Weakly Supervised Medical Image Segmentation via Superpixel-Guided Scribble Walking and Class-Wise Contrastive Regularization

Meng Zhou, Zhe Xu(✉), Kang Zhou, and Raymond Kai-yu Tong(✉)

The Chinese University of Hong Kong, Hong Kong, China
jackxz@link.cuhk.edu.hk
kytong@cuhk.edu.hk

Abstract. Deep learning-based segmentation typically requires a large amount of data with dense manual delineation, which is both time-consuming and expensive to obtain for medical images. Consequently, weakly supervised learning, which attempts to utilize sparse annotations such as scribbles for effective training, has garnered considerable attention. However, such scribble-supervision inherently lacks sufficient structural information, leading to two critical challenges: (i) while achieving good performance in overall overlap metrics such as Dice score, the existing methods struggle to perform satisfactory local prediction because no desired structural priors are accessible during training; (ii) the class feature distributions are inevitably less-compact due to sparse and extremely incomplete supervision, leading to poor generalizability. To address these, in this paper, we propose the SC-Net, a new scribble-supervised approach that combines **S**uperpixel-guided scribble walking with **C**lass-wise contrastive regularization. Specifically, the framework is built upon the recent dual-decoder backbone design, where predictions from two slightly different decoders are randomly mixed to provide auxiliary pseudo-label supervision. Besides the sparse and pseudo supervision, the scribbles walk towards unlabeled pixels guided by superpixel connectivity and image content to offer as much dense supervision as possible. Then, the class-wise contrastive regularization disconnects the feature manifolds of different classes to encourage the compactness of class feature distributions. We evaluate our approach on the public cardiac dataset ACDC and demonstrate the superiority of our method compared to recent scribble-supervised and semi-supervised learning methods with similar labeling efforts.

Keywords: Weakly-supervised Learning · Segmentation · Superpixel

1 Introduction

Accurately segmenting cardiac images is crucial for diagnosing and treating cardiovascular diseases. Recently, deep learning methods have greatly advanced

M. Zhou and Z. Xu—Equal contribution.

H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14221, pp. 137–147, 2023.
https://doi.org/10.1007/978-3-031-43895-0_13

cardiac image segmentation. However, most state-of-the-art segmentation models require a large scale of training samples with pixel-wise dense annotations, which are expensive and time-consuming to obtain. Thus, researchers are active in exploring other labour-efficient forms of annotations for effective training. For example, semi-supervised learning (SSL) [14,20,22,26–29,31] is one such approach that attempts to propagate labels from the limited labeled data to the abundant unlabeled data, typically via pseudo-labeling. However, due to limited diversity in the restricted labeled set, accurately propagating labels is very challenging [15]. As another form, weakly supervised learning (WSL), i.e., our focused scenario, utilizes sparse labels such as scribbles, bounding boxes, and points for effective training, wherein scribbles have gained significant attention due to their ease of annotation and flexibility in labeling irregular objects. Yet, an intuitive challenge is that the incomplete shape of cardiac in scribble annotations inherently lacks sufficient structural information, as illustrated in Fig. 1, which easily leads to (i) poor local prediction (e.g., poor boundary prediction with high 95% Hausdorff Distance) because no structural priors are provided during training; (ii) poor generalizability due to less-compact class feature distributions learned from extremely sparse supervision. Effectively training a cardiac segmentation model using scribble annotations remains an open challenge.

Related Work. A few efforts, not limited to medical images, have been made in scribble-supervised segmentation [4,5,9,11,13,15,19,32]. For example, Tang et al. [19] introduced a probabilistic graphical model, conditional random field (CRF), to regularize the spatial relationship between neighboring pixels in an image. Kim et al. [9] proposed another regularization loss based on level-set [17] to leverage the weak supervision. S2L [11] leverages label filtering to improve the pseudo labels generated by the scribble-trained model. USTM [13] adapts an uncertainty-aware mean-teacher [31] model in semi-supervised learning to leverage the unlabeled pixels. Zhang et al. [32] adapted a positive-unlabeled learning framework into this problem assisted by a global consistency term. Luo et al. [15] proposed a dual-decoder design where predictions from two slightly different decoders are randomly mixed to provide more reliable auxiliary pseudo-label supervision. Despite its effectiveness to some extent, the aforementioned two challenges still have not received adequate attention.

In this paper, we propose SC-Net, a new scribble-supervised approach that combines **S**uperpixel-guided scribble walking with **C**lass-wise contrastive regularization. The basic framework is built upon the recent dual-decoder backbone design [15]. Besides the sparse supervision (using partial cross-entropy loss) from scribbles, predictions from two slightly different decoders are randomly mixed to provide auxiliary pseudo-label supervision. This design helps to prevent the model from memorizing its own predictions and falling into a trivial solution during optimization. Then, we tackle the aforementioned inherent challenges with two schemes. Firstly, we propose a specialized mechanism to guide the scribbles to walk towards unlabeled pixels based on superpixel connectivity and image content, in order to augment the structural priors into the labels themselves. As such, better local predictions are achieved. Secondly, we propose a class-wise

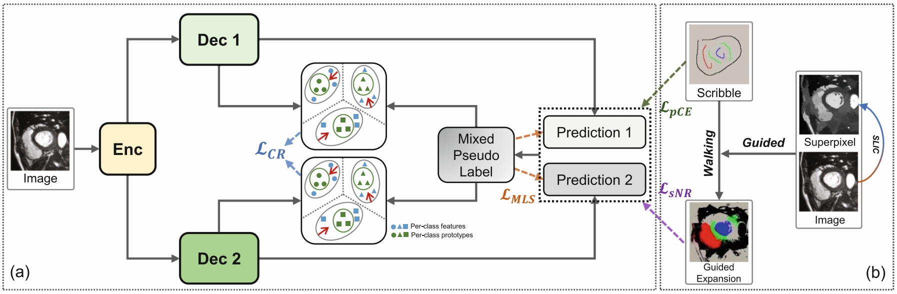

Fig. 1. Overview of SC-Net for scribble-supervised cardiac segmentation. (a) The framework consists of a shared encoder (Enc) and two independent and different decoders (Dec 1 and Dec 2). Structural priors are enriched by (b) superpixel-guided scribble walking strategy (Sect. 2.2). The class-wise contrastive regularization $\mathcal{L}_{CR}$ (Sect. 2.3) encourages the compactness of feature manifolds of different classes.

contrastive regularization term that leverages prototype contrastive learning to disconnect the feature manifolds of different classes, which addresses the issue of less-compact class feature distributions due to sparse supervision. We evaluate our approach on the public cardiac dataset ACDC and show that it achieves promising results, especially better boundary predictions, compared to recent scribble-supervised and semi-supervised methods with similar labeling efforts.

2 Methods

2.1 Preliminaries and Basic Framework

In the scribble-supervised setting, the dataset includes images and their corresponding scribble annotations. We denote an image as X with the scribble annotation $S = \{(s_r, y_r)\}$, where s_r is the pixel of scribble r, and $y_r \in \{0, 1, ..., C-1\}$ denotes the corresponding label with C possible classes at pixel s_r. As shown in Fig. 1, our framework is built upon a one-encoder-dual-decoder design [15], where the encoder (θ_{enc}) is shared and two decoders are independent and slightly different. Here, we denote the decoder 1 as θ_{Dec1} and the auxiliary decoder 2 as θ_{Dec2}. Compared to θ_{Dec1}, θ_{Dec2} introduces the dropout layer (ratio = 0.5) before each convolutional block to impose perturbations. In this framework, the supervised signals consist of a scribble-supervised loss and a pseudo-supervised self-training loss. For the former one, we adopt the commonly used partial cross-entropy loss for those scribble-containing pixels [11,19], formulated as:

$$\mathcal{L}_{pCE} = -0.5 \times (\sum_{c}\sum_{i \in S} \log\, p^{c}_{1(i)} + \sum_{c}\sum_{i \in S} \log\, p^{c}_{2(i)}), \qquad (1)$$

where $p^{c}_{1(i)}$ and $p^{c}_{2(i)}$ are the predicted probability of pixel i belonging to class c from the two decoders θ_{Dec1} and θ_{Dec2}, respectively. For the self-training loss, this dual-decoder framework randomly mix the predictions from the two different decoders to generate the ensemble pseudo label as: $\hat{y}_{ML} = \text{argmax}[\alpha \times p_1 + (1 - \alpha) \times p_2$, where $\alpha = random(0, 1)$. Such dynamically mixing scheme can increase the diversity of pseudo labels, which helps to prevent the model from memorizing its own single prediction and falling into a trivial solution during optimization [8]. As such, the self-training loss can be formulated as:

$$\mathcal{L}_{MLS} = 0.5 \times (\mathcal{L}_{Dice}(\hat{y}_{ML}, p_1) + \mathcal{L}_{Dice}(\hat{y}_{ML}, p_2)). \tag{2}$$

Despite its effectiveness, this framework still overlooks the aforementioned fundamental limitations of sparse scribble supervision: (i) although the mixed pseudo labels provide dense supervision, they still stems from the initial sparse guidance, making it difficult to provide accurate local structural information. Thus, we propose superpixel-guided scribble walking strategy (Sect. 2.2) to enrich structural priors for the initial supervision itself. (ii) Extremely sparse supervision inevitably leads to less-compact class feature distributions, resulting in poor generalizability to unseen test data. Thus, we further propose class-wise contrastive regularization (Sect. 2.3) to enhance the compactness of class embeddings.

2.2 Superpixel-Guided Scribble Walking

In order to enhance the structural information in our initial supervision, we utilize the superpixel of the image as a guide for propagating scribble annotations to unlabeled pixels, considering that it effectively groups pixels with similar characteristics within the uniform regions of an image and helps capture the class boundaries [30]. Specifically, we employ the simple linear iterative clustering (SLIC) algorithm [1] to generate the superpixels. The algorithm works by first dividing the image into a grid of equally-sized squares, then selecting a number of seed points within each square based on the desired number K of superpixels. Next, it iteratively assigns each pixel to the nearest seed point based on its color similarity and spatial proximity (distance). This process is repeated until the clustering converges or reaches a predefined number of iterations. Finally, the algorithm updates the location of the seed points to the centroid of the corresponding superpixel, and repeats until convergence. As such, the image is coarsely segmented into K clusters. To balance accuracy and computational efficiency, the number of iterations is empirically set to 10. K is set to 150. An example of superpixel is depicted in Fig. 1.

Then, guided by the obtained superpixel, the scribbles walk towards unlabeled pixels with the following mechanisms: (i) if the superpixel cluster overlaps with a scribble s_r, the label y_r of s_r walks towards to the pixels contained in this cluster; (ii) yet, if the superpixel cluster does not overlap any scribble or overlaps more than one scribble, the pixels within this cluster are not assigned any labels. As such, we denote the set of the superpixel-guided expanded label as $\{(x_{sp}, \hat{y}_{sp})\}$, where x_{sp} represents the pixel with the corresponding label $\hat{y}_{sp} \in \{0, 1, ..., C-1\}$. An expansion example can be also found in Fig. 1. Although we use strict walking constraints to expand the labels, superpixels are primarily based on color similarity and spatial proximity to seed points.

However, magnetic resonance imaging has less color information compared to natural images, and different organs often share similar intensity, leading to some inevitable label noises. Therefore, to alleviate the negative impact of the label noises, we adopt the noise-robust Dice loss [24] to supervise the models, formulated as:

$$\mathcal{L}_{sNR} = 0.5 \times (\frac{\sum_i^N |p_{1(i)} - \hat{y}_{sp(i)}|^\gamma}{\sum_i^N p_{1(i)}^2 + \sum_i^N \hat{y}_{sp(i)}^2 + \epsilon} + \frac{\sum_i^N |p_{2(i)} - \hat{y}_{sp(i)}|^\gamma}{\sum_i^N p_{2(i)}^2 + \sum_i^N \hat{y}_{sp(i)}^2 + \epsilon}), \tag{3}$$

where N is the number of label-containing pixels. $\hat{y}_{sp}$ is converted to one-hot representation. $p_{1(i)}$ and $p_{2(i)}$ are the predicted probabilities of pixel i from θ_{Dec1} and θ_{Dec2}, respectively. Following [24], $\epsilon = 10^{-5}$ and $\gamma = 1.5$. Note that when $\gamma = 2$, this loss will degrade into the typical Dice loss.

2.3 Class-Wise Contrastive Regularization

When using extremely sparse supervision, it is difficult for the model to learn compact class feature distributions, leading to poor generalizability. To address this, we propose a class-wise contrastive regularization term that leverages prototype contrastive learning to disconnect the feature manifolds of different classes, as illustrated in Fig. 1. Specifically, using the additional non-linear projection head, we derive two sets of projected features, namely F_1 and F_2, from decoder 1 and decoder 2, respectively. Then, we filter the projected features by comparing their respective categories with that of the mixed pseudo label $\hat{y}_{ML}$ and the current predictions from the two decoders. Only features that have matching categories are retained and denoted as $\dot{F}_1^c$ and $\dot{F}_2^c$, where superscript c indicates that such feature vectors correspond to class c. Then, we use C attention modules [2] to obtain ranking scores to sort the retained features and then the top-k features are selected as the class prototypes, where the class-c prototypes are denoted as $Z_1^c = \{z_1^c\}$ and $Z_2^c = \{z_2^c\}$. Note that we extract feature prototypes in an online fashion instead of retaining cross-epoch memories as in [2], since the latter can be computationally inefficient and memory-intensive. Then, we extract the features of each category $f_1^c \in F_1$ and $f_2^c \in F_2$ using the current predictions and encourage their proximity to the corresponding prototypes z_1^c and z_2^c. We adopt the cosine similarity to measure the proximity between the class features and the class prototypes. Taking decoder 1 as example, we define its class-wise contrastive regularization loss $\mathcal{L}_{CR}^{Dec1}$ as:

$$\mathcal{L}_{CR}^{Dec1}(f_1^c, Z_1^c) = \frac{1}{C}\frac{1}{N_{z_1^c}}\frac{1}{N_{f_1^c}}\sum_{c=1}^{C}\sum_{i=1}^{N_{z_1^c}}\sum_{j=1}^{N_{f_1^c}} w_{ij}(1 - \frac{< z_1^{c(i)}, f_1^{c(j)} >}{||z_1^{c(i)}||_2 \cdot ||f_1^{c(j)}||_2}), \tag{4}$$

where w_{ij} is obtained by normalizing the learnable attention weights (detailed in [2]). $N_{z_1^c}$ or $N_{f_1^c}$ is the number of prototypes or projected features of c-th class, respectively. Similarly, we obtain such regularization loss for decoder 2, denoted as $\mathcal{L}_{CR}^{Dec2}$. As such, the overall class-wise contrastive regularization loss is formulated as:

$$\mathcal{L}_{CR} = 0.5 \times (\mathcal{L}_{CR}^{Dec1} + \mathcal{L}_{CR}^{Dec2}). \tag{5}$$

Overall, the final loss of our SC-Net is summarized as:

$$\mathcal{L} = \mathcal{L}_{pCE} + \lambda_{MLS}\mathcal{L}_{MLS} + \lambda_{sNR}\mathcal{L}_{sNR} + \lambda_{CR}\mathcal{L}_{CR}, \tag{6}$$

where λ_{MLS}, λ_{sNR} and λ_{CR} are the trade-off weights. λ_{MLS} is set to 0.5, following [15]. λ_{sNR} is set to 0.005. λ_{CR} is scheduled with an iteration-dependent ramp-up function [10] with the maximal value of 0.01 suggested by [25].

3 Experiments and Results

Dataset. We evaluate our method on the public ACDC dataset [3], which consists of 200 short-axis cine-MRI scans from 100 patients. Each patient has two annotated scans from end-diastolic (ED) and end-systolic (ES) phases, where each scan has three structure labels, including right ventricle (RV), myocardium (Myo) and left venterícle (LV). The scribbles used in this work are manually annotated by Valvano et al. [21]. Considering the large thickness in this dataset, we perform 2D segmentation rather than 3D segmentation, following [15,21].

Implementation and Evaluation Metrics. The framework is implemented with PyTorch using an NVIDIA RTX 3090 GPU. We adopt UNet [18] as the backbone with extension to dual-branch design [15]. All the 2D slices are normalized to [0, 1] and resized to 256×256 pixels. Data augmentations, including random rotation, flipping and noise injection, are applied. The SGD optimizer is utilized with the momentum of 0.9 and weight decay is 10^{-4}, the poly learning rate strategy is employed [16]. We train the segmentation model for 60,000 iterations in total with a batch size of 12. During inference, the encoder in combination with the primary decoder (Dec 1) is utilized to segment each scan slice-by-slice and stack the resulting 2D slice predictions into a 3D volume. We adopt the commonly used Dice Coefficient (DSC) and 95% Hausdorff Distance (95HD) as the evaluation metrics. Five-fold cross-validation is employed. The code will be available at https://github.com/Lemonzhoumeng/SC-Net.

Comparison Study. We compare our proposed SC-Net with recent state-of-the-art alternative methods for annotation-efficient learning. Table 1 presents the quantitative results of different methods. All methods are implemented with the same backbone to ensure fairness. According to [15,21], the cost of scribble annotation for the entire ACDC training set is similar to that of dense pixel-level annotation for 10% of the training samples. Thus, we use 10% of the training samples (8 patients) as labeled data and the remaining 90% as unlabeled data to perform semi-supervised learning (SSL). Here, we compare popular semi-supervised approaches, including AdvEnt [23], DAN [34], MT [20] and UAMT [31], as well as the supervised-only (Sup) baseline (using 10% densely labeled data only). As observed, SC-Net achieves significantly better performance than

Table 1. Quantitative results of different methods via five-fold cross-validation. Standard deviations are shown in parentheses. The best mean results are marked in **bold**.

Setting	Method	RV		Myo		**LV**		Mean	
		DSC	95HD	DSC	95HD	DSC	95HD	DSC	95HD
SSL	AdvEnt [23] (10% label)	0.614(0.256)	21.3(19.8)	0.758(0.149)	8.4(8.6)	0.843(0.134)	12.4(19.3)	0.737(0.179)	14.0(15.9)
	DAN [34] (10% label)	0.655(0.260)	21.2(20.4)	0.768(0.171)	9.5(11.7)	0.833(0.178)	14.9(19.5)	0.752(0.203)	15.2(17.2)
	MT [20] (10% label)	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 [33] (10% label)	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 [12] (lower bound)	0.628(0.110)	178.5(27.1)	0.602 (0.090)	176.0(21.8)	0.710(0.142)	168.1(45.7)	0.647(0.114)	174.2(31.5)
	RW [6]	0.829(0.094)	12.4(19.5)	0.708(0.093)	12.9(8.6)	0.747(0.130)	12.0(14.8)	0.759(0.106)	12.5(14.3)
	USTM [13]	0.824(0.103)	40.4(47.8)	0.739 (0.075)	133.4(42.9)	0.782(0.178)	140.4(54.6)	0.782(0.121)	104.7(48.4)
	S2L [11]	0.821(0.097)	16.8(24.4)	0.786(0.067)	65.6(45.6)	0.845(0.127)	66.5(56.5)	0.817(0.097)	49.6(42.2)
	MLoss [9]	0.807(0.089)	13.4(21.1)	0.828(0.057)	29.8(41.5)	0.868(0.074)	55.1(61.6)	0.834(0.073)	32.8(41.4)
	EM [7]	0.815(0.119)	37.9(54.5)	0.803(0.059)	56.9(53.2)	0.887(0.071)	50.4(57.9)	0.834(0.083)	48.5(55.2)
	DBMS [15]	0.861(0.087)	8.3(13.0)	0.835(0.057)	10.3(19.7)	0.899(0.062)	11.0(20.9)	0.865(0.078)	9.9(17.8)
	Ours (w/o $\mathcal{L}_{sNR}$)	0.847(0.086)	7.8(13.8)	0.823(0.091)	8.9(18.5)	0.902(0.077)	10.4(18.4)	0.858(0.093)	8.9(16.9)
	Ours (w/o $\mathcal{L}_{CR}$)	0.850(0.079)	8.3(14.3)	0.819(0.078)	9.2(17.3)	0.889(0.058)	10.7(15.7)	0.853(0.076)	9.3(16.3)
	Ours (SC-Net)	**0.862(0.071)**	**4.6(3.8)**	**0.839(0.088)**	**6.7(16.3)**	**0.915(0.083)**	**8.1(14.1)**	**0.872(0.063)**	**6.5(13.9)**
SL	Sup (10% label)	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)
	Sup (full) (upper bound)	0.881(0.093)	6.9(10.9)	0.879 (0.039)	5.8(15.4)	0.935(0.065)	8.0(19.9)	0.898(0.066)	6.9(15.4)

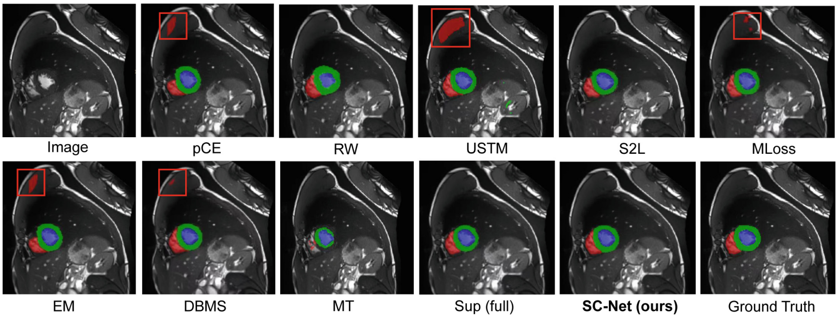

Fig. 2. Qualitative comparison of different methods.

the competing SSL methods, showing that when the annotation budget is similar, using scribble annotations can lead to better outcomes than pixel-wise annotations. Furthermore, we compare SC-Net with weakly-supervised learning (WSL) approaches on scribble annotated data, including pCE only [12] (lower bound), RW [6] (using random walker to produce additional label), USTM [13] (uncertainty-aware self-ensembling and transformation-consistent model), S2L [11] (Scribble2Label), MLoss [9] (Mumford-shah loss), EM [7] (entropy minimization) and DBMS [15] (dual-branch mixed supervision). Besides, the upper bound, i.e., supervised training with full dense annotations, is also presented. It can be observed that SC-Net achieves more promising results compared to existing methods. In comparison to DBMS, SC-Net exhibits a slight improvement in DSC, but a significant decrease in the 95HD metric (p<0.05). Furthermore, our method achieves slightly lower performance in DSC compared to the upper bound, but even slightly better results in 95HD. This indicates that our approach effectively addresses the inherent limitations of sparse supervision. Figure 2 presents exemplar qualitative results of our SC-Net and other methods. It can

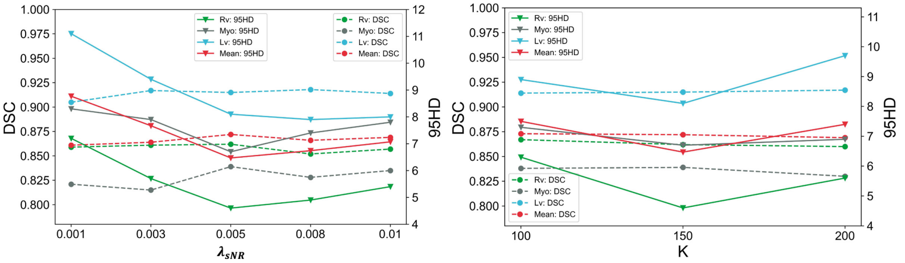

Fig. 3. Sensitivity analysis of λ_{sNR} and the cluster number K in superpixel generation.

be seen that the prediction of our SC-Net fit more accurately with the ground truth, especially in local details. Thanks to the more compact feature distributions, our method reduces false-positive predictions, as indicated in the red box.

Ablation Study. We perform an ablation study to investigate the effects of the two key components of our SC-Net. The results are also shown in Table 1. We found that the two components need to work together. When we remove $\mathcal{L}_{sNR}$, the performance degrades to some extent. This may be because it is difficult for the model to generate high-quality local pseudo-labels with only sparse supervision provided by scribbles, and class-wise contrastive regularization relies heavily on pseudo labels to separate class features. When we remove $\mathcal{L}_{CR}$, the performance also drops slightly. This is mainly because the generated superpixels inevitably contain errors, which can misguide the scribble walking. Yet, using $\mathcal{L}_{CR}$ can regularize the feature distribution between classes, reducing the impact of these errors. Meanwhile, the structure prior strengthened by superpixel guidance helps to provide higher-quality local pseudo labels to assist class-wise contrastive regularization. The two components complement each other, resulting in the best performance of our complete SC-Net.

Sensitivity Analysis. The superpixel-guided scribble walking plays an important role in our SC-Net. Thus, we conduct further assessments on the sensitivity of λ_{sNR}, which is used to weight $\mathcal{L}_{sNR}$, and the cluster number K used for superpixel generation. The results obtained by five-fold cross validation are presented in Fig. 3. As observed, increasing λ_{sNR} from 0.001 to 0.005 leads to better results in terms of both metrics. When λ_{sNR} is set to 0.01, the result exhibits only a slight decrease compared to 0.005 (0.872 vs. 0.867 in term of DSC). These observations show that our method is not so sensitive to λ_{sNR} within the empirical range. In practice, the optimal value of K depends on the characteristics of the input image, such as object complexity and texture. We find that our method is also not highly sensitive to K, but the optimal results are achieved when $K = 150$ for the cardiac MR images in our study.

4 Conclusion

In this work, we proposed the SC-Net towards effective weakly supervised medical image segmentation using scribble annotations. By combining superpixel-guided scribble walking with class-wise contrastive regularization, our approach alleviates two inherent challenges caused by sparse supervision, i.e., the lack of structural priors during training and less-compact class feature distributions. Comprehensive experiments on the public cardiac dataset ACDC demonstrated the superior performance of our method compared to recent scribble-supervised and semi-supervised methods with similar labeling efforts.

Acknowledgement. This research was supported by General Research Fund from Research Grant Council of Hong Kong (No. 14205419).

References

1. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S.: Slic superpixels compared to state-of-the-art superpixel methods. IEEE Trans. Pattern Anal. Mach. Intell. **34**(11), 2274–2282 (2012)
2. Alonso, I., Sabater, A., Ferstl, D., Montesano, L., Murillo, A.C.: Semi-supervised semantic segmentation with pixel-level contrastive learning from a class-wise memory bank. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 8219–8228 (2021)
3. 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)
4. Can, Y.B., et al.: Learning to segment medical images with scribble-supervision alone. In: Stoyanov, D., et al. (eds.) DLMIA/ML-CDS -2018. LNCS, vol. 11045, pp. 236–244. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00889-5_27
5. Chen, Q., Hong, Y.: Scribble2d5: Weakly-supervised volumetric image segmentation via scribble annotations. In: Medical Image Computing and Computer Assisted Intervention-MICCAI 2022: 25th International Conference, Singapore, 18–22 September 2022, Proceedings, Part VIII, pp. 234–243. Springer (2022). https://doi.org/10.1007/978-3-031-16452-1_23
6. Grady, L.: Random walks for image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **28**(11), 1768–1783 (2006)
7. Grandvalet, Y., Bengio, Y.: Semi-supervised learning by entropy minimization. In: Advances in Neural Information Processing Systems 17 (2004)
8. Huo, X., et al.: Atso: asynchronous teacher-student optimization for semi-supervised image segmentation. In: CVPR, pp. 1235–1244 (2021)
9. Kim, B., Ye, J.C.: Mumford-shah loss functional for image segmentation with deep learning. IEEE Trans. Image Process. **29**, 1856–1866 (2019)
10. Laine, S., Aila, T.: Temporal ensembling for semi-supervised learning. arXiv preprint arXiv:1610.02242 (2016)
11. 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

12. 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)
13. Liu, X.: Weakly supervised segmentation of covid19 infection with scribble annotation on ct images. Pattern Recogn. **122**, 108341 (2022)
14. Luo, X., Chen, J., Song, T., Chen, Y., Wang, G., Zhang, S.: Semi-supervised medical image segmentation through dual-task consistency. In: AAAI Conference on Artificial Intelligence (2021)
15. Luo, X., et al.: Scribble-supervised medical image segmentation via dual-branch network and dynamically mixed pseudo labels supervision. In: Medical Image Computing and Computer Assisted Intervention. pp. 528–538. Springer (2022). https://doi.org/10.1007/978-3-031-16431-6_50
16. 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
17. Mumford, D.B., Shah, J.: Optimal approximations by piecewise smooth functions and associated variational problems. In: Communications on Pure and Applied Mathematics (1989)
18. 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
19. Tang, M., Perazzi, F., Djelouah, A., Ayed, I.B., Schroers, C., Boykov, Y.: On regularized losses for weakly-supervised CNN segmentation. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) ECCV 2018. LNCS, vol. 11220, pp. 524–540. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01270-0_31
20. 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, pp. 1195–1204 (2017)
21. 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)
22. Verma, V., et al.: Interpolation consistency training for semi-supervised learning. Neural Netw. **145**, 90–106 (2022)
23. Vu, T.H., Jain, H., Bucher, M., Cord, M., Pérez, P.: ADVENT: adversarial entropy minimization for domain adaptation in semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2517–2526 (2019)
24. Wang, G., et al.: A noise-robust framework for automatic segmentation of COVID-19 pneumonia lesions from CT images. IEEE Trans. Med. Imaging **39**(8), 2653–2663 (2020)
25. Wu, Y., Wu, Z., Wu, Q., Ge, Z., Cai, J.: Exploring smoothness and class-separation for semi-supervised medical image segmentation. In: International Conference on Medical Image Computing and Computer Assisted Intervention. Springer (2022). https://doi.org/10.1007/978-3-031-16443-9_4
26. 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

27. Xu, Z., et al.: Anti-interference from noisy labels: mean-teacher-assisted confident learning for medical image segmentation. IEEE Trans. Med. Imaging (2022)
28. Xu, Z., et al.: Noisy labels are treasure: mean-teacher-assisted confident learning for hepatic vessel segmentation. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12901, pp. 3–13. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87193-2_1
29. Xu, Z., et al.: All-around real label supervision: Cyclic prototype consistency learning for semi-supervised medical image segmentation. IEEE J. Biomed. Health Inform. (2022)
30. Yi, S., Ma, H., Wang, X., Hu, T., Li, X., Wang, Y.: Weakly-supervised semantic segmentation with superpixel guided local and global consistency. Pattern Recogn. **124**, 108504 (2022)
31. 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
32. Zhang, K., Zhuang, X.: Shapepu: a new pu learning framework regularized by global consistency for scribble supervised cardiac segmentation. In: Medical Image Computing and Computer Assisted Intervention, pp. 162–172. Springer (2022). https://doi.org/10.1007/978-3-031-16452-1_16
33. Zhang, Y., Jiao, R., Liao, Q., Li, D., Zhang, J.: Uncertainty-guided mutual consistency learning for semi-supervised medical image segmentation. In: Artificial Intelligence in Medicine, p. 102476 (2022)
34. 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