Semi-supervised Class Imbalanced Deep Learning for Cardiac MRI Segmentation

Yuchen Yuan[1], Xi Wang[1(✉)], Xikai Yang[1], Ruijiang Li[2], and Pheng-Ann Heng[1,3]

[1] Department of Computer Science and Engineering, The Chinese University of Hong Kong, Shatin, Hong Kong
xiwang@cse.cuhk.edu.hk

[2] Department of Radiation Oncology, Stanford University School of Medicine, Stanford, USA

[3] Institute of Medical Intelligence and XR, The Chinese University of Hong Kong, Shatin, Hong Kong

Abstract. Despite great progress in semi-supervised learning (SSL) that leverages unlabeled data to improve the performance over fully supervised models, existing SSL approaches still fail to exhibit good results when faced with a severe class imbalance problem in medical image segmentation. In this work, we propose a novel Mean-teacher based class imbalanced learning framework for cardiac magnetic resonance imaging (MRI) segmentation, which can effectively conquer the problems of class imbalance and limited labeled data simultaneously. Specifically, in parallel to the traditional linear-based classifier, we additionally train a prototype-based classifier that makes dense predictions by matching test samples with a set of prototypes. The prototypes are iteratively updated by in-class features encoded in the entire sample set, which can better guide the model training by alleviating the class-wise bias exhibited in each individual sample. To reduce the noises in the pseudo labels, we propose a cascaded refining strategy by utilizing two multi-level tree filters that are built upon pairwise pixel similarity in terms of intensity values and semantic features. With the assistance of pixel affinities, soft pseudo labels are properly refined on-the-fly. Upon evaluation on ACDC and MMWHS, two cardiac MRI datasets with prominent class imbalance problem, the proposed method demonstrates the superiority compared to several state-of-the-art methods, especially in the case where few annotations are available (Code is available in https://github.com/IsYuchenYuan/SSCI).

Keywords: Semi-supervised learning · Medical image segmentation · Class imbalance · Prototype learning · Multi-level tree filters

1 Introduction

Recent progress in deep learning techniques, particularly convolutional neural networks (CNN), has shown enormous success in the field of 3D medical image

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-43901-8_44.

H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14223, pp. 459–469, 2023.
https://doi.org/10.1007/978-3-031-43901-8_44

segmentation [1]. However, these marvelous achievements are always accompanied by a high cost of substantial high-quality annotations, which are usually prohibitively difficult to obtain in the medical domain due to expensive time costs and highly-demanded expertise. Besides, class imbalance, caused by the huge variation of anatomical structure volumes, is another frequently occurring problem, further posing great challenges in automated and accurate 3D medical image segmentation with limited annotations.

To handle the label scarcity, a large bunch of semi-supervised learning (SSL) methods have been proposed by leveraging abundant unlabeled data to compensate for the annotation scarcity. Concretely, pseudo-labeling [2,3] and consistency regularization [4,5] are two effective SSL schemes. [6]. For example, Bai et al. [7] used the model trained on the labeled pairs to generate pseudo labels that serve as targets for the unlabeled data. To guarantee the quality of pseudo labels, Sedai et al. [8] used Bayesian deep learning [9] to measure the pixel-wise reliability of soft labels and suppressed regions with unreliable ones. Concurrently, Yu et al. designed a 3D uncertainty-aware semi-supervised MRI segmentation framework [10], where the uncertainty mechanism [9] is integrated with consistency regularization.

Despite their success, few of the SSL segmentation methods considered class imbalance that is naturally inherent in medical images. For example, in cardiac MRI images, *myocardium* is often missing or too small to be detected in apical slices [11], while other structures (e.g., the left ventricle blood cavity) usually have a large span in the whole volume. There are two possible hazards resulting from such severe class imbalance when annotations are limited. One is that the model can easily become biased towards the dominant classes. The class-wise bias substantially affects model convergence during training and the generalization on the test domain [11]. The other is that inadequate learning from the tail classes could introduce more noises to pseudo labels, which will dramatically accumulate during the training phase and destroy the model training consequently.

To solve the problems above, we propose a novel semi-supervised class imbalanced deep learning approach on top of the Mean-teacher framework [12] for MRI segmentation. Firstly, an extra prototype-based classifier is introduced into the student model in parallel to the traditional linear-based classifier. With prototype learning, each in-class features encoded in the entire training set can be leveraged as the prior knowledge to assist the learning of each individual, conducive to accurate segmentation for tail classes. Besides, two multi-level tree filters modeling the pairwise pixel similarity in terms of intensity and semantic features are employed to refine pseudo labels. Through a comprehensive evaluation on two 3D cardiac MRI datasets with different experimental settings, our method achieves state-of-the-art results. Notably, the outstanding superiority of our method is exhibited (2.2% and 10% improvement) in the case where only quite a few labels are available (1.25% for ACDC and 10% for MMWHS).

2 Methods

Problem Setting. Given limited labeled data $\mathcal{D}_L = \{X_i, Y_i\}_{i=1}^{L}$ and unlabeled data $\mathcal{D}_U = \{X_i\}_{i=L+1}^{L+U}$, $X_i \in \mathbb{R}^{H \times W \times D}$ represents the original image and $Y_i =$

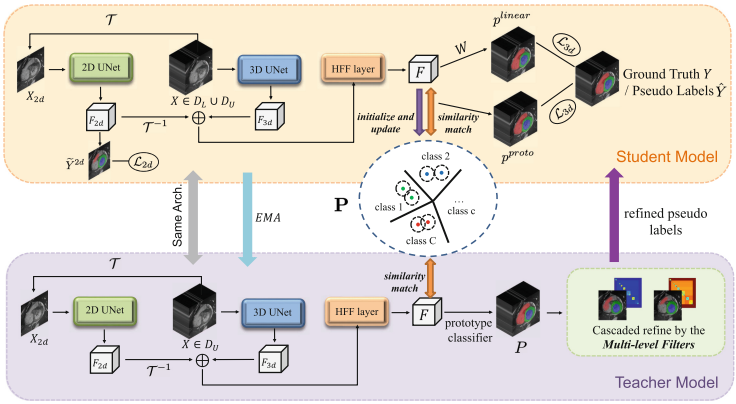

Fig. 1. Overview of the proposed semi-supervised class imbalance learning framework.

$\{0, 1, ..., C\}^{H \times W \times D}$ is the corresponding ground truth of C classes of objects in total. The given data are severely class imbalanced.

Overview. The overview of the proposed method is shown in Fig. 1. Our approach is built on top of the Mean-teacher [12] framework and adopts hybrid UNet (H-UNet) as backbone. Following [13], the networks learn through the student branch only, and the teacher model is update by using an Exponential Moving Average (EMA). The student model is trained with labeled pairs and pseudo labels generated by the teacher model. We train a linear- and a prototype-based classifier in parallel to alleviate class imbalance. Moreover, to guarantee the label quality, two multi-level tree filters (TFs) constructed based on the pixel affinity in terms of intensity and semantic features are employed to refine the teacher model's predictions from the prototype-based classifier.

2.1 H-UNet for 3D Medical Image Segmentation

The backbone model H-UNet is a variant of H-DenseUNet [14]. Given n samples in a training batch $X \in \mathbb{R}^{n \times 1 \times H \times W \times D}$, the 3D UNet produces features $F_{3d} \in \mathbb{R}^{n \times m \times H \times W \times D}$. Meanwhile, the 2D slices $X_{2d} \in \mathbb{R}^{nD \times 1 \times H \times W}$ obtained by performing the transformation operation $\mathcal{T}$ [14] on the original inputs are fed into the 2D UNet to generate features $F_{2d} \in \mathbb{R}^{nD \times m \times H \times W}$. By conducting inverse transformation operation, $\mathcal{T}^{-1}$, F_{2d} is aligned with F_{3d} before the two features are added and fed into the hybrid feature fusion (HFF) layer to yield the hybrid features F. We apply supervision on both 2D UNet and 3D UNet to train the H-UNet.

2.2 Class Imbalance Alleviation via Prototype-Based Classifier

Prototype-based classifier, leveraging prototypes instead of a parameterized predictor to make predictions, presents efficacy for semantic segmentation [15] and

class imbalance alleviation [16]. Therefore, we introduce an extra prototype-based classifier (PC) into the student model in parallel to a linear-based classifier (LC). Specifically, PC makes dense predictions by matching the normalized hybrid feature F with the nearest prototypes. Each prototype is the feature aggregation of several training pixels belonging to the same class. We denote $\mathbf{P} = \{\boldsymbol{p}_{c,k}\}_{c,k=1}^{C,K}$ as a set of prototypes associated with C classes for segmentation. Note that each class may have more than one prototype (i.e., K) due to the intra-class heterogeneity exhibited in medical images [17]. Given X_i, PC produces the probability distribution of pixel a over the C classes:

$$p^{proto}(Y_i^{'}[a] = c|X_i) = \frac{exp(s_{a,c})}{\sum_{c'}^{C} exp(s_{a,c'})}, \text{with } s_{a,c} = max\{\mathcal{S}(\|F_i[a]\|_2, \boldsymbol{p}_{c,k})\}_{k=1}^{K}, \tag{1}$$

where $s_{a,c} \in [-1, 1]$ denotes the *pixel-class* similarity between pixel a and its closest prototype of class c. $\mathcal{S}(,)$ is the similarity measure (i.e., *cosine similarity*). $F_i[a]$ denotes the extracted features of pixel a, and $Y_i^{'}[a]$ denotes the predicted probability. $\|\cdot\|_2$ stands for the ℓ_2 normalization.

Meanwhile, the features F are also fed into LC parameterized by $W = [\mathbf{w}_1, \cdots, \mathbf{w}_C] \in \mathbb{R}^{C \times m}$; $\mathbf{w}_c \in \mathbb{R}^m$ is a learnable projection vector for class c. The probability distribution of pixel a estimated by LC is defined as:

$$p^{linear}(Y_i^{'}[a] = c|X_i) = \frac{exp(\mathbf{w}_c^T F_i[a])}{\sum_{c'=1}^{C} exp(\mathbf{w}_{c'}^T F_i[a])}. \tag{2}$$

These two classifiers complete each other during training. In the early training phase, LC dominates knowledge learning and provides PC with discriminative features to initialize and update the prototypes (See Sect. 2.4). With the addition of PC, the feature embedding space is further regularized, along with intra-class features being more compact and inter-class features being more separated, which in turn benefits the learning of LC.

2.3 Pseudo Label Refinement via Multi-level Tree Filters

Rather than directly using the teacher's high-confidence prediction to generate pseudo labels [13], we propose a cascaded refining strategy (CRS) to improve the label quality in a slice-by-slice manner on 3D volumes by using TFs [18], which proceeds with the following three steps as depicted in Fig. 2.

Graph Construction: First, we respectively represent the topology of the low-level original unlabeled image and the high-level hybrid features as two 4-connected planar graphs: $\mathcal{G}^* = \{\mathcal{V}^*, \mathcal{E}^*\}$ where $\mathcal{V}^*$ is the vertex set associated with each pixel and $\mathcal{E}^*$ is the edge set, and $* \in \{low, high\}$. The weight of the edge connecting two adjacent nodes a and b indicates their dissimilarity, which is defined by:

$$w_{a,b}^{low} = w_{b,a}^{low} = d(I[a], I[b]), \quad w_{a,b}^{high} = w_{b,a}^{high} = d(F[a], F[b]), \tag{3}$$

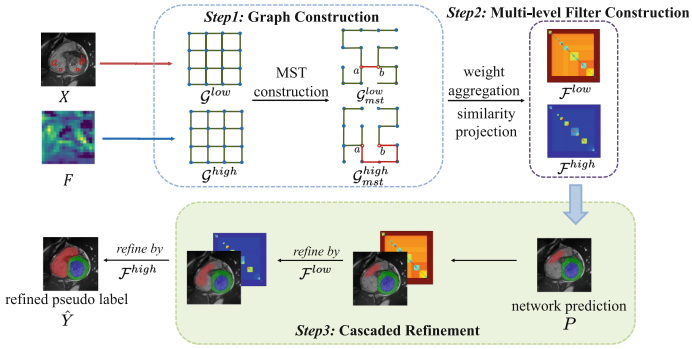

Fig. 2. The illustration of the cascaded refining strategy by using multi-level tree filters.

where $I[a] \in \mathbb{R}$ is the intensity of pixel a and $F[a] \in \mathbb{R}^m$ is the hybrid features. $d(\cdot,\cdot)$ denotes the squared Euclidean distance. Then, an edge pruning strategy is performed on $\mathcal{G}^{low}$ and $\mathcal{G}^{high}$ by sequentially removing the edges with substantial dissimilarity (i.e., larger distance) from $\mathcal{E}^*$ to generate two *minimum spanning trees* (MSTs) [19], $\mathcal{G}^{low}_{mst}$ and $\mathcal{G}^{high}_{mst}$, in which vertices are preferentially connected with similar ones.

Multi-level Filter Construction: Based on the two MSTs, we build the low-level TF $\mathcal{F}^{low}$ and the high-level TF $\mathcal{F}^{high}$. The filter weight $\mathcal{F}^*_{a,b}$ of any two nodes is obtained by aggregating the MST edge weights along the path between the two nodes [20]:

$$\mathcal{F}^*_{a,b} = \frac{1}{z_a} S_{\mathcal{G}^*_{mst}}(\mathbb{E}^*_{a,b}), \quad \text{where} \quad S_{\mathcal{G}^*_{mst}}(\mathbb{E}^*_{a,b}) = exp(-\sum_{\forall (q,o)\in \mathbb{E}^*_{a,b}} w^*_{q,o}). \tag{4}$$

Here, $\mathbb{E}^*_{a,b}$ is the edge set in the path from node a to node b. $S_{\mathcal{G}^*_{mst}}(\cdot)$ maps the distance of two vertices into a positive scalar which measures the pixel affinity. z_a is a normalization factor, which is the summation of the similarity between node a and all other nodes in the MST.

Cascaded Refinement: Lastly, we refine the teacher's prediction P with the two filters in a cascade manner to acquire high-quality pseudo labels $\hat{Y}$:

$$\hat{Y} = \mathsf{R}\big(\mathcal{F}^{high}, \mathsf{R}(\mathcal{F}^{low}, P)\big), \quad \text{with} \quad \mathsf{R}(\mathcal{F}^*, P_a) = \sum_{\forall b \in \mathcal{V}^*} \mathcal{F}^*_{a,b} P_b, \tag{5}$$

where $\mathsf{R}(\cdot,\cdot)$ is the refinement process where each unlabeled pixel can contribute to the refinement for other pixels with a contribution proportional to their similarity. By exploiting the multi-level complementary features, i.e., the object boundary information encoded in $\mathcal{F}^{low}$ and the semantic similarity encoded in $\mathcal{F}^{high}$, CRS shows superiority when a single TF fails in some cases. E.g., when two pixels of different classes have similar intensity values (pixel a and b in Fig. 2),

$\mathcal{F}^{low}$ will model them as affinity pairs, which is not expected. Fortunately, $\mathcal{F}^{high}$ can suppress the mutual interference between them according to their distinct semantic features, thus ensuring refinement efficacy (See Supplementary Fig. 1).

2.4 Network Training and Prototype Update

The framework is trained by utilizing both the labeled data D_L and the unlabeled data D_U. For the labeled data, the supervised loss is defined as:

$$\mathcal{L}^l = \frac{1}{L}\sum_{i=1}^{L}\left(l(\widetilde{Y}_i^{2D}, Y_i^{2D}) + l(p^{proto}(Y_i^{'}|X_i), Y_i) + l(p^{linear}(Y_i^{'}|X_i), Y_i)\right), \quad (6)$$

where $\widetilde{Y}^{2D}$ is the prediction generated by the 2D UNet in H-UNet, and Y^{2D} is the transformation of the 3D ground truth Y by conducting $\mathcal{T}$ (See Sect. 2.1). l is the weighted sum of cross entropy loss and dice loss.

For the unlabeled data, the student model is trained with the pseudo labels $\hat{Y}$ refined by the proposed CRS. The unsupervised loss is defined as:

$$\mathcal{L}^u = \frac{1}{U}\sum_{i=L+1}^{L+U}\left(l(\widetilde{Y}_i^{2D}, \hat{Y}_i^{2D}) + l(p^{proto}(Y_i^{'}|X_i), \hat{Y}_i) + l(p^{linear}(Y_i^{'}|X_i), \hat{Y}_i)\right). \quad (7)$$

Following [10], we use a time-dependent Gaussian warming up function to control the balance between the supervised and unsupervised loss.

Prototype Initialize and Update: The initialization of prototypes determines the PC performance. To get better prototypes, we first pretrain the network with the LC solely. Then, we collect pixelwise deep features and use K-means [21] to generate K subclusters for each class. Finally, initial prototypes of each class are obtained by averaging the features in each subcluster. Following [15], after each training iteration, we first conduct online clustering by few steps of Sinkhorn-Knopp iteration [22] and let the prototypes evolve continuously with the clustering features of both the labeled and unlabeled samples:

$$\boldsymbol{p}_{c,k} \leftarrow \alpha \boldsymbol{p}_{c,k} + (1-\alpha)\left\|F_{c,k}\right\|_2, \quad (8)$$

where α is a momentum coefficient, i.e., 0.999. $\left\|F_{c,k}\right\|_2$ are the normalized features of pixels which belong to the c-th class and are closest to the k-th subcluster. Since the prototypes are iteratively updated by the in-class features encoded in the entire sample set, such prior knowledge can alleviate the class-wise bias exhibited in each individual sample, which can better guide the model training.

3 Experiments and Results

3.1 Datasets and Implementation Details

Datasets and Evaluation Metrics: The model is evaluated on two public cardiac MRI datasets: (1) The ACDC[1] [23] contains 100 patients' scans, with expert annotations for 3 structures: left ventricle (LV), myocardium (MYO), and right ventricle (RV); (2) The MMWHS[2] [24] consists of 20 3D cardiac MRIs with annotations for 7 structures: LV, RV, MYO, left atrium (LA), right atrium (RA), ascending aorta (AA), and pulmonary artery (PA). We adopt Dice Similarity Score (DSC) to evaluate model performance.

Pre-processing: Ahead of training, we removed the top 2% of the highest intensity value to reduce artifact impact for MMWHS dataset, followed by z-score normalization. As ACDC has a low through-plane resolution of 5–10 mm, we resample all 2D slices to a fixed in-plane resolution of 1×1 mm with a size of 256×256 as the training samples. For MMWHS, we randomly crop 3D cubes with a size of $80 \times 80 \times 80$ for training.

Implementation Details: We split the datasets into the training and validation set randomly at a ratio of 4:1 [11]. To assess the model performance on different label percentages, we respectively take 1.25%, 2.5%, and 10% data from the ACDC training set and 10%, 20%, and 40% data from the MMWHS training set as the labeled data and the rest as unlabeled. The training details for the two datasets are provided in Section A of the Supplementary material.

3.2 Comparison with State-of-the-Arts

Table 1. Comparison with state-of-art methods on the ACDC and MMWHS datasets

Methods	Avg DSC (ACDC)			Avg DSC (MMWHS)		
	L = 1.25%	L = 2.5%	L = 10%	L = 10%	L = 20%	L = 40%
Self train [7]	0.717	0.856	0.874	0.529	0.630	0.792
Data Aug [25]	0.657	0.833	0.865	0.513	0.679	0.791
Mixmatch [26]	0.616	0.812	0.852	0.550	0.689	0.792
Global+Local CL [27]	0.739	0.800	0.885	0.561	0.687	0.779
Class-wise Sampling [11]	0.746	0.842	0.889	0.626	**0.791**	0.815
Proposed	**0.768**	**0.869**	**0.907**	**0.732**	0.789	**0.840**

We compare our method with several state-of-the-art frameworks for semi-supervised medical image segmentation. The results are reported in Table 1.

[1] https://www.creatis.insa-lyon.fr/Challenge/acdc/databases.html.
[2] https://zmiclab.github.io/zxh/0/mmwhs/.

Among these, the proposed method outperforms competing methods across all datasets and label percentages, except for a slightly lower DSC than method Class-wise Sampling [11] by using 20% labeled MMWHS data. However, method [11] fails to produce satisfactory results using very few annotations (L = 1.25% in ACDC and L = 10% in MMWHS). In contrast, our method exhibits quite consistent performance on different label percentages and has substantial improvement when only few annotations are available, with more than 10% increase by using 10% labeled MMWHS. This indicates the benefit of fully leveraging unlabeled data via the proposed PC and CRS, in which every pixel in the unlabeled data can properly contribute to the model training and the refinement for pseudo labels. Methods like Data Aug [25] and contrastive learning of global and local features [27] do not explicitly address the class imbalance problem, thus they have inferior performance on both datasets. Also, these two methods both require a complex pretraining process, which is inefficient and time-consuming. Mixmatch [26], mixing labeled and unlabeled data using MixUp in SSL, cannot work very well for the datasets with a severe class imbalance problem, mainly because simply mixing two images could oppositely increase difficulty in the less-dominant classes learning.

3.3 Ablation Studies

Table 2. Ablation studies on the efficacy of different components in the proposed method. The mean and standard deviation of the testing results are reported.

Methods		LC	PC	CRS	Avg DSC	Dice of heart substructures ↑						
						MYO	LA	LV	RA	RV	AA	PA
3D UNet		✓			0.770	0.756(.08)	0.662(.08)	0.896(.02)	0.740(.13)	0.854(.04)	0.743(.06)	0.738(.02)
		✓	✓	✓	0.800	0.793(.04)	0.751(.02)	0.909(.02)	0.798(.06)	0.837(.09)	0.789(.06)	0.725(.04)
H-UNet	①	✓			0.782	0.757(.06)	0.758(.02)	0.897(.03)	0.730(.08)	0.805(.07)	0.775(.08)	0.756(.03)
	②	✓		✓	0.799	0.776(.04)	0.763(.01)	0.893(.02)	0.809(.07)	0.786(.05)	0.789(.08)	**0.775(.01)**
	③		✓		0.797	0.762(.06)	0.750(.02)	0.909(.02)	0.789(.08)	0.835(.04)	0.786(.05)	0.751(.03)
	④	✓	✓		0.825	0.800(.03)	**0.800(.02)**	0.910(.02)	0.843(.04)	0.864(.04)	0.789(.09)	0.766(.03)
	⑤	✓	✓	✓	**0.840**	**0.808(.07)**	0.794(.01)	**0.927(.01)**	**0.860(.05)**	**0.897(.03)**	**0.838(.04)**	0.755(.04)

We conducted ablation analyses on MMWHS by using 40% labeled data to investigate the contribution of the proposed key components (PC and CRS) and the choice of the backbone network. All the models are trained in the SSL way, and we use the thresholded pseudo-labeling [3] (threshold = 0.8) to generate pseudo labels for the models when CRS is absent. Table 2 presents the results of several variants by using different combinations of the proposed key components and two backbone networks. Compared with 3D UNet, H-UNet (model ①) shows more consistent results on different substructures and better performance regardless of the use of the proposed components, with a 1.2% and 4.0% increase in average DSC with and without PC and CRS respectively, indicating the merits of exploiting both intra-slice and inter-slice information for 3D

medical image segmentation. When thresholded pseudo-labeling is replaced with CRS, model ② improves the average DSC by 1.7% with a smaller deviation for each heart substructure among all test samples, suggesting the effectiveness of CRS. PC alleviates the class imbalance problem by leveraging the prior knowledge of the entire sample set to assist individual learning. This is justified by the improvements of 1.5% on average DSC and better results on tail classes (e.g., RA, MYO) brought by substituting LC with PC (model ③). Moreover, the performance is further boosted when these two classifiers are simultaneously adopted (model ④), outperforming LC and PC by a margin of 4.3% and 2.8% respectively. This is because the mutual promotion between the two classifiers could provide a better regularized feature space for pixel predictions. With CRS integrated, PC and LC are trained with more reliable pseudo labels. The DSC is further improved by 1.5% (model ⑤), arriving at the highest value (0.840). Noticeably, when 3D UNet is equipped with all the components, the average DSC is improved by 3.0% with performance soar on tail classes. Such consistent efficacy indicates the potential of using the proposed PC and CRS to endow any segmentation backbones with the capability of addressing the class imbalance problem.

Please refer to Section B in the Supplementary material for more results in terms of average symmetric surface distance (ASSD) in voxel and the qualitative analyses of multi-level tree filters.

4 Conclusion

The scarcity of pixel-level annotations affects the performance of deep neural networks for medical image segmentation. Moreover, the class imbalanced problem existing in the medical data can further exacerbate the model degradation. To address the problems, we propose a novel semi-supervised class imbalanced learning approach by additionally introducing the prototype-based classifier into the student model and constructing two multi-level tree filters to refine the pseudo labels for more robust learning. Experiments conducted on two public cardiac MRI datasets demonstrate the superiority of the proposed method.

Acknowledgement. This work described in this paper was supported in part by the Shenzhen Portion of Shenzhen-Hong Kong Science and Technology Innovation Cooperation Zone under HZQB-KCZYB-20200089. The work was also partially supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project Number: T45-401/22-N) and by a grant from the Hong Kong Innovation and Technology Fund (Project Number: MHP/085/21).

References

1. Hesamian, M.H., Jia, W., He, X., Kennedy, P.: Deep learning techniques for medical image segmentation: achievements and challenges. J. Digit. Imaging **32**, 582–596 (2019)
2. Lee, D.-H., et al.: Pseudo-label: the simple and efficient semi-supervised learning method for deep neural networks. In: Workshop on Challenges in Representation Learning. ICML, vol. 3, p. 896 (2013)
3. Zhang, B., et al.: Flexmatch: boosting semi-supervised learning with curriculum pseudo labeling. Adv. Neural. Inf. Process. Syst. **34**, 18408–18419 (2021)
4. Laine, S., Aila, T.: Temporal ensembling for semi-supervised learning, arXiv preprint arXiv:1610.02242 (2016)
5. Zhou, D., Bousquet, O., Lal, T., Weston, J., Schölkopf, B.: Learning with local and global consistency. In: Advances in Neural Information Processing Systems, vol. 16 (2003)
6. Tajbakhsh, N., Jeyaseelan, L., Li, Q., Chiang, J.N., Wu, Z., Ding, X.: Embracing imperfect datasets: a review of deep learning solutions for medical image segmentation. Med. Image Anal. **63**, 101693 (2020)
7. Bai, W., et al.: Semi-supervised learning for network-based cardiac MR image segmentation. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) MICCAI 2017. LNCS, vol. 10434, pp. 253–260. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66185-8_29
8. Sedai, S., et al.: Uncertainty guided semi-supervised segmentation of retinal layers in OCT images. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11764, pp. 282–290. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32239-7_32
9. Gal, Y., Ghahramani, Z.: Dropout as a Bayesian approximation: representing model uncertainty in deep learning. In: International Conference on Machine Learning, pp. 1050–1059. PMLR (2016)
10. 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: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11765, pp. 605–613. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32245-8_67
11. Basak, H., Ghosal, S., Sarkar, R.: Addressing class imbalance in semi-supervised image segmentation: a study on cardiac MRI. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) MICCAI 2022. LNCS, vol. 13438, pp. 224–233. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16452-1_22
12. Tarvainen, A., Valpola, H.: Mean teachers are better role models: weight-averaged consistency targets improve semi-supervised deep learning results. In: Advances in Neural Information Processing Systems, vol. 30 (2017)
13. Sohn, K., et al.: Fixmatch: simplifying semi-supervised learning with consistency and confidence. Adv. Neural. Inf. Process. Syst. **33**, 596–608 (2020)
14. Li, X., Chen, H., Qi, X., Dou, Q., Fu, C.-W., Heng, P.-A.: H-DenseUNet: hybrid densely connected UNet for liver and tumor segmentation from CT volumes. IEEE Trans. Med. Imaging **37**(12), 2663–2674 (2018)
15. Zhou, T., Wang, W., Konukoglu, E., Van Gool, L.: Rethinking semantic segmentation: a prototype view. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2582–2593 (2022)
16. Huang, C., Wu, X., Zhang, X., Lin, S., Chawla, N.V.: Deep prototypical networks for imbalanced time series classification under data scarcity. In: Proceedings of the 28th ACM International Conference on Information and Knowledge Management, pp. 2141–2144 (2019)

17. Pham, D.L., Xu, C., Prince, J.L.: Current methods in medical image segmentation. Annu. Rev. Biomed. Eng. **2**(1), 315–337 (2000)
18. Song, L., Li, Y., Li, Z., Yu, G., Sun, H., Sun, J., Zheng, N.: Learnable tree filter for structure-preserving feature transform. In: Advances in Neural Information Processing Systems, vol. 32 (2019)
19. Kruskal, J.B.: On the shortest spanning subtree of a graph and the traveling salesman problem. Proc. Am. Math. Soc. **7**(1), 48–50 (1956)
20. Yang, Q.: Stereo matching using tree filtering. IEEE Trans. Pattern Anal. Mach. Intell. **37**(4), 834–846 (2014)
21. Krishna, K., Murty, M.N.: Genetic k-means algorithm. IEEE Trans. Syst. Man Cybern. Part B (Cybern.) **29**(3), 433–439 (1999)
22. Cuturi, M.: Sinkhorn distances: lightspeed computation of optimal transport. In: Advances in Neural Information Processing Systems, vol. 26 (2013)
23. Bernard, O., et al.: Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: is the problem solved? IEEE Trans. Med. Imaging **37**(11), 2514–2525 (2018)
24. Zhuang, X., Shen, J.: Multi-scale patch and multi-modality atlases for whole heart segmentation of MRI. Med. Image Anal. **31**, 77–87 (2016)
25. Chaitanya, K., Karani, N., Baumgartner, C.F., Becker, A., Donati, O., Konukoglu, E.: Semi-supervised and task-driven data augmentation. In: Chung, A.C.S., Gee, J.C., Yushkevich, P.A., Bao, S. (eds.) IPMI 2019. LNCS, vol. 11492, pp. 29–41. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-20351-1_3
26. Berthelot, D., Carlini, N., Goodfellow, I., Papernot, N., Oliver, A., Raffel, C.A.: Mixmatch: a holistic approach to semi-supervised learning. In: Advances in Neural Information Processing Systems, vol. 32 (2019)
27. Chaitanya, K., Erdil, E., Karani, N., Konukoglu, E.: Contrastive learning of global and local features for medical image segmentation with limited annotations. Adv. Neural. Inf. Process. Syst. **33**, 12546–12558 (2020)