similar semantic information with an unlabeled image, then estimates pixel-level similarity between unlabeled features and labeled prototypes, thereby rectifying the pseudo labels with reliable pixel-level precision. CPCL [24] introduces a cyclic prototype consistency learning framework to exploit unlabeled data and enhance the prototype representation.

Overall, prototype learning has much room for improvement in semi-supervised segmentation. As voxel-level averaging is only reliable for labeled data, current prototype learning approaches rely on labeled data and a small amount of unlabeled data, or learn prototypes separately for labeled and unlabeled data. In this way, they may not fully represent the distribution of the embedding space. Here raises the question:

Can we capture the embedding distribution by considering all voxels, including both labeled and unlabeled, and exploit the knowledge of the entire dataset?

To answer it, we propose to learn fused prototypes through uncertainty-based attention pooling. The fused prototypes represent the most representative and informative examples from both the labeled and unlabeled data for each class. The main contributions of our work can be summarized as follows:

1) We develop a novel uncertainty-informed prototype consistency learning framework, UPCoL, by considering voxel-level consistency in both latent feature space (i.e., prototype) and decision space.
2) Different from previous studies, we design a fused prototype learning scheme, which jointly learns from labeled and unlabeled data embeddings.
3) For stable prototype learning, we propose a new entropy measure to qualify the reliability of unlabeled voxel and an attention-weighted strategy for fusion.
4) We apply UPCoL to two-class and three-class segmentation tasks. UPCoL outperforms the SOTA SSL methods by large margins.

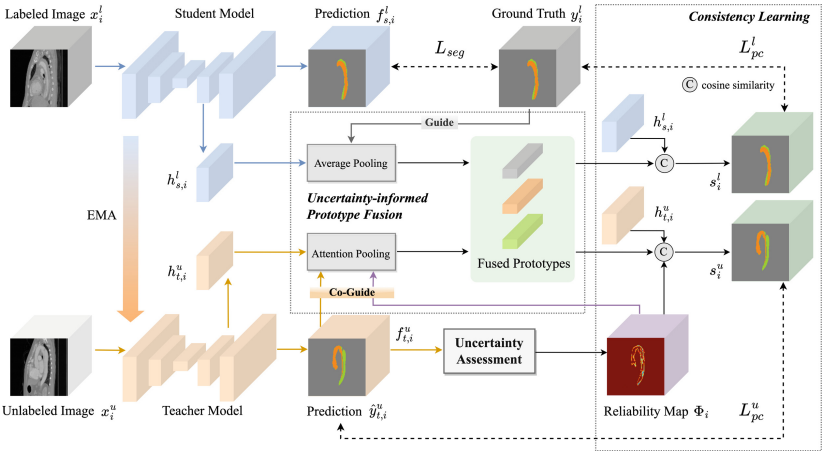

Fig. 1. Overview of UPCoL (using Aortic Dissection segmentation for illustration). The labeled images go through the student model for supervised learning, while the unlabeled images go through the teacher model for segmentation and uncertainty estimation. In Uncertainty-informed Prototype Fusion module, we utilize the reliability map to fuse prototypes learned from labeled and unlabeled embeddings. The similarity between fused prototypes and feature embeddings at each spatial location is then measured for consistency learning.

2 Methodology

Given a dataset $\mathcal{D} = \{\mathcal{D}^l, \mathcal{D}^u\}$, the labeled set $\mathcal{D}^l = \{x_i^l, y_i^l\}_{i=1}^N$ contains N samples, and the unlabeled set $\mathcal{D}^u = \{x_i^u\}_{i=N}^{N+M}$ contains M samples, where

$x_i^l, x_i^u \in \mathbb{R}^{H\times W\times D}$ represent the input with height H, width W, depth D, and $y_i^u \in \{0, 1, ..., C-1\}^{H\times W\times D}$. The proposed framework UPCoL includes a student model and a self-ensembling teacher model, each consisting of a representation head h and a segmentation head f. Figure 1 shows the overview.

2.1 Uncertainty Assessment

To assess the uncertainty at both voxel-level and geometry-level, we adopt the same ensemble of classifiers as [22] using different loss functions, i.e., cross-entropy loss, focal loss [9], Dice loss, and IoU loss. Unlike [22], which simply differentiates certain and uncertain regions based on the result discrepancy of four classifiers, we use the average of four prediction results and define an entropy-based measure to quantify the reliability for each voxel. Specifically, let $f_{t,i}^{(x,y,z)} \in \mathcal{R}^C$ denote the softmax probability for voxel at position (x, y, z) in i-th unlabeled image yielded by the segmentation head of the teacher model, where the segmentation result is the average over multiple classifiers (AMC), and C is the number of classes. The entropy is formulated in Eq. (1),

$$\mathcal{H}(f_{t,i}^{(x,y,z)}) = -\sum_{c=0}^{C-1} f_{t,i}^{(x,y,z)}(c) \log f_{t,i}^{(x,y,z)}(c). \tag{1}$$

Intuitively, voxels with high entropy are ambiguous. Thus, a reliability map can be defined accordingly, denoted by $\Phi_i^{(x,y,z)}$, which enables the model to assign varying degrees of importance to voxels,

$$\Phi_i^{(x,y,z)} = \frac{1}{H\times W\times D}(1 - \frac{\mathcal{H}(f_{t,i}^{(x,y,z)})}{\sum_{x,y,z}\mathcal{H}(f_{t,i}^{(x,y,z)})}). \tag{2}$$

2.2 Prototype Consistency Learning

Uncertainty-Informed Prototype Fusion. The prototypes from labeled and unlabeled data are first extracted separately. Both of them originate from the feature maps of the 3rd-layer decoder, which are upsampled to the same size as segmentation labels by trilinear interpolation. Let $h_{s,i}^l$ be the output feature by the representation head of the student model for the i-th labeled image, and $h_{t,i}^u$ be the hidden feature by the teacher representation head for the i-th unlabeled image. $\mathcal{B}_l$ and $\mathcal{B}_u$ denote the batch sizes of labeled and unlabeled set respectively, and (x, y, z) denotes voxel coordinate. For labeled prototype, the feature maps are masked directly using ground truth labels, and the prototype of class c is computed via masked average pooling [17,29]:

$$p_c^l = \frac{1}{\mathcal{B}_l}\sum_{i=1}^{\mathcal{B}_l}\frac{\sum_{x,y,z} h_{s,i}^{l(x,y,z)}\mathbb{1}\left[y_i^{l(x,y,z)} = c\right]}{\sum_{x,y,z}\mathbb{1}\left[y_i^{l(x,y,z)} = c\right]}. \tag{3}$$

For unlabeled data, instead of simply averaging features from the same predicted class, UPCoL obtains the prototypes in an uncertainty-informed manner, i.e., using a masked attention pooling based on each voxel's reliability:

$$p_c^u = \frac{1}{\mathcal{B}_u} \sum_{i=1}^{\mathcal{B}_u} \frac{\sum_{x,y,z} h_{t,i}^{u(x,y,z)} \Phi_i^{(x,y,z)} \mathbb{1}\left[\hat{y}_i^{u(x,y,z)} = c\right]}{\sum_{x,y,z} \mathbb{1}\left[\hat{y}_i^{u(x,y,z)} = c\right]}. \tag{4}$$

Temporal Ensembling technique in Mean-Teacher architecture enhances model performance and augments the predictive label quality [16], leading to a progressive improvement in the reliability of predictive labels and the refinement of unlabeled prototypes throughout the training process. Thus, we adopt a nonlinear updating strategy to adjust the proportion of unlabeled prototypes for fusing the labeled prototypes and unlabeled prototypes, i.e.,

$$p_c = \lambda_{lab} p_c^l + \lambda_{unlab} p_c^u, \tag{5}$$

where $\lambda_{lab} = \frac{1}{1+\lambda_{con}}$, $\lambda_{unlab} = \frac{\lambda_{con}}{1+\lambda_{con}}$, and λ_{con} is the widely-used time-dependent Gaussian warming up function [16]. During the training process, the proportion of labeled prototypes decreases from 1 to 0.5, while the proportion of unlabeled prototypes increases from 0 to 0.5. This adjustment strategy ensures that labeled prototypes remain the primary source of information during training, even as the model gradually gives more attention to the unlabeled prototypes.

Consistency Learning. We adopt non-parametric metric learning to obtain representative prototypes for each semantic class. The feature-to-prototype similarity is employed to approximate the probability of voxels in each class,

$$s_i^{(x,y,z)} = \mathrm{CosSim}\left(h_i^{(x,y,z)}, p_c\right) = \frac{h_i^{(x,y,z)} \cdot p_c}{\max(\left\|h_i^{(x,y,z)}\right\|_2 \cdot \|p_c\|_2, \epsilon)} \tag{6}$$

where the value of ϵ is fixed to $1e^{-8}$, and $\mathrm{CosSim}(\cdot)$ denotes cosine similarity. To ensure the accuarcy of the prototype, prototype-based predictions for labeled voxels expect to close to ground truth. And the prototype-based predictions for unlabeled voxels expect to close to segmentor prediction since prototype predictions are considered to be reliable aid. Then the prototype consistency losses for labeled and unlabeled samples are defined respectively as:

$$\mathcal{L}_{pc}^l = \mathcal{L}_{CE}(s_i^l, y_i^l), \qquad \mathcal{L}_{pc}^u = \sum_{x,y,z} \Phi_i^{(x,y,z)} \mathcal{L}_{CE}(s_i^{u(x,y,z)}, \hat{y}_i^{u(x,y,z)}), \tag{7}$$

where $\hat{y}_i^{u(x,y,z)}$ is the student model prediction of the i-th unlabeled sample at (x, y, z). Equation (7) is a variant of expanding training set by pseudo labels of the unlabeled data commonly adopted in SSL, with the difference that we use

the reliability-aware pseudo labels at the voxel level. Finally, the total loss of our UPCoL network is shown in Eq. (8),

$$\mathcal{L} = \mathcal{L}_{seg} + \mathcal{L}_{pc}^{l} + \lambda_{con}\mathcal{L}_{pc}^{u}. \tag{8}$$

3 Experimental Results

Datasets. We evaluate our approach on three datasets: the pancreas dataset (82 CTA scans), the left atrium dataset (100 MR images), and a multi-center dataset for type B aortic dissection (TBAD, 124 CTA scans). The pancreas and left atrium datasets are preprocessed following previous studies [7,10,21,22,27,28]. The TBAD dataset is well-annotated by experienced radiologists, with 100 scans for training and 24 for test, including both public data [25] and data collected by our team. The dataset was resampled to 1mm^3 and resized to $128 \times 128 \times 128$, in accordance with [2]. For all three datasets, we use only 20% of the training data with labels and normalize the voxel intensities to zero mean and unit variance.

Implementation Details. We adopt V-Net [13] as the backbone, and use the results of V-Nets trained with 20% and 100% labeled data as the lower and upper bounds, respectively. For the Mean-Teacher framework, the student network is trained for $10k$ iterations using Adam optimizer and learning rate 0.001, while the teacher is updated with exponential moving average (EMA) of the student's parameters. The batch size is 3, including 1 labeled and 2 unlabeled samples. Following [7,10,22], we randomly crop cubes of size $96 \times 96 \times 96$ and $112 \times 112 \times 80$ for the pancreas and left atrium datasets, respectively. In addition, we use 3-fold cross validation and apply data augmentation by rotation within the range $(-10°, 10°)$ and zoom factors within the range $(0.9, 1.1)$ for TBAD dataset during training, as proposed in [2,4]. Four performance metrics are adopted, i.e., Dice coefficient (Dice), Jaccard Index (Jac), 95% Hausdorff Distance (95HD), and Average Symmetric Surface Distance (ASD).

Results on the Left Atrium Dataset and Pancreas-CT Dataset. We compare UPCoL with nine SOTA SSL methods, including consistency-based (MT [16], MC-Net [21], ASE-Net [6], URPC [12]), uncertainty-based (UA-MT [27], MC-Net+ [20]), and divergence-based (DTC [10], SimCVD [26], CoraNet [15]). The results (Table 1) demonstrate a substantial performance gap between the lower and upper bounds due to the limited labeled data. Remarkably, our proposed framework outperforms the theoretical upper bound (the second row) in terms of Dice, Jac, 95HD on the left atrium dataset, and 95HD, ASD on the pancreas dataset, suggesting that the unlabeled knowledge extracted from deep features is reliable and well complements the information that is not captured in the fully supervised prediction phase.

Results on the Aortic Dissection Dataset. We compare with two SOTA SSL methods, the uncertainty-based method FUSSNet [22] and embedding-based method URPC [12], as well as two common SSL approaches, MT [16] and UA-MT [27]. As shown in Table 2, the proposed UPCoL obtains the best segmentation results and outperforms fully-supervised V-Net over 6% on Dice score, but

Table 1. Performance comparison on the Left Atrium and Pancreas datasets.

Method	Left Atrium						Pancreas-CT					
	Lb	Unlb	Dice	Jac	95HD	ASD	Lb	Unlb	Dice	Jac	95HD	ASD
V-Net	16	0	84.41	73.54	19.94	5.32	12	0	70.63	56.72	22.54	6.29
V-Net	80	0	91.42	84.27	5.15	1.50	62	0	82.60	70.81	5.61	1.33
MT [16] (NIPS'17)	16	64	86.00	76.27	9.75	2.80	12	50	75.85	61.98	12.59	3.40
UA-MT [27] (MICCAI'19)	16	64	88.88	80.21	7.32	2.26	12	50	77.26	63.28	11.90	3.06
DTC [10] (AAAI'21)	16	64	89.42	80.98	7.32	2.10	12	50	76.27	62.82	8.70	2.20
MC-Net [21] (MICCAI'21)	16	64	90.34	82.48	6.00	1.77	12	50	78.17	65.22	6.90	1.55
MC-Net+ [20] (MIA'22)	16	64	91.07	83.67	5.84	1.67	12	50	80.59	68.08	6.47	1.74
ASE-Net [6] (TMI'22)	16	64	90.29	82.76	7.18	1.64	-	-	-	-	-	-
CoraNet [15] (TMI'21)	-	-	-	-	-	-	12	50	79.49	67.10	11.10	3.06
SimCVD [26] (TMI'22)	16	64	90.85	83.80	6.03	1.86	-	-	-	-	-	-
URPC [12] (MIA'22)	-	-	-	-	-	-	12	50	80.02	67.30	8.51	1.98
UPCoL(Ours)	16	64	**91.69**	**84.69**	**4.87**	**1.56**	12	50	**81.78**	**69.66**	**3.78**	**0.63**

'Lb' and 'Unlb' denote the number of labeled samples and unlabeled samples, respectively.

Table 2. Performance comparison on the Aortic Dissection dataset.

Method	Lb	Unlb	Dice(%) ↑			Jaccard(%) ↑			95HD(voxel) ↓			ASD(voxel) ↓		
			TL	FL	Mean	TL	FL	Mean	TL	FL	Mean	TL	FL	Mean
V-Net	20	0	55.51	48.98	52.25	39.81	34.79	37.30	7.24	10.17	8.71	1.27	3.19	2.23
V-Net	100	0	75.98	64.02	70.00	61.89	50.05	55.97	3.16	7.56	5.36	0.48	2.44	1.46
MT [16]	20	80	57.62	49.95	53.78	41.57	35.52	38.54	6.00	8.98	7.49	0.97	2.77	1.87
UA-MT [27]	20	80	70.91	60.66	65.78	56.15	46.24	51.20	4.44	7.94	6.19	0.83	2.37	1.60
FUSSNet [22]	20	80	79.73	65.32	72.53	67.31	51.74	59.52	3.46	7.87	5.67	0.61	2.93	1.77
URPC [12]	20	80	81.84	69.15	75.50	70.35	57.00	63.68	4.41	9.13	6.77	0.93	**1.11**	**1.02**
UPCoL	20	80	**82.65**	**69.74**	**76.19**	**71.49**	**57.42**	**64.45**	**2.82**	**6.81**	**4.82**	**0.43**	2.22	1.33

only requires 20% labels. The accuracy of FUSSNet [22], URPC [12], and UPCoL surpassing the upper bound demonstrates the effectiveness of uncertainty and embedding-based approaches in exploiting the latent information of the data, particularly in challenging classification tasks. We further visualize the segmentation results on test data of different methods in Fig. 2. As can be seen, UPCoL achieves superior segmentation performance with fewer false positives and superior capability in capturing intricate geometric features, such as the vessel walls between True Lumen (TL) and False Lumen (FL), and effectively smoothing out rough portions of the manual annotation.

Ablation Study. Here we investigate the contribution of key components, including the mean-teacher architecture (MT), the average multi-classifier (AMC) (to yield segmentation results), and the prototype learning (PL) strategy. As shown in Table 3, the MT model, which enforces a consistency cost between the predictions of student model and teacher model, outperforms vanilla V-Net by a large margin (over 5% on Dice), and AMC can also enhance the MT's performance (over 2% on Dice). Compared to the consistency cost in original MT

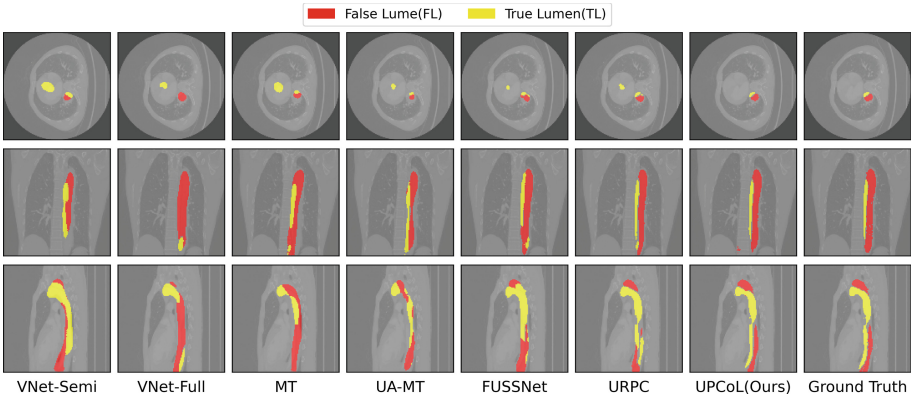

Fig. 2. Visualization of segmentation results on Type B Aortic Dissection dataset.

model [16] (used by MT and MT+AMC), the prototype consistency leads to better performance. Especially, we compare different prototype learning strategies in three methods, MT+PL, CPCL*, and UPCoL, which have the same backbone and AMC module. MT+PL performs prototype learning only for labeled data, CPCL* learns prototypes for labeled and unlabeled data separately using the same strategy proposed in [24], and UPCoL learns fused prototypes. As can be seen, prototype learning using unlabeled data is beneficial for performance improvement, but it requires a well-designed mechanism. Here CPCL* is slightly worse than using only labeled data for prototype learning, potentially due to the isolation of updating labeled and unlabeled prototypes, which may hinder their interaction and prevent the full utilization of knowledge. This highlights the importance of fusing labeled and unlabeled prototypes. UPCoL, on the other hand, successfully merges labeled and unlabeled prototypes through the use of reliability maps, resulting in SOTA performance. This demonstrates the effectiveness of uncertainty-based reliability assessment and prototype fusion in

Table 3. Ablation study results on the pancreas dataset.

Method		Component indication				Metric			
		MT	PL			Dice (%)	Jaccard (%)	95HD (voxel)	ASD (voxel)
			Lb	Unlb	Fuse				
W/o Consistency						70.63	56.72	22.54	6.29
Prediction Consistency	MT	✓				75.85	61.98	12.59	3.40
	MT + AMC	✓				77.97	64.46	9.80	2.52
Prototype Consistency	MT + PL	✓	✓			80.50	68.11	4.49	0.74
	CPCL*	✓	✓	✓		80.08	67.29	7.68	2.08
	UPCoL	✓	✓	✓	✓	**81.78**	**69.66**	**3.78**	**0.63**

'Lb', 'Unlb' and 'Fuse' denote the labeled, unlabeled, and fused prototypes, respectively.

fully leveraging both labeled and unlabeled information. In the Supplementary Materials, visualized results showcase improved predicted labels and unlabeled prototypes as training progresses.

4 Conclusion

This paper presents a novel framework, UPCoL, for semi-supervised segmentation that effectively addresses the issue of label sparsity through uncertainty-based prototype consistency learning. To better utilize unlabeled data, UPCoL employs a quantitative uncertainty measure at the voxel level to assign degrees of attention. UPCoL achieves a careful and effective fusion of unlabeled data with labeled data in the prototype learning process, which leads to exceptional performance on both 2-class and 3-class medical image segmentation tasks. As future work, a possible extension is to allow multiple prototypes for a class with diversified semantic concepts, and a memory-bank-like mechanism could be introduced to learn prototypes from large sample pools more efficiently.

Acknowledgements. This work was supported by the National Natural Science Foundation of China (Nos. 61972251 and 62272300).

References

1. Bai, Wenjia, et al.: Semi-supervised learning for network-based cardiac MR image segmentation. In: Descoteaux, Maxime, Maier-Hein, Lena, Franz, Alfred, Jannin, Pierre, Collins, D. Louis., Duchesne, Simon (eds.) MICCAI 2017. LNCS, vol. 10434, pp. 253–260. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66185-8_29
2. Cao, L., et al.: Fully automatic segmentation of type b aortic dissection from CTA images enabled by deep learning. Europ. J. Radiol. **121**, 108713 (2019)
3. Dong, N., Xing, E.P.: Few-shot semantic segmentation with prototype learning. In: BMVC, vol. 3 (2018)
4. Fantazzini, A., et al.: 3d automatic segmentation of aortic computed tomography angiography combining multi-view 2d convolutional neural networks. Cardiovascular Eng. Technol. **11**, 576–586 (2020)
5. Hang, Wenlong, et al.: Local and global structure-aware entropy regularized mean teacher model for 3D left atrium segmentation. In: Martel, Anne L.., Abolmaesumi, Purang, Stoyanov, Danail, Mateus, Diana, Zuluaga, Maria A.., Zhou, S. Kevin., Racoceanu, Daniel, Joskowicz, Leo (eds.) MICCAI 2020. LNCS, vol. 12261, pp. 562–571. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59710-8_55
6. Lei, T., Zhang, D., Du, X., Wang, X., Wan, Y., Nandi, A.K.: Semi-supervised medical image segmentation using adversarial consistency learning and dynamic convolution network. IEEE Trans. Med. Imaging (2022)
7. Li, Shuailin, Zhang, Chuyu, He, Xuming: Shape-aware semi-supervised 3d semantic segmentation for medical images. In: Martel, Anne L.., Abolmaesumi, Purang, Stoyanov, Danail, Mateus, Diana, Zuluaga, Maria A.., Zhou, S. Kevin., Racoceanu, Daniel, Joskowicz, Leo (eds.) MICCAI 2020. LNCS, vol. 12261, pp. 552–561. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59710-8_54

8. Li, X., Yu, L., Chen, H., Fu, C.W., Xing, L., Heng, P.A.: Transformation-consistent self-ensembling model for semisupervised medical image segmentation. IEEE Trans. Neural Networks Learn. Syst. **32**(2), 523–534 (2020)
9. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2980–2988 (2017)
10. Luo, X., Chen, J., Song, T., Wang, G.: Semi-supervised medical image segmentation through dual-task consistency. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 35, pp. 8801–8809 (2021)
11. Luo, X., Liao, W., Chen, J., Song, T., Chen, Y., Zhang, S., Chen, N., Wang, G., Zhang, S.: Efficient semi-supervised gross target volume of nasopharyngeal carcinoma segmentation via uncertainty rectified pyramid consistency. In: MICCAI 2021. pp. 318–329. Springer (2021)
12. Luo, X., Wang, G., Liao, W., Chen, J., Song, T., Chen, Y., Zhang, S., Metaxas, D.N., Zhang, S.: Semi-supervised medical image segmentation via uncertainty rectified pyramid consistency. Medical Image Analysis **80**, 102517 (2022)
13. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 2016 fourth international conference on 3D vision (3DV). pp. 565–571. Ieee (2016)
14. Nie, D., Gao, Y., Wang, L., Shen, D.: Asdnet: attention based semi-supervised deep networks for medical image segmentation. In: MICCAI 2018. pp. 370–378. Springer (2018)
15. Shi, Y., Zhang, J., Ling, T., Lu, J., Zheng, Y., Yu, Q., Qi, L., Gao, Y.: Inconsistency-aware uncertainty estimation for semi-supervised medical image segmentation. IEEE transactions on medical imaging **41**(3), 608–620 (2021)
16. Tarvainen, A., Valpola, H.: Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. Advances in neural information processing systems 30 (2017)
17. 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)
18. Wang, Y., Wang, H., Shen, Y., Fei, J., Li, W., Jin, G., Wu, L., Zhao, R., Le, X.: Semi-supervised semantic segmentation using unreliable pseudo-labels. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4248–4257 (2022)
19. Wu, L., Fang, L., He, X., He, M., Ma, J., Zhong, Z.: Querying labeled for unlabeled: Cross-image semantic consistency guided semi-supervised semantic segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence (2023)
20. Wu, Y., Ge, Z., Zhang, D., Xu, M., Zhang, L., Xia, Y., Cai, J.: Mutual consistency learning for semi-supervised medical image segmentation. Medical Image Analysis **81**, 102530 (2022)
21. Wu, Y., Xu, M., Ge, Z., Cai, J., Zhang, L.: Semi-supervised left atrium segmentation with mutual consistency training. In: MICCAI 2021. pp. 297–306. Springer (2021)
22. Xiang, J., Qiu, P., Yang, Y.: Fussnet: Fusing two sources of uncertainty for semi-supervised medical image segmentation. In: MICCAI 2022. pp. 481–491. Springer (2022)
23. Xie, Q., Luong, M.T., Hovy, E., Le, Q.V.: Self-training with noisy student improves imagenet classification. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 10687–10698 (2020)

24. Xu, Z., Wang, Y., Lu, D., Yu, L., Yan, J., Luo, J., Ma, K., Zheng, Y., Tong, R.K.y.: All-around real label supervision: Cyclic prototype consistency learning for semi-supervised medical image segmentation. IEEE Journal of Biomedical and Health Informatics 26(7), 3174–3184 (2022)
25. Yao, Z., Xie, W., Zhang, J., Dong, Y., Qiu, H., Yuan, H., Jia, Q., Wang, T., Shi, Y., Zhuang, J., et al.: Imagetbad: A 3d computed tomography angiography image dataset for automatic segmentation of type-b aortic dissection. Frontiers in Physiology p. 1611 (2021)
26. You, C., Zhou, Y., Zhao, R., Staib, L., Duncan, J.S.: Simcvd: Simple contrastive voxel-wise representation distillation for semi-supervised medical image segmentation. IEEE Transactions on Medical Imaging **41**(9), 2228–2237 (2022)
27. Yu, L., Wang, S., Li, X., Fu, C.W., Heng, P.A.: Uncertainty-aware self-ensembling model for semi-supervised 3d left atrium segmentation. In: MICCAI 2019. pp. 605–613. Springer (2019)
28. Zeng, X., Huang, R., Zhong, Y., Sun, D., Han, C., Lin, D., Ni, D., Wang, Y.: Reciprocal learning for semi-supervised segmentation. In: MICCAI 2021. pp. 352–361. Springer (2021)
29. Zhang, X., Wei, Y., Yang, Y., Huang, T.S.: Sg-one: Similarity guidance network for one-shot semantic segmentation. IEEE transactions on cybernetics **50**(9), 3855–3865 (2020)