Minimal-Supervised Medical Image Segmentation via Vector Quantization Memory

Yanyu Xu, Menghan Zhou, Yangqin Feng, Xinxing Xu(✉), Huazhu Fu, Rick Siow Mong Goh, and Yong Liu

Institute of High Performance Computing (IHPC), Agency for Science, Technology and Research (A*STAR), 1 Fusionopolis Way, #16-16 Connexis, Singapore 138632, Republic of Singapore
{xu_yanyu,zhou_menghan,feng_yangqin,xuxinx,fu_huazhu,gohsm,liuyong}@ihpc.a-star.edu.sg

Abstract. Medical imaging segmentation is a critical key task for computer-assisted diagnosis and disease monitoring. However, collecting a large-scale medical dataset with well-annotation is time-consuming and requires domain knowledge. Reducing the number of annotations poses two challenges: obtaining sufficient supervision and generating high-quality pseudo labels. To address these, we propose a universal framework for annotation-efficient medical segmentation, which is capable of handling both scribble-supervised and point-supervised segmentation. Our approach includes an auxiliary reconstruction branch that provides more supervision and backwards sufficient gradients for learning visual representations. Besides, a novel pseudo label generation branch utilizes the Vector Quantization (VQ) bank to store texture-oriented and global features for generating pseudo labels. To boost the model training, we generate the high-quality pseudo labels by mixing the segmentation prediction and pseudo labels from the VQ bank. The experimental results on the ACDC MRI segmentation dataset demonstrate effectiveness of our designed method. We obtain a comparable performance (0.86 vs. 0.87 DSC score) with a few points.

Keywords: Annotation-efficient Learning · Vector Quantization

1 Introduction

The medical imaging segmentation plays a crucial role in the computer-assisted diagnosis and monitoring of diseases. In recent years, deep neural networks have demonstrated remarkable results in automatic medical segmentation [3,10,22]. However, the process of collecting large-scale and sufficiently annotated medical datasets remains expensive and tedious, requiring domain knowledge and clinical experience. To mitigate the annotation cost, various techniques have been developed to train models using as few annotations as possible, including semi-supervised learning [1,18,20], and weakly supervised learning [5,6,19,26].

H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14222, pp. 625–636, 2023.
https://doi.org/10.1007/978-3-031-43898-1_60

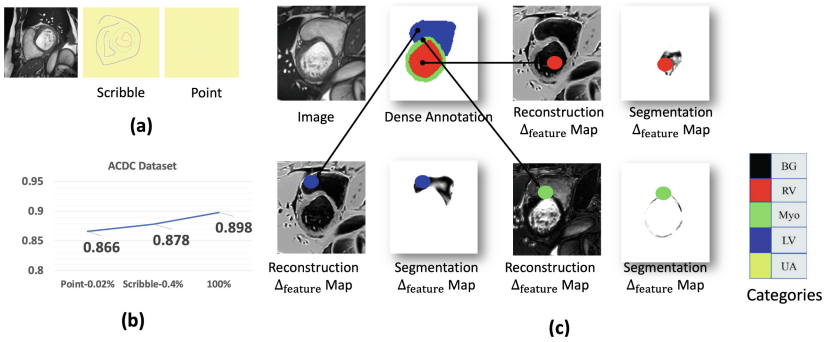

Fig. 1. Illustrations: (a) Examples of point and scribble annotations; (b) Performance comparison; (c) Investigation of similar feature patterns in multi-task branches. The reconstruction and segmentation $\Delta_{feature}$ maps are feature distance between annotated points and the rest of the regions for each class on both the segmentation and reconstruction feature maps. The blue, red, and green dots is annotated points. BG, Myo, LV, RV, and UA are background, myocardium, left ventricle, right ventricle, and unannotated pixels.

In this study, we focus on annotation-efficient learning and propose a universal framework for training segmentation models with scribble and point annotation.

Reducing the number of annotations from dense annotations to scribbles or even points poses two challenges in segmentation: (1) how to obtain sufficient supervision to train a network and (2) how to generate high-quality pseudo labels. In the context of scribble-supervised segmentation, various segmentation methods have been explored, including machine learning or other algorithms [7,23,28], as well as deep learning networks [11,13–15,26,32]. However, these methods only use scribble annotations, excluding point annotations, and their performance remains inferior to training with dense annotations, limiting their practical use in clinical settings. Pseudo labeling [12] is widely used to generate supervision signals for unlabeled images/pixels from imperfect annotations [4, 30]. Recently, some works [17,30,31] have demonstrated that semi-supervised learning can benefit from high-quality pseudo labels. In this study, we propose generating pseudo labels by randomly mixing prediction and texture-oriented pseudo label, which can address the inherent weakness of the previous methods.

This study aims to address the challenges of obtaining sufficient supervisions and generating high-quality pseudo labels. Previous works [33,34] have demonstrated repetitive patterns in texture and feature spaces under single-task learning in both natural and medical images. This observation raises the question of whether similar feature patterns exist in multi-task branches. To investigate this question, we conducted experiments and validated our findings in Fig. 1. We first trained a network with segmentation and reconstruction branches and computed the feature distance distributions between one point and the rest of the regions for each class on both the segmentation and reconstruction feature maps. The features are extracted from the last conv layers in their branches.

Figure 1 (c) shows there are similar patterns between reconstruction and segmentation feature distance maps for each class at the global level. Black color indicates smaller distances. The feature distances in segmentation maps appear to be cleaner than those in the recon maps. It suggests that segmentation features possess task-specific information, while recon features can be seen as a broader set with segmentation information.

Taking inspiration from the similar feature patterns observed in segmentation and reconstruction features, we propose a novel framework that utilizes a memory bank to generate pseudo labels. Our framework consists of an encoder that extracts visual features, as well as two decoders: one for segmenting target objects using scribble or point annotations, and another for reconstructing the input image. To address the challenge of seeking sufficient supervision, we employ the reconstruction branch as an auxiliary task to provide additional supervision and enable the network to learn visual representations. To tackle the challenge of generating high-quality pseudo labels, we use a VQ memory bank to store texture-oriented and global features, which we use to generate the pseudo labels. We then combine information from the global dataset and local image to generate improved, confident pseudo labels.

The contributions of this work can be summarized as follows. **Firstly**, a universal framework for annotation-efficient medical segmentation is proposed, which is capable of handling both scribble-supervised and point-supervised segmentation. **Secondly**, an auxiliary reconstruction branch is employed to provide more supervision and backwards sufficient gradients to learn visual representations. **Thirdly**, a novel pseudo label generation method from memory bank is proposed, which utilizes the VQ memory bank to store texture-oriented and global features to generate high-quality pseudo labels. To boost the model training, we generate high-quality pseudo labels by mixing the segmentation prediction and pseudo labels from the VQ bank. **Finally**, experimental results on public MRI segmentation datasets demonstrate the effectiveness of the proposed method. Specifically, our method outperforms existing scribble-supervised segmentation approaches on the ACDC dataset and also achieves better performance than several semi-supervised methods.

2 Method

In this study, we focus on the problem of annotation-efficient medical image segmentation and propose a universal and adaptable framework for both scribble-supervised and point-supervised learning, as illustrated in Fig. 2. These annotations involve only a subset of pixels in the image and present two challenges: seeking sufficient supervisions to train the network and generating high-quality pseudo labels. To overcome these challenges, we draw inspiration from the recent success of self-supervised learning and propose a framework that includes a reconstruction branch as an auxiliary task and a novel pseudo label generation method using VQ bank memory.

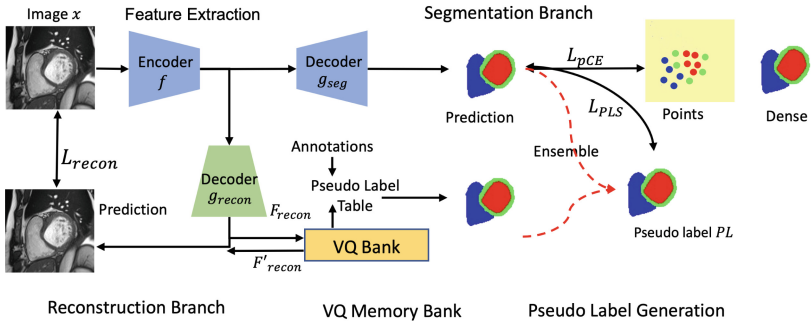

Fig. 2. Overview of the proposed method. It consists of a segmentation task and an auxiliary reconstruction task. One encoder f is used to extract visual features, and one decoder g_{seg} learns from scribble or point annotations to segment target objects, as well as one decoder g_{recon} reconstructs the input image. The memory bank in the reconstruction branch is utilized to generate the pseudo labels, which are then used to assist in the training of the segmentation branch.

Overview: The proposed framework for annotation-efficient medical image segmentation is illustrated in Fig. 2, which consists of a segmentation task and an auxiliary reconstruction task. Firstly, visual features are extracted using one encoder f, and then fed into one decoder g_{seg} to learn from scribble or point annotations to segment target objects, as well as one decoder g_{recon} to reconstruct the input image. The memory bank in the reconstruction branch is utilized to generate the pseudo labels, which are then used to assist in the training of the segmentation branch. The entire network is trained in an end-to-end manner.

Feature Extraction: In this work, we employ a U-Net [22] as the encoder to extract features from the input image x. The size of the input patch is $H \times W$, and the resulting feature map F has the same size as the input patch. It is worth noting that the U-Net backbone used in our work can be replaced with other state-of-the-art structures. Our focus is on designing a universal framework for annotation-efficient medical segmentation, rather than on optimizing the network architecture for a specific task.

Segmentation Branch: The segmentation branch g_{seg} takes the feature map F as input and produces the final segmentation masks based on the available scribble or point annotations. Following recent works such as [13] [24] and [19], we utilize the partial cross-entropy loss to train the decoder $L_{pCE}(y, s) = -\sum_c \sum_{i \in \omega_s} \log y_i^c$, where s denotes the annotation set with reduced annotation efficiency, and y_i^c is the predicted probability of pixel i belonging to class c. The set of labeled pixels in s is denoted by ω_s. To note that the number of pixels s in point annotations is much less than that in scribble annotations, with around $s < 10$ for each class.

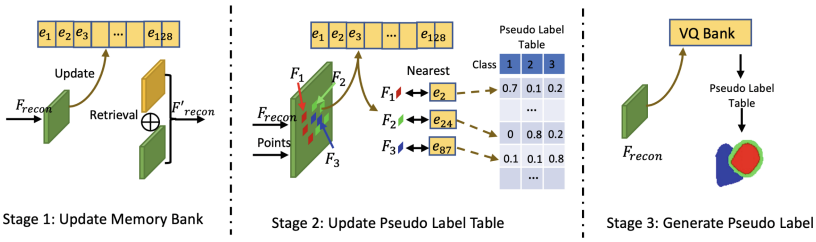

Fig. 3. Overview of the VQ Memory Bank.

Reconstruction Branch: To address the first challenge of seeking sufficient supervisions to train a network with reduced annotations, we propose an auxiliary reconstruction branch. This branch is designed to add more supervision and provide sufficient gradients for learning visual representations. The reconstruction branch has the same decoder structure as the segmentation branch, except for the final prediction layer. We employ the mean squared error loss for the reconstruction task, given by $L_{recon} = |x - y_{recon}|_F^2$, where x is the input image and y_{recon} is the predicted image.

VQ Memory Bank: Motivated by the similar feature patterns observed in medical images, we utilize the Vector Quantization (VQ) memory bank to store texture-oriented and global features, which are then employed for pseudo label generation. The pseudo label generation process involves three stages, as illustrated in Fig. 3.

Memory Bank Definition. In accordance with the VQVAE framework [27] [34], we use a memory bank E to encode and store the visual features on reconstruction branch of the entire dataset. The memory bank E is defined as a dictionary of latent vectors $E := e_1, e_2, ..., e_n$, where $e_i \in R^{1\times 64}$ represents the stored feature in the dictionary and $n = 512$ is the total size of the memory.

Memory Update Stage. The feature map F_{recon} is obtained from the last layer in the reconstruction branch and is utilized to update the VQ memory bank and retrieve an augmented feature $\hat{F}_{recon}$. For each spatial location $f_j \in R^{1\times 64}$ in $F_{recon} \in R^{64\times 256\times 256}$, we use L2 is used to compute the distance between f_j and e_k and find the nearest feature $e_i \in R^{1\times 64}$ in the VQ memory bank, as follows: $\hat{f}_j = e_i, i = \arg\min_k |f_j - e_k|_2^2$. Following [27], we use the VQ loss to update the memory bank and encoder, $L_{VQ} = |\text{sg}[f] - e|_2^2 + |f - \text{sg}[e]|_2^2$, where sg denotes the stop-gradient operator.

Pseudo Label Table Update Stage. The second stage mainly updates a pseudo label table, using the labelled regions on the reconstruction features and assigning pseudo labels on memory vectors. In particular, it uses the labelled pixels and their corresponding reconstruction features and finds the nearest vectors in the

memory bank. As shown in Fig. 3, for the features extracted from the first class regions F_1, its nearest memory vector is e_2. Then, we need to assign probability $[1, 0, 0]$ on e_2 and $[0.7, 0.1, 0.2]$ is stored probabilities after doing Exponentially Moving Average (EMA) and delay is 0.9 in our implementation. We do the same thing for the rest labelled pixels. The pseudo label table is updated on each iteration and records the average values for each vectors.

Pseudo Label Generation Stage. The third stage utilizes the pseudo label table to generate the pseudo labels. It takes the feature map F_{recon} as inputs, then finds their nearest memory vectors, and retrieve the pseudo label according to the vector indices. The generated pseudo label is generated by the repetitive texture patterns on the reconstruction branch, which would include the segmenation information as well as other things.

Pseudo Label Generation: The generation of pseudo labels from the reconstruction branch is based on a texture-oriented and global view, as the memory bank stores the features extracted from the entire dataset. However, relying solely on it may not be sufficient, and it is necessary to incorporate more segmentation-specific information from the segmentation branch. Therefore, we leverage both approaches to enhance the model training.

To incorporate both the segmentation-specific information and the texture-oriented and global information, we dynamically mix the predictions y_1 from the segmentation branch and the pseudo labels y_2 from the VQ memory bank to generate the final pseudo labels y^* [36] [19]. Specifically, we use the following equation: $y^* = argmax[\alpha \times y_1 + (1-\alpha) \times y_2]$, where α is uniformly sampled from $[0, 1]$. The $argmax$ function is used to generate hard pseudo labels. We then use the generated y^* to supervise y_1 and assist in the network training. The pseudo label loss is defined as $L_{pl}(PL, y_1) = 0.5 \times L_{dice}(y^*, y_1)$, where L_{dice} is the dice loss, which can be substituted with other segmentation loss functions such as cross-entropy loss.

Loss Function: Finally, our loss function is calculated as

$$L = L_{pCE} + L_{recon} + L_{PLS}(PL, y_1) + \lambda_{VQ} L_{VQ}, \tag{1}$$

where λ_{VQ} is hyper weights with $\lambda_{VQ} = 0.1$.

3 Experiment

3.1 Experimental Setting

We use the PyTorch [21] platform to implement our model with the following parameter settings: mini-batch size (32), learning rate (3.0e-2), and the number of iterations (60000). We employ the default initialization of PyTorch (1.8.0) to initialize the model. We evaluated our proposed universal framework on scribble

Table 1. Performance Comparisons on the ACDC dataset. All results are based on the 5-fold cross-validation with same backbone (UNet). Mean and standard variance values of 3D DSC and HD_{95} (mm) are presented in this table

Type	Method	RV		Myo		LV		Mean	
		DSC	HD	DSC	HD	DSC	HD	DSC	HD
SSL	PS	0.659(0.261)	26.8(30.4)	0.724(0.176)	16.0(21.6)	0.790(0.205)	24.5(30.4)	0.724(0.214)	22.5(27.5)
	DAN	0.639(0.26)	20.6(21.4)	0.764(0.144)	9.4(12.4)	0.825(0.186)	15.9(20.8)	0.743(0.197)	15.3(18.2)
	AdvEnt	0.615(0.296)	20.2(19.4)	0.760(0.151)	8.5(8.3)	0.848(0.159)	11.7(18.1)	0.741(0.202)	13.5(15.3)
	MT	0.653(0.271)	18.6(22.0)	0.785(0.118)	11.4(17.0)	0.846(0.153)	19.0(26.7)	0.761(0.180)	16.3(21.9)
	UAMT	0.660(0.267)	22.3(22.9)	0.773(0.129)	10.3(14.8)	0.847(0.157)	17.1(23.9)	0.760(0.185)	16.6(20.5)
WSL	pCE	0.625(0.16)	187.2(35.2)	0.668(0.095)	165.1(34.4)	0.766(0.156)	167.7(55.0)	0.686(0.137)	173.3(41.5)
	RW	0.813(0.113)	11.1(17.3)	0.708(0.066)	9.8(8.9)	0.844(0.091)	9.2(13.0)	0.788(0.09)	10.0(13.1)
	USTM	0.815(0.115)	54.7(65.7)	0.756(0.081)	112.2(54.1)	0.785(0.162)	139.6(57.7)	0.786(0.119)	102.2(59.2)
	S2L	0.833(0.103)	14.6(30.9)	0.806(0.069)	37.1(49.4)	0.856(0.121)	65.2(65.1)	0.832(0.098)	38.9(48.5)
	MLoss	0.809(0.093)	17.1(30.8)	0.832(0.055)	28.2(43.2)	0.876(0.093)	37.9(59.6)	0.839(0.080)	27.7(44.5)
	EM	0.839(0.108)	25.7(44.5)	0.812(0.062)	47.4(50.6)	0.887(0.099)	43.8(57.6)	0.846(0.089)	39.0(50.9)
	RLoss	0.856(0.101)	7.9(12.6)	0.817(0.054)	6.0(6.9)	0.896(0.086)	7.0(13.5)	0.856(0.080)	6.9(11.0)
	WSL4MI	0.861(0.096)	7.9(12.5)	0.842(0.054)	9.7(23.2)	0.913(0.082)	12.1(27.2)	0.872(0.077)	9.9(21.0)
	Ours-points	0.843(0.002)	4.7(8.8)	0.842(0.001)	9.0(30.8)	0.916(0.001)	9.7(27.7)	0.866(0.001)	5.1(8.2)
	Ours-scribbles	**0.858(0.001)**	**3.4(4.9)**	**0.857(0.001)**	**3.7(3.2)**	**0.919(0.001)**	**4.3(4.0)**	**0.881(0.001)**	**3.8(2.7)**
FSL	FullSup	0.882(0.095)	6.9(10.8)	0.883(0.042)	5.9(15.2)	0.930(0.074)	8.1(20.9)	0.898(0.070)	7.0(15.6)

and point annotations using the ACDC dataset [2]. The dataset comprises 200 short-axis cine-MRI scans collected from 100 patients, with each patient having two annotated end-diastolic (ED) and end-systolic (ES) phases scans. Each scan has three structures with dense annotation, namely, the right ventricle (RV), myocardium (Myo), and left ventricle (LV). Following previous studies [1,19,26] and consistent with the dataset's convention, we performed 2D slice segmentation instead of 3D volume segmentation. Scribble annotations are simulated by ITK-SNAP. To simulate point annotations, we randomly generated five points for each class. During testing, we predicted the segmentation slice by slice and combined them to form a 3D volume.

3.2 Performance Comparisons

We conducted an evaluation of our model on the ACDC dataset, utilizing the 3D Dice Coefficient (DSC) and the 95% Hausdorff Distance (95) as the metrics. In this study, we compare our proposed model with various state-of-the-art methods and designed baselines. These include: (1) scribble-supervised segmentation methods such as pCE only [15] (lower bound), the model using pseudo labels generated by Random Walker (RW) [7], Uncertainty-aware Self-ensembling and Transformation-consistent Model (USTM) [16], Scribble2Label (S2L) [13], Mumford-shah Loss (MLoss) [11], Entropy Minimization (EM) [8] and Regularized Loss (RLoss) [24]; (2) widely-used semi-supervised segmentation methods, including Deep Adversarial Network (DAN) [37], Adversarial Entropy Minimization (AdvEnt) [29], Mean Teacher (MT) [25], and Uncertainty Aware Mean Teacher (UAMT) [35]. Additionally, we also conduct partially supervised (PS) learning, where only 10% labeled data is used to train the networks.

Table 2. Ablation study.

Method	RV		Myo		LV		Mean	
	DSC	HD	DSC	HD	DSC	HD	DSC	HD
UNet	0.494(0.008)	129.1(119.9)	0.439(0.001)	120.6(41.2)	0.687(0.010)	99.3(254.9)	0.540(0.004)	116.3(495.3)
UNet-PL	0.165(0.002)	125.1(120.5)	0.183(0.001)	119.8(135.6)	0.411(0.010)	122.4(176.4)	0.254(0.002)	122.4(130.1)
UNet-VQ	0.815(0.001)	23.3(188.8)	0.795(0.001)	33.4(179.5)	0.873(0.002)	23.79(61.6)	0.834(0.001)	25.3(43.2)
UNet-add	0.838(0.001)	8.6(20.5)	0.817(0.001)	18.4(42.6)	0.881(0.003)	13.5(69.3)	0.847(0.001)	17.9(33.1)
Point-2	0.835(0.002)	4.9(3.9)	0.817(0.001)	13.1(0.9)	0.902(0.001)	11.7(11.1)	0.851(0.001)	9.9(1.2)
Point-5	0843(0.001)	4.7(8.7)	0.841(0.001)	9.1(30.8)	0.91(0.001)	9.7(27.2)	0.866(0.001)	5.2(8.2)
Point-10	0.858(0.001)	3.7(10.6)	0.846(0.001)	4.7(8.7)	0.919(0.001)	6.2(39.8)	0.874(0.001)	4.9(15.1)
Scribble	0.858(0.001)	3.4(4.9)	0.858(0.001)	3.7(3.2)	0.920(0.001)	4.3(4.0)	0.878(0.001)	3.8(2.7)

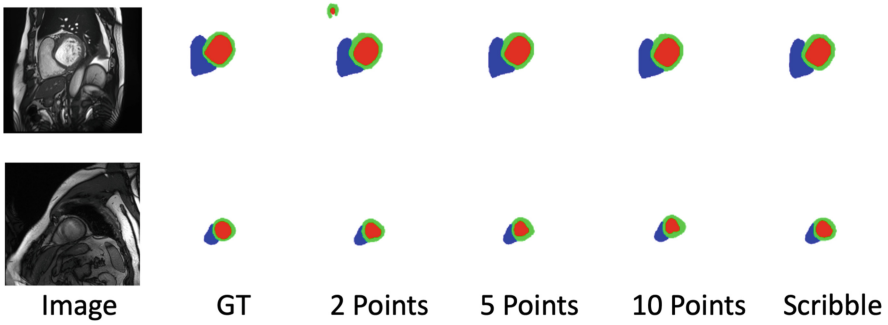

Fig. 4. Qualitative comparison of our method using different level of annotations.

Table 1 presents the performance comparisons of our proposed method with state-of-the-art methods on the ACDC dataset. Our proposed method employing scribble annotations achieves superior performance over existing semi-supervised and weakly-supervised methods. Furthermore, our proposed method utilizing a few point annotations demonstrates comparable performance with weakly-supervised methods and outperforms semi-supervised segmentation methods by a significant margin. Despite achieving slightly lower performance than fully supervised methods, our proposed method requires much lower annotation costs. We present the qualitative comparison results in Fig. 4. The visual analysis of the results indicates that our proposed methods using scribble even point annotation perform well in terms of visual similarity with the ground truth. These results demonstrate the effectiveness of using scribble or point annotations as a potential way to reduce the annotation cost.

3.3 Ablation Study

We investigate the effect of our proposed method on the ACDC datasets.

Effect of the Auxiliary Task: We designed a baseline by removing the reconstruction branch and VQ memory. The results in Table 2 show a significant

performance gap, indicating the importance of the reconstruction branch in stabilizing the training process. We also use local pixel-wise contrastive learning [9] to replace reconstruction and keep the rest same. Results are 0.85 for point.

Effect of Pseudo Labels: We also designed a baseline using only the predictions as pseudo labels. In Table 2, the performance drop highlights the effectiveness of the texture-orient and global information in the VQ memory bank. The size and dimension of embedding of VQ bank are 512 and 64 the default setting in VQVAE. Model results (sizes of bank 64, 256, 512) are 0.865, 0.866, 0.866. We find 20–24 vectors are commonly used as clusters of 95% features and can set 64 as bank size to save memory.

Effect of Different Levels of Aannotations: We also evaluated the impact of using different levels of annotations in point-supervised learning, ranging from more annotations, *e.g.* 10 points, to fewer annotations, *e.g.* 2 points. The results in Table 2 indicate that clicking points is a promising data annotation approach to reduce annotation costs. Overall, our findings suggest that the proposed universal framework could effectively leverage different types of annotations and provide high-quality segmentation results with less annotation costs.

4 Conclusion

In this study, we introduce a universal framework for annotation-efficient medical segmentation. Our framework leverages an auxiliary reconstruction branch to provide additional supervision to learn visual representations and a novel pseudo label generation method from memory bank, which utilizes the VQ memory bank to store global features to generate high-quality pseudo labels. We evaluate the proposed method on a publicly available MRI segmentation dataset, and the experimental results demonstrate its effectiveness.

Acknowledgement. This research/project is supported by the National Research Foundation, Singapore under its AI Singapore Programme (AISG Award No: AISG2-TC-2021-003) This work was supported by the Agency for Science, Technology and Research (A*STAR) through its AME Programmatic Funding Scheme Under Project A20H4b0141. This work was partially supported by A*STAR Central Research Fund "A Secure and Privacy Preserving AI Platform for Digital Health"

References

1. 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

2. Bernard, O., Lalande, A., Zotti, C., 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)
3. Chen, J., et al.: Transunet: transformers make strong encoders for medical image segmentation. arXiv preprint arXiv:2102.04306 (2021)
4. Chen, X., Yuan, Y., Zeng, G., Wang, J.: Semi-supervised semantic segmentation with cross pseudo supervision. In: CVPR, pp. 2613–2622 (2021)
5. Dolz, J., Desrosiers, C., Ayed, I.B.: Teach me to segment with mixed supervision: confident students become masters. In: Feragen, A., Sommer, S., Schnabel, J., Nielsen, M. (eds.) IPMI 2021. LNCS, vol. 12729, pp. 517–529. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-78191-0_40
6. Dorent, R., et al.: Inter extreme points geodesics for end-to-end weakly supervised image segmentation. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12902, pp. 615–624. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87196-3_57
7. Grady, L.: Random walks for image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **28**(11), 1768–1783 (2006)
8. Grandvalet, Y., Bengio, Y.: Semi-supervised learning by entropy minimization. Adv. Neural Inf. Process. Syst. **17** (2004)
9. Hu, X., Zeng, D., Xu, X., Shi, Y.: Semi-supervised contrastive learning for label-efficient medical image segmentation. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12902, pp. 481–490. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87196-3_45
10. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. Nat. Methods **18**(2), 203–211 (2021)
11. Kim, B., Ye, J.C.: Mumford-shah loss functional for image segmentation with deep learning. IEEE Trans. Image Process. **29**, 1856–1866 (2019)
12. 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)
13. Lee, H., Jeong, W.-K.: Scribble2Label: scribble-supervised cell segmentation via self-generating pseudo-labels with consistency. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12261, pp. 14–23. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59710-8_2
14. Li, S., et al.: Few-shot domain adaptation with polymorphic transformers. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12902, pp. 330–340. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87196-3_31
15. Lin, D., Dai, J., Jia, J., He, K., Sun, J.: Scribblesup: scribble-supervised convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3159–3167 (2016)
16. Liu, X., et al.: Weakly supervised segmentation of covid19 infection with scribble annotation on ct images. Pattern Recogn. **122**, 108341 (2022)
17. Luo, W., Yang, M.: Semi-supervised semantic segmentation via strong-weak dual-branch network. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) ECCV 2020. LNCS, vol. 12350, pp. 784–800. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58558-7_46
18. 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)

19. Luo, X., et al.: Scribble-supervised medical image segmentation via dual-branch network and dynamically mixed pseudo labels supervision. arXiv preprint arXiv:2203.02106 (2022)
20. Luo, X., et al.: Efficient semi-supervised gross target volume of nasopharyngeal carcinoma segmentation via uncertainty rectified pyramid consistency. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12902, pp. 318–329. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87196-3_30
21. Paszke, A., Gross, S., Massa, F., et al.: Pytorch: an imperative style, high-performance deep learning library. Adv. Neural Inf. Process. Syst. **32**, 8024–8035 (2019)
22. 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
23. Rother, C., Kolmogorov, V., Blake, A.: "grabcut" interactive foreground extraction using iterated graph cuts. ACM Trans. Graph. (TOG) **23**(3), 309–314 (2004)
24. Tang, M., Perazzi, F., Djelouah, A., Ben Ayed, I., Schroers, C., Boykov, Y.: On regularized losses for weakly-supervised cnn segmentation. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 507–522 (2018)
25. Tarvainen, A., Valpola, H.: Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. Adv. Neural Inf. Process. Syst. **30**, 1195–1204 (2017)
26. Valvano, G., Leo, A., Tsaftaris, S.A.: Learning to segment from scribbles using multi-scale adversarial attention gates. IEEE Trans. Med. Imaging **40**(8), 1990–2001 (2021)
27. Van Den Oord, A., Vinyals, O., et al.: Neural discrete representation learning. In: NeurIPS (2017)
28. Vezhnevets, V., Konouchine, V.: Growcut: interactive multi-label nd image segmentation by cellular automata. In: Proceedings of Graphicon, vol. 1, pp. 150–156. Citeseer (2005)
29. Vu, T.H., Jain, H., Bucher, M., Cord, M., Pérez, P.: Advent: adversarial entropy minimization for domain adaptation in semantic segmentation. In: CVPR, pp. 2517–2526 (2019)
30. Wang, X., Gao, J., Long, M., Wang, J.: Self-tuning for data-efficient deep learning. In: International Conference on Machine Learning, pp. 10738–10748. PMLR (2021)
31. Wu, Y., Xu, M., Ge, Z., Cai, J., Zhang, L.: Semi-supervised left atrium segmentation with mutual consistency training. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12902, pp. 297–306. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87196-3_28
32. Xu, Y., Xu, X., Fu, H., Wang, M., Goh, R.S.M., Liu, Y.: Facing annotation redundancy: Oct layer segmentation with only 10 annotated pixels per layer. In: Xu, X., Li, X., Mahapatra, D., Cheng, L., Petitjean, C., Fu, H. (eds.) REMIA 2022. LNCS, pp. 126–136. Springer, Heidelberg (2022). https://doi.org/10.1007/978-3-031-16876-5_13
33. Xu, Y., et al.: Partially-supervised learning for vessel segmentation in ocular images. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12901, pp. 271–281. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87193-2_26
34. Xu, Y., et al.: Crowd counting with partial annotations in an image. In: ICCV, pp. 15570–15579 (2021)

35. 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
36. Zhang, H., Cisse, M., Dauphin, Y.N., Lopez-Paz, D.: mixup: Beyond empirical risk minimization. arXiv preprint arXiv:1710.09412 (2017)
37. Zhang, Y., Yang, L., Chen, J., Fredericksen, M., Hughes, D.P., Chen, D.Z.: Deep adversarial networks for biomedical image segmentation utilizing unannotated images. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) MICCAI 2017. LNCS, vol. 10435, pp. 408–416. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66179-7_47