Scribble-Supervised Medical Image Segmentation via Dual-Branch Network and Dynamically Mixed Pseudo Labels Supervision

Xiangde Luo[1,2], Minhao Hu[3], Wenjun Liao[1], Shuwei Zhai[1], Tao Song[3],
Guotai Wang[1,2(✉)], and Shaoting Zhang[1,2]

[1]University of Electronic Science and Technology of China, Chengdu, China
[2]Shanghai AI Lab, Shanghai, China
[3]SenseTime Research, Shanghai, China
https://github.com/HiLab-git/WSL4MIS
xiangde.luo@std.uestc.edu.cn and guotai.wang@uestc.edu.cn

Abstract. Medical image segmentation plays an irreplaceable role in computer-assisted diagnosis, treatment planning and following-up. Collecting and annotating a large-scale dataset is crucial to training a powerful segmentation model, but producing high-quality segmentation masks is an expensive and time-consuming procedure. Recently, weakly-supervised learning that uses sparse annotations (points, scribbles, bounding boxes) for network training has achieved encouraging performance and shown the potential for annotation cost reduction. However, due to the limited supervision signal of sparse annotations, it is still challenging to employ them for networks training directly. In this work, we propose a simple yet efficient scribble-supervised image segmentation method and apply it to cardiac MRI segmentation. Specifically, we employ a dual-branch network with one encoder and two slightly different decoders for image segmentation and dynamically mix the two decoders' predictions to generate pseudo labels for auxiliary supervision. By combining the scribble supervision and auxiliary pseudo labels supervision, the dual-branch network can efficiently learn from scribble annotations end-to-end. Experiments on the public ACDC dataset show that our method performs better than current scribble-supervised segmentation methods and also outperforms several semi-supervised segmentation methods.

Keywords: Weakly-supervised learning · scribble annotation · pseudo labels

1 Introduction

Recently, Convolutional Neural Networks (CNNs) and Transformers have achieved encouraging results in automatic medical image segmentation [5, 12, 25]. Most of them need large-scale images with accurate pixel-level dense annotations to train models. However, collecting a large-scale and carefully annotated medical image dataset is still an expensive and time-consuming journey, as it requires domain knowledge and clinical experience [19,21]. Recently, many efforts have been made to reduce the annotation cost for models training to alleviate this issue. For example, semi-supervised learning (SSL) combines a few labeled data and massive unlabeled data for network training [1, 19, 20].

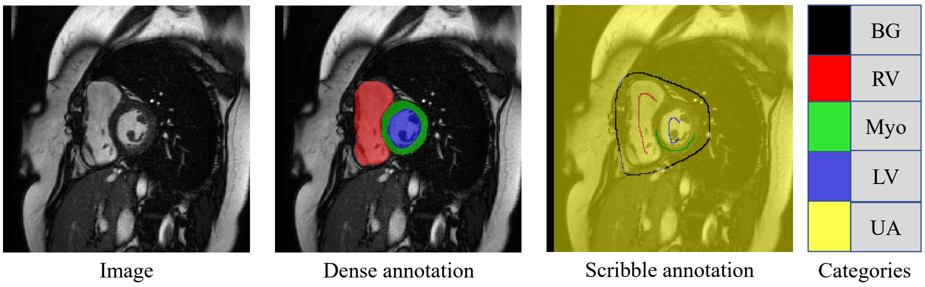

Image Dense annotation Scribble annotation Categories

Fig. 1. Examples of dense and scribble annotations. BG, RV, Myo, LV, and UA represent the background, right ventricle, myocardium, left ventricle, and unannotated pixels respectively.

Weakly supervised learning (WSL) uses sparse annotations to train models rather than dense annotations [7,8,29]. Considering collecting sparse annotations (points, scribbles and bounding boxes) is easier than dense annotations [16] and scribbles have better generality to annotate complex objects than bounding boxes and points [16,29] (Example in Fig. 1). This work focuses on exploring scribble annotations to train high-performance medical image segmentation networks efficiently and robustly.

Scribble-Supervised Segmentation: Using scribble annotations to segment objects has been studied for many years. Before the deep learning era, combining user-provided sparse annotations and machine learning or other algorithms was the most popular and general segmentation method, such as GraphCuts [3], GrabCut [26], Random Walker [9], GrowCut [30], ITK-SNAP [36], Slic-Seg [32], etc. Recently, deep learning with convolutional neural networks or transformers can learn to segment from dense annotations and then inference automatically. So, it is desirable to train powerful segmentation networks using scribble annotations. To achieve this goal, Lin et al [16] proposed a graphical-based method to propagate information from scribbles to unannotated pixels and train models jointly. After that, Tang et al [27] introduced a Conditional Random Field (CRF) regularization loss to train segmentation networks directly. For medical images, Can et al [4] proposed an iterative framework to train models with scribbles. At first, they seeded the scribbles into the Random Walker [9] to produce the initial segmentation. Then, they used the initial segmentation to train the model and refine the model's prediction with CRF for the network retraining. Finally, they repeated the second procedure several times for powerful segmentation models. Kim et al [13] proposed a level set-based [22] regularization function to train deep networks with weak annotations. Lee et al [15] combined pseudo-labeling and label filtering to generate reliable labels for network training with scribble supervisions. Liu et al [17] presented a unified weakly-supervised framework to train networks from scribble annotations, which consists of an uncertainty-aware mean teacher and a transformation-consistent strategy. More recently, Valvano et al [29] proposed multi-scale adversarial attention gates to train models with mixed scribble and dense annotations. Although these attempts have saved the annotation cost by using scribble annotations, the performance is still lower than training with dense annotations, limiting the applicability in clinical practice.

Pseudo Labels for Segmentation: Pseudo labeling [14] is widely used to generate supervision signals for unlabeled images/pixels. The main idea is utilizing imperfect annotations to produce high-quality and reliable pseudo labels for network training [6,33]. Recently, some works have demonstrated [18,33,34] that semi-supervised learning can benefit from high-quality pseudo labels. For weakly-supervised learning, Lee et al [15] showed that generating pseudo labels by ensembling predictions at a temporal level can boost performance. *Nevertheless, recent work [11] points out the inherent weakness of these methods that the model retains the prediction from itself and thus resists updates.* Recently, some works resort to perturbation-consistency strategy for semi-supervised learning [23,34], where the main branch is assisted by auxiliary branches that are typically perturbed and encouraged to produce similar predictions to the main branch. In this work, we assume that generating pseudo labels by mixing multiple predictions randomly can go against the above inherent weakness, as these auxiliary branches are added perturbations and do not enable interaction with each other.

Motivated by these observations, we present a simple yet efficient approach to learning from scribble annotations. Particularly, we employ a dual branches network (one encoder and two slightly different decoders) as the segmentation network. To learn from scribbles, the dual branches network is supervised by the partially cross-entropy loss (pCE), which only considers annotated pixels' gradient for back-propagation and ignores unlabeled pixels. At the same time, we employ the two predictions to generate hard pseudo labels for more substantial and more reliable supervision signals than scribbles. Afterward, we combine scribbles supervision and pseudo labels supervision to train the segmentation network end-to-end. Differently from threshold-based methods [15,34], we generate hard pseudo labels by dynamically mixing two branches' predictions, which can help against the inherent weakness [11]. Such a strategy imposes the segmentation network to produce high-quality pseudo labels for unannotated pixels. We evaluate our method on a public scribble-supervised benchmark ACDC [2]. Experiments results show that our proposed method outperforms existing scribble-supervised methods when using the same scribble annotations and also performs better than semi-supervised methods when taking similar annotation budgets.

The contributions of this work are two-fold. Firstly, we propose a dual-branch network and a dynamically mixed pseudo labeling strategy to train segmentation models with scribble annotations. Specifically, we generate high-quality hard pseudo labels by randomly mixing the two branches' outputs and use the generated pseudo labels to supervise the network training end-to-end. (2) Extensive experiments on the public cardiac MRI segmentation dataset (ACDC) demonstrate the effectiveness of the proposed method. Our method has achieved better performance on the ACDC dataset than existing scribble-supervised segmentation approaches and also outperformed several semi-supervised segmentation methods with similar annotation costs.

2 Method

The proposed framework for scribble-supervised medical image segmentation is depicted in Fig. 2. We firstly employ a network with one encoder and two slightly different decoders to learn from scribble annotations to segment target objects. At the same time, we utilize the two branches' outputs to generate hard pseudo labels that are used to

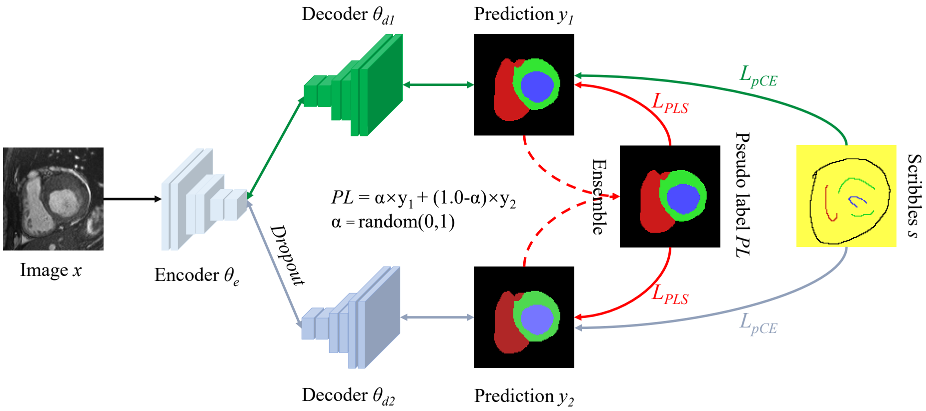

Fig. 2. Overview of the proposed method. The framework consists of an encoder (θ_e), the main decoder (θ_{d1}), and an auxiliary decoder (θ_{d2}) and is trained with scribble annotations separately (L_{pCE}). At the same time, the hard pseudo label is generated by dynamically mixing two decoders' outputs and used as the pseudo labels supervision for further network training (L_{PLS}).

assist the network training. Note that the training procedure is in an end-to-end manner rather than the multi-stage [4] or iterative refinement strategies [16].

2.1 Learning from Scribbles

For general scribble-supervised learning, the available dataset consists of images and scribble annotations , where the scribble is a set of pixels with a category or unknown label. Previous work [4] uses interactive segmentation methods [9] to propagate annotated pixels to the whole image for a rough segmentation and then train deep networks with the segmentation in a fully-supervised manner. Recently, there are much better alternatives [15, 27], e.g., using scribbles to train CNNs directly by minimizing a partial cross-entropy loss:

$$L_{pCE}(y, s) = -\sum_{c}\sum_{i \in \omega_s} \log y_i^c \tag{1}$$

where s represents the one-hot scribble annotations. y_i^c is the predicted probability of pixel i belonging class c. ω_s is the set of labeled pixels in s.

2.2 Dual-branch Network

The proposed network ($f(\theta_e, \theta_{d1}, \theta_{d2})$) is composed of a shared encoder (θ_e) for feature extraction and two independent and different decoders (θ_{d1}, θ_{d2}) for segmentation and supplementary training (see Fig. 2). We embed a perturbed decoder into the general UNet [25], where the dropout [23] is used to introduce perturbation at the feature level. This design has two advantages: (1) It can be against the inherent weakness of pseudo-label in the single branch network [11], as the two branches' outputs are different due to the feature perturbation. (2) It can generate pseudo-label by two outputs ensemble but does not require training two networks, and the encoder benefits from the two individual supervisions to boost the feature extraction ability [23, 34]. It is worthy to point out that some recent works used similar architecture for the consistency training [19, 23, 34] or knowledge distillation [7]. There are many significant differences in the learning scenarios and supervision strategies. Firstly, [19, 23, 34] concentrate on semi-supervised learning and [7] focus on knowledge distillation but we aim to scribble-supervised segmentation. Secondly, they employ consistency regularization to supervise networks, but we randomly mix two outputs to generate hard pseudo labels for fully supervised learning. These differences lead to different training, optimization strategies, and results.

2.3 Dynamically Mixed Pseudo Labels

Based on the dual-branch network, we further exploit the two decoders' outputs to boost the model training. We generate the hard pseudo labels by mixing two predictions dynamically, like mixup [37]. The dynamically mixed pseudo labels (*PL*) generation strategy is defined as:

$$PL = argmax[\alpha \times y_1 + (1.0 - \alpha) \times y_2], \; \alpha = random(0, 1) \tag{2}$$

where y_1 and y_2 are outputs of decoder 1 and 2, respectively. α is randomly generated in (0, 1) at each iteration. This strategy boosts the diversity of pseudo labels and avoids the inherent weakness of the pseudo labeling strategy (remembering itself predictions without updating) [11]. *argmax* is used to generate hard pseudo labels. Compared with consistency learning [23, 34], this strategy cuts off the gradient between θ_{d1} and θ_{d2} to maintain their independence rather than enforce consistency directly. In this way, the supervision signal is enlarged from a few pixels to the whole image, as the scribbled pixels are propagated to all unlabeled pixels by the dynamically mixed pseudo labeling. Then, we further employ the generated *PL* to supervise θ_{d1} and θ_{d2} separately to assist the network training. The *P*seudo *L*abels *S*upervision (*PLS*) is defined as :

$$L_{PLS}(PL, y_1, y_2) = 0.5 \times (L_{Dice}(PL, y_1) + L_{Dice}(PL, y_2)) \tag{3}$$

where L_{Dice} is the widely-used dice loss and also can be replaced by cross-entropy loss or other segmentation loss functions. Finally, the proposed network can be trained with scribble annotations by minimizing the following joint object function:

$$L_{total} = \underbrace{0.5 \times (L_{pCE}(y_1, s) + L_{pCE}(y_2, s))}_{scribble\ supervision} + \lambda \times \underbrace{L_{PLS}(PL, y_1, y_2)}_{pseudo\ labels\ supervision} \tag{4}$$

λ is a weight factor to balance the supervision of scribbles and pseudo labels.

3 Experiment and Results

3.1 Experimental Details

Dataset: We evaluate the proposed method on the training set of ACDC [2] via five-fold cross-validation. This dataset is publicly available, with 200 short-axis cine-MRI

Table 1. Comparison with existing weakly-/semi-supervised methods on the ACDC dataset. All results are based on the *5-fold cross-validation* with same backbone (UNet). Mean and standard variance (in parentheses) values of 3D *DSC* and HD_{95} (mm) are presented in this table. * denotes p-value < 0.05 (paired t-test) when comparing with the second place method (RLoss [27]).

Type	Method	**RV**		**Myo**		**LV**		**Mean**	
		DSC	HD_{95}	*DSC*	HD_{95}	*DSC*	HD_{95}	*DSC*	HD_{95}
WSL	pCE [16]	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 [9]	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 [17]	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 [15]	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 [13]	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 [10]	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 [27]	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)**
	Ours	**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)
SSL	PS [25]	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 [38]	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 [31]	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 [28]	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 [35]	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)
FSL	FullSup [25]	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)

scans from 100 patients, and each patient has two annotated end-diastolic (ED), and end-systolic (ES) phases scans. And each scan has three structures' dense annotation, including the right ventricle (RV), myocardium (Myo), and left ventricle (LV). Recently, Valvano et al [29] provided the scribble annotation for each scan manually. Following previous works [1, 29], we employ the 2D slice segmentation rather than 3D volume segmentation, as the thickness is too large.

Implementation Details: We employed the UNet [25] as the base segmentation network architecture, and we further extended the basic UNet to dual branches network by embedding an auxiliary decoder. We added the dropout layer (ratio=0.5) before each conv-block of the auxiliary decoder to introduce perturbations. We implemented and ran our proposed and other comparison methods by PyTorch [24] on a cluster with 8 TiTAN 1080TI GPUs. For the network training, we first re-scaled the intensity of each slice to 0-1. Then, random rotation, random flipping, random noise were used to enlarge the training set, and the augmented image was resized to 256 × 256 as the network input. We used the SGD (weight decay = 10^{-4}, momentum = 0.9) to minimize the joint object function Eq. 4 for the model optimization. The poly learning rate strategy was used to adjust the learning rate online [20]. The batch size, total iterations, and λ are set to 12, 60k, and 0.5, respectively. For testing, we produced predictions slice by slice and stacked them into a 3D volume. For a fair comparison, we used the primary decoder's output as the final result during the inference stage and did not use any post-processing method. Note that all experiments were conducted in the same experimental setting. The 3D Dice Coefficient (DSC) and 95% Hausdorff Distance (HD_{95}) are used as evaluation metrics. All code, data, and details of existing and proposed methods at: `https://github.com/HiLab-git/WSL4MIS`.

3.2 Results

Comparison with Other Methods: Firstly, we compared our method with seven scribble-supervised segmentation methods with the same set of scribbles: 1) pCE only [16] (lower bound), 2) using pxeudo label generated by Random Walker (RW) [9], 3) Uncertainty-aware Self-ensembling and Transformation-consistent Model (USTM) [17], 4) Scribble2Label (S2L) [15], 5) Mumford–shah Loss (MLoss) [13], 6) Entropy Minimization (EM) [10], 7) Regularized Loss (RLoss) [27]. The first section of Table 1 lists the quantitative comparison of the proposed with seven existing weakly supervised learning methods. It can be found that our method achieved the best performance in terms of mean *DSC* (p-value < 0.05) and second place in the HD_{95} metric than other methods. Afterward, we further compared our method with other popular annotation-efficient segmentation methods, e.g., semi-supervised learning methods. Following [8], we investigated the performance difference of these approaches when using very similar annotation costs. To do so, we trained networks with partially supervised and semi-supervised fashions, respectively. We used a 10% training set (8 patients) as labeled data and the remaining as unlabeled data, as the scribble annotation also takes similar annotation costs [29]. For partially supervised (*PS*) learning, we used the 10% labeled data to train networks only. For semi-supervised learning, we combined the 10% labeled data and 90% unlabeled data to train models jointly. We further employed four widely-used semi-supervised segmentation methods for comparison: 1) Deep Adversarial Network (DAN) [38], 2) Adversarial Entropy Minimization (AdvEnt) [31], 3) Mean Teacher (MT) [28], and Uncertainty Aware Mean Teacher (UAMT) [35]. The quantitative comparison is presented in the second section of Table 1. It shows that the scribbled annotation can achieve better results than pixel-wise annotation when taking a similar annotation budget. Moreover, our weakly-supervised method significantly outperforms existing semi-supervised methods in the cardiac MR segmentation. Finally, we also investigated the upper bound when using all mask annotation to train models (FullSup) in the last row of Table 1. It can be found that our method is slightly inferior compared with fully supervised learning with pixel-wise annotation. But our method requires fewer annotation costs than pixel-wise annotation. Fig. 3 shows the segmentation results obtained by existing and our methods, and the corresponding ground truth on the ACDC dataset (patient026_frame01). We can observe that the result obtained by our method is more similar to the ground truth than the others. It further shows that drawing scribble is a potential data annotation approach to reduce annotation costs.

Sensitivity analysis of λ***:*** The study was conducted to assess the sensitivity of λ in Eq. 4. Particularly, the *PLS* term plays a crucial role in the proposed framework, as it controls the usage of the pseudo labels during the network training. We investigated the segmentation performance of the proposed framework when the λ is set to $\{0.01, 0.1, 0.2, 0.3, 0.5, 1.0\}$. Fig. 4 shows the evolution of the segmentation result of RV, Myo, LV, and their average results, all these results are based on the 5-fold cross-validation. It can be observed that increasing λ from 0.01 to 0.5 leads to better performance in terms of both *DSC* and HD_{95}. When the λ is set to 1.0, the segmentation result just decreases slightly compared with 0.5 (0.872 vs 0.870 in term of mean *DSC*). These observations show that the proposed method is not sensitive to λ.

Ablation Study: We further investigated the effect of using different supervision approaches for the dual-branch network: 1) Consistency Regularization (*CR*) [7] that encourages the two predictions to be similar, directly; 2) Cross Pseudo Supervision (*CPS*) [6, 34] that uses one decoder's output as the pseudo label to supervise the other one; 3) the proposed approach dynamically mixes two outputs to generate hard pseudo labels for two decoders training separately. We trained the dual-branch network with

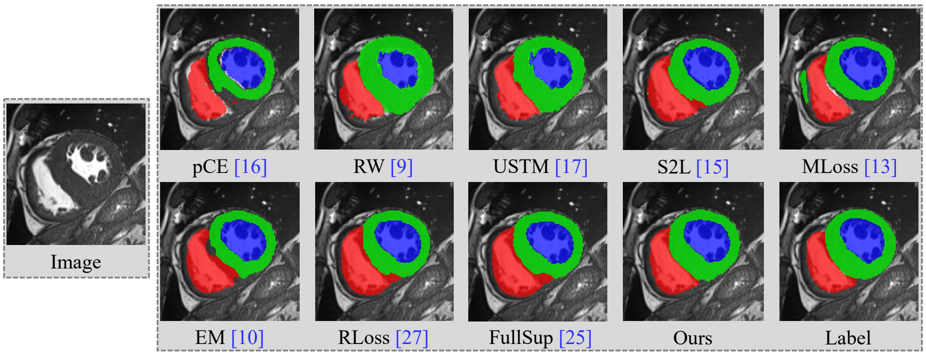

Fig. 3. Qualitative comparison of our proposed method and several existing ways.

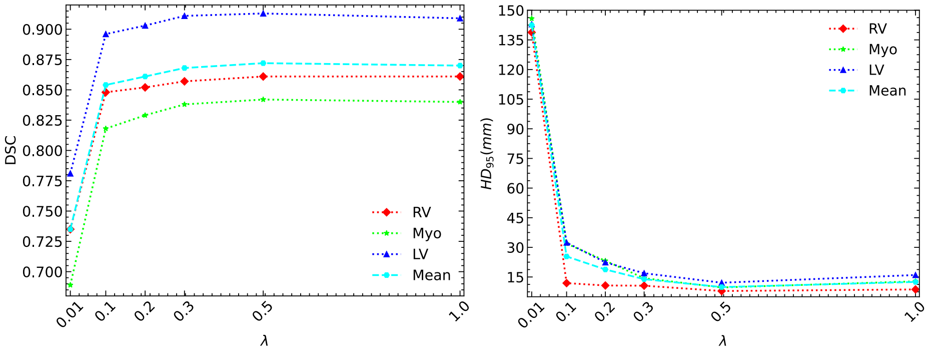

Fig. 4. Sensitivity analysis of hyper-parameter λ.

scribbles and the above supervision strategies. The quantitative evaluation results are presented in Table 2. It can be observed that compared with *CR* and *CPS*, using our proposed *PLS* leads to the best performance. Moreover, we also investigated the performance when α is set to a fixed value (0.5) and dynamic values. The result demonstrates the effectiveness of the proposed dynamically mixing strategy. In addition, we found that the main (θ_{d1}) and auxiliary (θ_{d2}) decoders achieve very similar results.

4 Conclusion

In this paper, we presented pseudo labels supervision strategy for scribble-supervised medical image segmentation. A dual-branch network is employed to learn from scribble annotations in an end-to-end manner. Based on the dual-branch network, a dynamically mixed pseudo labeling strategy was presented to propagate the scribble annotations to the whole image and supervise the network training. Experiments on a public cardiac MR image segmentation dataset (ACDC) demonstrated the effectiveness of the proposed method, where it outperformed seven recent scribble-supervised segmentation methods using the same scribble annotations and four semi-supervised segmentation

Table 2. Ablation study on different supervision strategies for the dual-branch network. Single denotes the baseline UNet [25] with *pCE* only. *CR* means consistency regularization between the main and auxiliary decoders [7]. *CPS* is the cross pseudo supervision strategy in [6, 34]. *Ours* is proposed *PLS*, θ_{d1} and θ_{d2} mean the prediction of main and auxiliary decoders, respectively.

Method	**RV**		**Myo**		**LV**		**Mean**	
	DSC	HD_{95}	*DSC*	HD_{95}	*DSC*	HD_{95}	*DSC*	HD_{95}
Single [16]	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)
Dual+***CR*** [7]	0.844(0.106)	20.1(37.2)	0.798(0.07)	62.2(55.7)	0.873(0.101)	63.4(65.5)	0.838(0.092)	48.6(52.8)
Dual+***CPS*** [6,34]	0.849(0.099)	12.4(25.6)	0.833(0.056)	19.3(33.5)	0.905(0.091)	18.3(35.8)	0.863(0.082)	16.6(31.6)
Ours (α=0.5, θ_{d1})	0.855(0.101)	8.6(13.9)	0.837(0.053)	13.6(29.1)	0.908(0.086)	15.8(34.1)	0.866(0.08)	12.6(25.7)
Ours (α=random, θ_{d1})	**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 (α=random, θ_{d2})	0.861(0.098)	**7.3(10.3)**	0.840(0.058)	10.9(24.5)	0.911(0.086)	**11.3(26.4)**	0.871(0.08)	**9.8(20.4)**

methods with very similar annotation costs. In the future, we will extend and evaluate the proposed method on other challenging medical image segmentation tasks.

5 Acknowledgments

This work was supported by the Beijing Nova Program (Z201100006820064).

References

1. Bai, W., Oktay, O., Sinclair, M., Suzuki, H., Rajchl, M., Tarroni, G., Glocker, B., King, A., Matthews, P.M., Rueckert, D.: Semi-supervised learning for network-based cardiac MR image segmentation. In: MICCAI. pp. 253–260. Springer (2017)
2. Bernard, O., Lalande, A., Zotti, C., Cervenansky, F., Yang, X., Heng, P.A., Cetin, I., et al.: Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: is the problem solved? TMI 37(11), 2514–2525 (2018)
3. Boykov, Y.Y., Jolly, M.P.: Interactive graph cuts for optimal boundary & region segmentation of objects in ND images. In: ICCV. vol. 1, pp. 105–112. IEEE (2001)
4. Can, Y.B., Chaitanya, K., Mustafa, B., Koch, L.M., Konukoglu, E., Baumgartner, C.F.: Learning to segment medical images with scribble-supervision alone. In: DLMIA, pp. 236–244. Springer (2018)
5. Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A.L., Zhou, Y.: Transunet: Transformers make strong encoders for medical image segmentation. arXiv preprint arXiv:2102.04306 (2021)
6. Chen, X., Yuan, Y., Zeng, G., Wang, J.: Semi-supervised semantic segmentation with cross pseudo supervision. In: CVPR. pp. 2613–2622 (2021)
7. Dolz, J., Desrosiers, C., Ayed, I.B.: Teach me to segment with mixed supervision: Confident students become masters. In: IPMI. pp. 517–529. Springer (2021)
8. Dorent, R., Joutard, S., Shapey, J., Kujawa, A., Modat, M., Ourselin, S., Vercauteren, T.: Inter extreme points geodesics for end-to-end weakly supervised image segmentation. In: MICCAI. pp. 615–624. Springer (2021)
9. Grady, L.: Random walks for image segmentation. TPAMI (11), 1768–1783 (2006)
10. Grandvalet, Y., Bengio, Y., et al.: Semi-supervised learning by entropy minimization. NeurIPS 367, 281–296 (2005)
11. Huo, X., Xie, L., He, J., Yang, Z., Zhou, W., Li, H., Tian, Q.: ATSO: Asynchronous teacher-student optimization for semi-supervised image segmentation. In: CVPR. pp. 1235–1244 (2021)

12. 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. Nature methods 18(2), 203–211 (2021)
13. Kim, B., Ye, J.C.: Mumford-Shah loss functional for image segmentation with deep learning. IEEE Transactions on Image Processing 29, 1856–1866 (2019)
14. Lee, D.H., et al.: Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In: ICML. vol. 3, p. 896 (2013)
15. Lee, H., Jeong, W.K.: Scribble2Label: Scribble-supervised cell segmentation via self-generating pseudo-labels with consistency. In: MICCAI. pp. 14–23. Springer (2020)
16. Lin, D., Dai, J., Jia, J., He, K., Sun, J.: ScribbleSup: Scribble-supervised convolutional networks for semantic segmentation. In: CVPR. pp. 3159–3167 (2016)
17. Liu, X., Yuan, Q., Gao, Y., He, K., Wang, S., Tang, X., Tang, J., Shen, D.: Weakly supervised segmentation of covid19 infection with scribble annotation on CT images. PR 122, 108341 (2022)
18. Luo, W., Yang, M.: Semi-supervised semantic segmentation via strong-weak dual-branch network. In: ECCV. pp. 784–800. Springer (2020)
19. Luo, X., Chen, J., Song, T., Wang, G.: Semi-supervised medical image segmentation through dual-task consistency. AAAI 35(10), 8801–8809 (2021)
20. 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. pp. 318–329 (2021)
21. Luo, X., Wang, G., Song, T., Zhang, J., Aertsen, M., Deprest, J., Ourselin, S., Vercauteren, T., Zhang, S.: MIDeepSeg: Minimally interactive segmentation of unseen objects from medical images using deep learning. MedIA 72, 102102 (2021)
22. Mumford, D.B., Shah, J.: Optimal approximations by piecewise smooth functions and associated variational problems. Communications on pure and applied mathematics (1989)
23. Ouali, Y., Hudelot, C., Tami, M.: Semi-supervised semantic segmentation with cross-consistency training. In: CVPR. pp. 12674–12684 (2020)
24. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al.: Pytorch: An imperative style, high-performance deep learning library. In: NeurIPS. pp. 8026–8037 (2019)
25. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: MICCAI. pp. 234–241 (2015)
26. Rother, C., Kolmogorov, V., Blake, A.: GrabCut: interactive foreground extraction using iterated graph cuts. TOG 23(3), 309–314 (2004)
27. Tang, M., Perazzi, F., Djelouah, A., Ben Ayed, I., Schroers, C., Boykov, Y.: On regularized losses for weakly-supervised cnn segmentation. In: ECCV. pp. 507–522 (2018)
28. Tarvainen, A., Valpola, H.: Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In: NeurIPS. pp. 1195–1204 (2017)
29. Valvano, G., Leo, A., Tsaftaris, S.A.: Learning to segment from scribbles using multi-scale adversarial attention gates. TMI (2021)
30. Vezhnevets, V., Konouchine, V.: GrowCut: Interactive multi-label ND image segmentation by cellular automata. Graphicon 1(4), 150–156 (2005)
31. 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)
32. Wang, G., Zuluaga, M.A., Pratt, R., Aertsen, M., Doel, T., Klusmann, M., David, A.L., Deprest, J., Vercauteren, T., Ourselin, S.: Slic-Seg: A minimally interactive segmentation of the placenta from sparse and motion-corrupted fetal MRI in multiple views. MedIA 34, 137–147 (2016)

33. Wang, X., Gao, J., Long, M., Wang, J.: Self-tuning for data-efficient deep learning. In: ICML. pp. 10738–10748. PMLR (2021)
34. Wu, Y., Xu, M., Ge, Z., Cai, J., Zhang, L.: Semi-supervised left atrium segmentation with mutual consistency training. In: MICCAI. pp. 297–306. Springer (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: MICCAI. pp. 605–613. Springer (2019)
36. Yushkevich, P.A., Piven, J., Hazlett, H.C., Smith, R.G., Ho, S., Gee, J.C., Gerig, G.: User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. Neuroimage 31(3), 1116–1128 (2006)
37. Zhang, H., Cisse, M., Dauphin, Y.N., Lopez-Paz, D.: mixup: Beyond empirical risk minimization. arXiv preprint arXiv:1710.09412 (2017)
38. 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: MICCAI. pp. 408–416. Springer (2017)