2025 International Conference on Intelligent Computing
July 26-29, Ningbo, China
https://www.ic-icc.cn/2025/index.php

Region Features Propagation with Class-Aware Contrastive Learning for Weakly Supervised Cardiac Segmentation

YuXiao[1,2], Ping Wang[2(✉)], Xiuyang Zhao[2], Dongmei Niu[2], Jinshuo Zhang[3]

[1] Shandong Key Laboratory of Ubiquitous Intelligent Computing, University of Jinan, Jinan, China
[2] School of Information Science and Engineering, University of Jinan, Jinan, China
ise_wangp@ujn.edu.cn
3 School of Mathematics, Shandong University, Jinan, Shandong, China

Abstract. Cardiac segmentation is crucial for analyzing heart structure and function, providing essential support for clinical diagnosis and treatment planning. However, obtaining fully annotated images is both costly and time-consuming. Scribble annotations, which utilize simple lines instead of pixel-wise annotations, offer a cost-effective alternative but lack sufficient information, making segmentation network training challenging. To tackle this challenge, we introduce ScribbleCorrNet (SCN), an innovative architecture designed for medical image segmentation under scribble-based supervision. SCN employs Correlation-Aware Label Enhancement (CALE) strategies, introducing two key mechanisms: (i) pixel affinity propagation (PAP), which propagates high-confidence pixels using pairwise similarities in a correlation map, and (ii) region shape refinement (RSR), which refines pseudo-labels by leveraging shape information encoded in the correlation map. Additionally, a class-aware contrastive learning (CACL) mechanism enhances intra-class consistency and inter-class separation. Experiments on the ACDC2017 and MSCMR datasets demonstrate SCN's superior performance compared to existing scribble-based segmentation methods.

Keywords: Cardiac segmentation, scribble annotation, weakly supervised learning, contrastive learning.

1 Introduction

Cardiac MRI segmentation is crucial for diagnosing cardiac conditions and guiding interventions, requiring accurate delineation of structures such as ventricles and myocardium [1]. fully supervised learning relies on pixel-level labels [2], [3], which are labor-intensive and subject to anatomical differences between patients [4], [5]. Weakly supervised learning (WSL) addresses these challenges by using sparse annotations (e.g., scribbles, image-level labels) to reduce annotation costs while maintaining performance [6-10]. However, sparse annotations such as scribbles limit model performance due to insufficient supervision [11-13]. To address this problem, we designed CALE that integrates two new strategies: pixel affinity propagation (PAP) for capturing fine-

grained pixel relationships and region shape refinement (RSR) for global shape-based pseudo-label refinement. In addition, the confidence-guided threshold mechanism (CGTM) iteratively updates the high-confidence region masks, while the class-aware contrastive learning (CACL) enhances feature discrimination through intra-class consistency and inter-class separability. Our contributions include:

- CALE: We propose PAP to exploit pairwise pixel relationships and Region Shape Refinement (RSR) to integrate global shape information, enhancing pseudo label quality and segmentation accuracy.
- CGTM: We propose CGTM which can iteratively updates high confidence region masks, ensuring reliable supervision with minimal annotation requirements.
- CACL: The proposed CACL enhances the model's ability to distinguish different classes under weakly annotation scene and improved segmentation performance.
- Comprehensive experiments conducted on public cardiac MRI datasets validate the efficacy of SCN. Our method surpasses prior methods based on scribble annotations and various semi-supervised techniques, attaining comparable performance metrics.

2 RELATED WORK

2.1 Weakly-Supervised Segmentation

Instead of relying on labor-intensive pixel-level annotations, necessitating techniques that utilize sparse annotations. Weakly supervised learning (WSL) reduces dependency on dense labels, leveraging sparse inputs like scribbles [15], image-level labels [16], or bounding boxes [17], [18]. Scribble-based methods have gained attention. Luo et al. [8] proposed a dual-branch network with hybrid pseudo-label supervision to improve segmentation quality, while Lin et al. [11] introduced ScribbleSup, enhancing performance with regularization loss. Tang et al. [19] further explored regularization loss for generalization, and Valvano et al. [12] used adversarial learning to stabilize pseudo-label quality. Pseudo-label refinement approaches, such as Scribble2Label [20] and Beyond Weakly Supervised [21], mitigate annotation noise through iterative consistency constraints. However, these methods remain vulnerable to initial pseudo-label quality, limiting their ability to capture global structural information. To address this, we propose a dynamic correlation graph-based label propagation approach, combining pixel-wise local consistency and region-wise global shape refinement. This dual strategy dynamically updates high-confidence masks, improving pseudo-label accuracy and segmentation performance.

2.2 Contrastive Learning

Self-supervised learning (SSL), particularly contrastive learning, offers an effective way to learn feature representations from unlabeled data through unsupervised loss

functions [23]. Contrastive learning pulls semantically similar samples (positive pairs) closer and pushes dissimilar ones (negative pairs) apart in latent space [24], [25]. Early methods like SimCLR [24] relied on data augmentation for positive pairs, but this often blur inter-class boundaries in fine-grained tasks like segmentation. Recent advancements, such as CPC [26] and supervised contrastive learning (SCL) [27], improved feature extraction using true or pseudo-labels. However, these methods struggle with pixel-level detail capture or require high-quality pseudo-labels, limiting their efficacy in weakly supervised scenarios.

To overcome these challenges, we propose a contrastive learning approach based on weakly supervised segmentation. We propose a class-aware contrastive learning (CACL) strategy specifically designed for sparse annotation scenarios. CACL introduces category mean vectors computed from sparse annotations, enabling more robust intra-category feature aggregation and inter-class separation.

3 METHOD

In this section, we introduce ScribbleCorrNet (SCN), a weakly supervised segmentation framework for medical images using sparse scribble annotations. As shown in Fig.1, the framework integrates three core components: pixel affinity propagation (PAP), region shape refinement (RSR), and class-aware contrastive learning (CACL). PAP captures fine-grained pixel relationships to generate pseudo-labels, RSR refines these labels using global shape constraints, and CACL enhances feature discrimination. This combination enables accurate segmentation under scribble annotated scene.

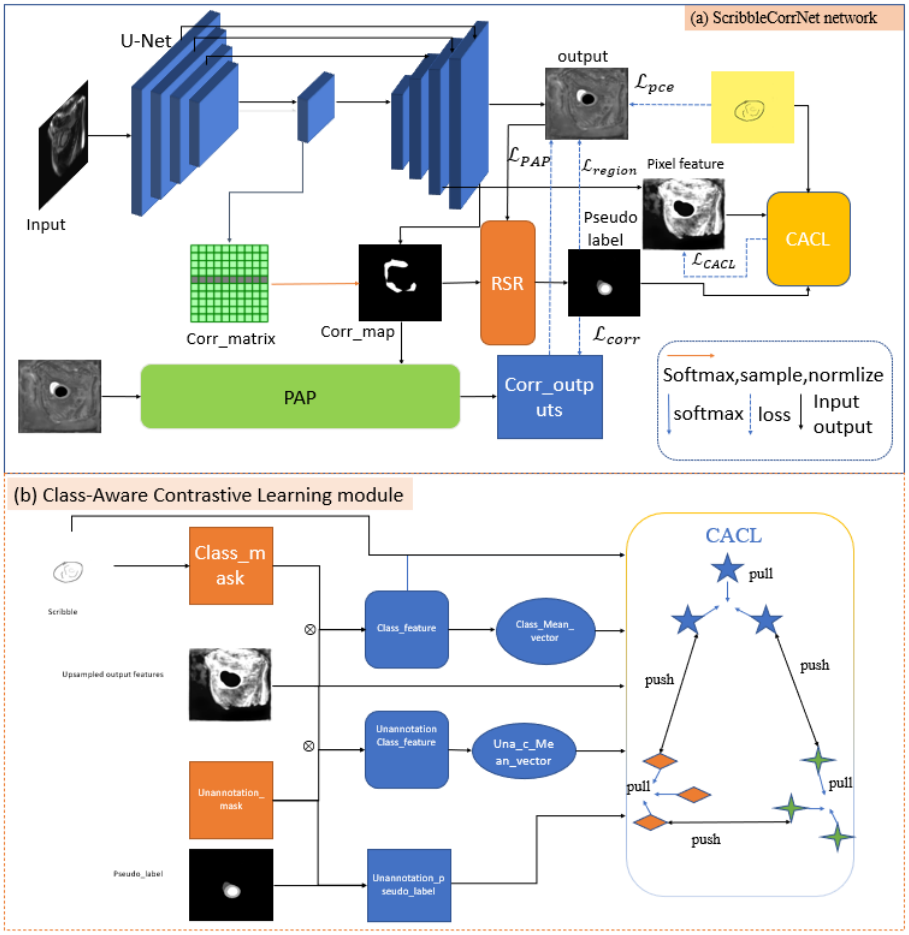

Fig. 1. Overview of the formulated ScribbleCorrNet(SCN) framework. (a) The main network integrates pixel affinity propagation (PAP), region shape refinement (RSR), and class-aware contrastive learning (CACL). (b) The CACL module ensures class-aware feature alignment through contrastive learning.

3.1 Correlation-Aware Label Enhancement

Pixel Affinity Propagation. To propagate sparse scribble annotations to unannotated regions, we propose PAP. Given an encoder-extracted feature representation $f \in R^{H\times W\times D}$, where H, W, and D represent height, width and the number of feature channels, pixel - level semantic affinities are obtained via a correlation map that is softmax - normalized:

$$C_{i,j} = \frac{softmax(f_i \cdot f_j)}{\sqrt{D}} \tag{1}$$

where $f_i \cdot f_j$ represents the dot product of the feature vectors at pixels *i* and *j*. In order to enhance the model's ability to identify pairwise similarities, we embed the information of the correlation matrix *C* into the model's prediction $\widehat{Y}_i$ to obtain the hard pseudo labels $Y_{i,corr}$ after pixel propagation.

$$Y_{i,corr} = C_{i,j} \cdot \widehat{Y}_i \tag{1}$$

here, $C_{i,j}$ denotes the correlation map, and $Y_{i,corr} \in R^{H*W*C}$ integrate semantic similarity to enable smooth, context-aware diffusion of supervisory signals from annotated to unannotated regions. To ensure consistency, we introduce the PAP loss, which measures the discrepancy between the model's predictions $\widehat{Y}_i$ and propagated pseudo-labels $Y_{i,corr}$ via cross-entropy. This loss mitigates the limitations of sparse scribble annotations by aligning predictions with refined pseudo-labels.

$$\mathcal{L}_{PAP} = -\frac{1}{N}\sum_{i=1}^{N} \log\left(\frac{\exp(\widehat{Y}_i, Y_{i,corr})}{\sum_{c=1}^{C}\exp(\widehat{Y}_i)}\right) \tag{3}$$

here, $\widehat{Y}_i$ represents the model predictions for the i-th pixel, and $Y_{i,corr}$ denotes the pseudo-label category index for the i-th pixel. N is the total number of pixels.

Region Shape Refinement. To incorporate shape information from the correlation map C, we propose RSR, which refines pseudo-labels for unlabeled regions using high-confidence region-level statistics. Each row c in C encodes the similarity of a pixel to all others, capturing class-agnostic shape patterns. We refine pseudo-labels by integrating high-confidence region statistics with normalized rows of C, leveraging shape priors to enhance boundary accuracy. This process is formalized as follows:

$$\hat{c} = f_2\left(\mathbb{I}\left(\frac{\hat{c}-\min(c)}{max(c)-\min(c)}\right) > \tau\right) \tag{4}$$

where $f_2(\cdot)$ is a shape-matching function, and $\hat{c} \in R^{H*W}$ represents the binary shape mask, embedding shape information. To identify high-confidence regions, we compute a high confidence mask M_i and calculate the overlap ratio r between $\hat{c}$ and the high-confidence regions within M_i.

$$M_{i,j} = \begin{cases} 1, if \ \max\left(P_{i,j} > \tau_{region}\right) \\ 0, \qquad\qquad otherwise \end{cases} \tag{5}$$

$$r = \frac{|\hat{C} \cap M_i|}{|M_i|} \tag{6}$$

here, the threshold τ_{region} is used to filter high-confidence predicted pixels for generating the pseudo-labels. the shape mask $\hat{c}$ is used to refine the pseudo-labels $F(x)_i$. This process refines the pseudo-labels by identifying the most dominant class c^* within the high-confidence region. The dominant class is determined as follows:

$$c^* = argmax_{l \in L} G(l) \tag{7}$$

here, L represents the set of all unique classes in $F(x)_i$. M_i is the high-confidence mask, and $\mathbb{I}(\cdot)$ is the indicator function. After identifying, the pseudo-labels are updated by

propagating this dominant class into the unlabeled regions, thereby expanding the high-confidence regions:

$$F(x)_i = \begin{cases} c^*, if\ \hat{c}_i = 1, \\ F(x)_i, otherwise, \end{cases} M_i = M_i \cup \hat{c}_i \tag{8}$$

To ensure the effectiveness of updated pseudo-labels, we introduce two complementary loss terms that supervise distinct aspects of model predictions. The first term is the propagation consistency loss, which enforces consistency between propagated outputs $Y_{i,corr}$ and refined pseudo-labels $F(x)_i$. This loss aligns predictions with updated pseudo-labels, ensuring robustness of the propagation process. It is defined as:

$$\mathcal{L}_{PCL} = \frac{1}{|N|}\left(\mathcal{L}_c\left(Y_{i,corr}, F(x)_i\right)\right) \cdot M_i \tag{9}$$

where, $\mathcal{L}_c$ is the cross-entropy loss; $F(x)_i$ is the updated high confidence pseudo-label; $Y_{i,corr}$ is the output pseudo-label after pixel propagation and M_i is the high-confidence pixel mask, indicating the high-confidence pixel location.

The second loss term supervises the model's final predictions $\hat{Y}_i$ using the same updated pseudo-labels $F(x)_i$. Unlike the pixel propagation supervision, this loss directly guides the model's overall segmentation predictions, encouraging consistency between the model's outputs $\hat{Y}_i$ and the refined pseudo labels $F(x)$. The region-level supervision loss is defined as:

$$\mathcal{L}_{region} = \frac{1}{|N|}\left(\mathcal{L}_c\left(\hat{Y}_i, F(x)_i\right)\right) \cdot M_i \tag{10}$$

The total RSR loss is as follows:

$$\mathcal{L}_{RSR} = \mathcal{L}_{PCL} + \mathcal{L}_{region} \tag{11}$$

The RSR strategy enhances pseudo-label reliability and context consistency by integrating shape information from correlation graphs with statistical information from high-confidence regions. This dual approach improves model perception of unlabeled areas and significantly elevates pseudo-label quality, thereby boosting overall segmentation performance.

3.2 Confidence-Guided Thresholding Mechanism

Traditional medical image segmentation methods often rely on fixed thresholds to identify high-confidence regions [30], [31], but these thresholds are suboptimal: overly strict thresholds underutilize unlabeled data, while loose ones degrade prediction accuracy. Inspired by [32], we propose the confidence-guided thresholding mechanism (CGTM), which adaptively updates high-confidence region masks by iteratively correlating each pixel with the entire feature map. This mechanism dynamically aligns thresholds with data characteristics and model predictions, enhancing pseudo-label re-

liability. CGTM further refines masks by integrating shape information from the correlation map and scribble annotations, enabling precise label propagation. This approach effectively captures complex anatomical structures and outperforms static or neighborhood-based methods, achieving superior segmentation performance in challenging medical tasks. After multiple experimental verifications, the initial threshold τ_0 of CGTM is set to 0.85, which not only avoids mistakenly incorporating a large number of low-quality predictions into pseudo-supervision in the early stage, but also is not too harsh, resulting in too few pseudo-labels and difficulty in training. CGTM is updated iteratively based on model output $\hat{Y}_i$. We employ the exponential moving average (EMA) to optimize τ to ensure the smoothness of threshold updates and avoid excessive fluctuations. Each update is defined as:

$$\Delta\tau = \frac{1}{|L|}\Sigma\sum_{l\in L}\max\left[1\left(\hat{Y}_i = l\right)\odot_{max}^{c}\left(\hat{Y}_i\right)\right] \tag{12}$$

where $\max^{c}(\cdot)$ denotes taking the maximum value along the max channel dimension. L is the set of all unique classes. We take the maximum confidence of all classes and take their average as the increment in each iteration. We found that such a threshold update strategy is effective.

3.3 Class-Aware Contrastive Learning

In weakly supervised semantic segmentation with sparse annotations, inter-category discrimination and intra-category consistency are critical for effective feature representation. However, sparse annotations limit direct supervision, leaving unlabeled regions ambiguous. While pseudo-labels help mitigate this issue, they risk introducing noise. To address these challenges, we propose CACL, which leverages category mean vectors to enhance intra-class feature aggregation and inter-class separation, thereby strengthening global feature constraints. Specifically: let $F_i \epsilon R^{B*D*H*W}$ denote the feature output of the model's last upsampling layer, and represent sparse annotations. Here, *B*, *D*, *H*, *W* denote batch size, feature dimension, height, and width, respectively. For each category c, let M_C denote the set of pixels labeled as class c. The category mean vector μ_c is defined as:

$$\mu_c = \frac{\sum_i M_c \odot F_i}{\sum_i M_c + \epsilon} \tag{13}$$

among them, ε is a smoothing term to prevent division by zero. To incorporate both labeled and unlabeled regions, we calculate the contrastive loss separately for the sparse annotations and pseudo labeled regions. For the labeled regions, we calculate the similarity $S_{i,c}$ of each pixel feature F_i with the mean vector of all categories, which is defined as:

$$S_{i,c} = \frac{F_i \cdot \mu_c}{T} \tag{14}$$

Here, the temperature parameter T governs the sharpness of the distribution. We ensure intra-class consistency by encouraging the pixel feature F_i to maximize the similarity with its true class mean vector μ_c:

$$\mathcal{L}_{scribble} = -\frac{1}{M_{y_i}} \sum_{i \in M_{y_i}} log \frac{e^{S_{i,y_i}}}{\sum_{c=1} e^{S_{i,c}}} \tag{15}$$

here, S_{i,y_i} represents the similarity between pixel i and its true category y_i, which is used to measure category consistency. M_{y_i} represents the set of labeled pixels of category y_i. For the unlabeled regions, we use pseudo labels $F(x)_i \epsilon R^{B*H*W}$ which are generated through region propagation and confidence thresholding, to supervise the unannotated regions. The contrastive loss for pseudo-labeled regions is similarly defined as:

$$\mathcal{L}_{unlabel} = -\frac{1}{|M_{\hat{y}_i}|} \sum_{i \in \hat{y}_i} log \frac{e^{S_{i,\hat{y}_i}}}{\sum_{c=1}^{c} e^{S_{i,c}}} \tag{16}$$

here, $S_{i,\hat{y}_i}$ represents the similarity between pixel i and its true category $\hat{y}_i$, which is used to measure category consistency. $M_{\hat{y}_i}$ represents the set of labeled pixels of category $\hat{y}_i$. Finally, we combine the contrastive loss from both labeled and pseudo-labeled regions to define the total CACL as:

$$\mathcal{L}_{CACL} = \mathcal{L}_{scribble} + \mathcal{L}_{unlabel} \tag{17}$$

In the weakly supervised medical image segmentation task, the category-aware contrast learning strategy significantly improves the model performance, especially the performance on sparsely labeled data sets verifies its effectiveness.

3.4 Loss Function

The total loss function can be written as:

$$\mathcal{L}_{total} = \lambda_1 \mathcal{L}_{pce} + \lambda_2 \mathcal{L}_{PAP} + \lambda_3 \mathcal{L}_{RSR} + \mathcal{L}_4 \mathcal{L}_{CACL} \tag{18}$$

here, $\mathcal{L}_{pce}$ denotes the partial cross-entropy loss supervised by scribble annotations. $\mathcal{L}_{pce}$ is defined as follows:

$$\mathcal{L}_{pce} = -\sum_{c} \sum_{i \in \omega_i} y_{i,s}^{c} log y_{i,p}^{c} \tag{19}$$

where, $y_{i,p}^{c}$ is the predicted probability of pixel i belonging to class c, and ω_s is the set of pixels in the scribble. $y_{i,s}^{c}$ represents the label of pixel i for category c in the scribble annotations.

4 EXPERIMENTS

4.1 Dataset

We validate our method on two public datasets, ACDC2017[33] and MSCMR [34].
ACDC2017. Contains 1,902 MRI slices at end-diastole and end-systole of 100 patients, with three cardiac structures (left/right ventricle, myocardium) annotated. Based on previous studies, the dataset is divided into training, validation, and testing parts in a ratio of 6:2:2. Augmentation strategies (random rotation, flipping, and resizing to 256×256) are applied during training.
MSCMR. Contains 686 late gadolinium-enhanced MRI slices of 45 patients with cardiomyopathy. Among them, 20 patients have both scribbles and pixel-level annotations of cardiac structures. Five of these patients are used for validation and 15 patients are used for testing. The 25 patients with only scribble labels are used for training.

4.2 Evaluation Metrics

We quantitatively evaluate segmentation performance using standard metrics: The Dice Similarity Coefficient (DSC) and the 95th percentile Hausdorff Distance (95HD) are employed as evaluation metrics. The DSC quantifies the volumetric agreement between predicted segmentations and reference annotations, while 95HD quantifies the maximum boundary deviation between predictions and labels. Metrics are computed for each cardiac structure and averaged to provide overall performance scores.

4.3 Implementation Details.

Our method is implemented using the U-Net framework and trained on NVIDIA TITAN V GPUs with PyTorch. For the purpose of ensuring comparability and fairness across experiments, all experiments adopted the same training protocol such as the enhancement strategy, and the backbone architecture adopted the backbone architecture of the original method. The experiment only takes seven or eight hours to complete the model training. All input images are resized to 256×256 pixels to match network input requirements. Training employs the Adam optimizer with an initial learning rate of 1×10^{-3}, adjusted dynamically using a StepLR scheduler. The batch size is set to 24, and training proceeds for 30,000 iterations. Optimized loss weights $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ are set to [1, 0.3, 0.25, 0.1] following empirical hyperparameter tuning.

Table 1. Performance comparison on the ACDC2017 dataset

Method	RV		MYO		LV		Mean	
	DSC↑	95HD↓	DSC↑	95HD↓	DSC↑	95HD↓	DSC↑	95HD↓
PCE[13]	66.53	111.26	61.39	104.89	76.14	106.59	68.02	107.58
S2l[19]	86.66	22.03	79.25	48.94	80.31	72.72	82.07	47.89
MLoss[34]	85.67	3.63	83.80	1.33	90.98	3.91	86.82	2.96
RLoss[35]	86.98	3.13	81.94	2.46	91.80	3.37	86.91	2.99
MELoss[37]	86.94	12.09	81.35	23.65	90.05	23.08	86.12	19.61
USTM[33]	86.52	6.61	75.52	65.21	80.55	72.91	80.86	48.24
DMPLS[8]	86.1	7.90	84.20	9.70	91.3	12.10	87.20	9.90
Fddseg[36]	85.82	1.49	85.95	2.27	90.32	5.41	87.36	3.05
ours	87.81	1.33	85.96	3.53	91.76	4.23	88.51	3.03

Table 2. Performance comparison on the MSCMR dataset

Method	RV		MYO		LV		Mean	
	DSC↑	95HD↓	DSC↑	95HD↓	DSC↑	95HD↓	DSC↑	95HD↓
PCE[13]	77.49	181.95	57.06	206.50	73.72	201.36	69.42	196.60
S2l[19]	77.79	169.33	81.80	24.59	89.04	88.37	82.88	94.29
MLoss[34]	87.28	3.04	83.39	2.39	92.28	2.30	87.65	2.58
RLoss[35]	86.98	3.13	81.94	2.46	91.80	3.37	86.91	2.99
MELoss[37]	87.07	32.39	82.70	19.95	90.74	54.06	86.84	35.47
USTM[33]	72.63	186.37	58.12	186.29	80.99	171.72	70.58	180.46
DMPLS[8]	87.59	3.11	82.20	2.51	91.99	2.30	87.25	6.15
Fddseg[36]	88.73	2.56	85.11	2.38	92.41	2.67	88.75	2.53
ours	88.65	2.31	85.25	2.29	92.92	2.32	88.94	2.31

4.4 Comparison with Other Methods

We evaluate our method against eight state-of-the-art methods ([13], [19], [35–39]) using the same scribble annotations. Quantitative results on the ACDC2017 and MSCMR datasets (Tables 1 and 2) show that our method achieves the highest average Dice similarity coefficient (DSC, $p<0.05$) and 95HD, performing well in leveraging scribble annotations for accurate segmentation.

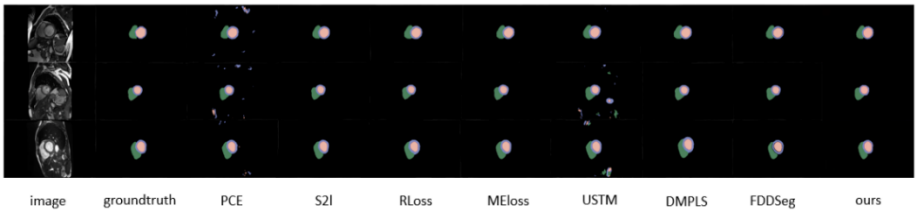

Fig. 2. Visual comparison between our proposed method and other weakly supervised techniques on the ACDC2017 dataset.

As shown in Fig 2, Visual comparisons further highlight our method's advantages. Pixel-level contrastive learning enhances feature discrimination, enabling better separation of similar and dissimilar regions. Additionally, dynamic threshold updates and region propagation strategies generate accurate high-confidence pseudo-labels, improving boundary precision in challenging anatomical regions. These mechanisms collectively ensure superior structural consistency, outperforming other methods in segmentation tasks.

Table 3. ABLATION EXPERIMENT on the ACDC2017 dataset

Method	RV		MYO		LV		Mean	
	DSC↑	95HD↓	DSC↑	95HD↓	DSC↑	95HD↓	DSC↑	95HD↓
PCE[13]	66.53	111.26	61.39	104.89	76.14	106.59	68.02	107.58
PCE+CALE	87.79	1.39	85.75	5.68	91.52	8.01	88.35	5.03
PCE+CALE+ CACL	87.*81*	1.38	85.96	3.48	91.76	4.23	88.51	3.03

4.5 Ablation Study

We conducted ablation experiments on the ACDC2017 datasets (Table 3) to evaluate the contribution of each component in our framework. On ACDC2017, the PCE baseline achieved a mean DSC of 68.02%. Adding the label propagation strategy improved this to 88.35%, demonstrating its effectiveness in enhancing supervision and segmentation performance. Further integrating the class-aware contrastive loss (CACL) increased the mean DSC to 88.51%, underscoring its role in optimizing inter-class feature separation.

5 CONCLUSION

This study presents a weakly supervised method for cardiac substructure segmentation using sparse scribble annotations. Specifically, the CALE module, which includes PAP and RSR, propagates scribble labels to unlabeled regions by leveraging semantic affinity and shape priors, effectively improving pseudo-label quality and boundary accuracy. The CGTM adaptively updates high-confidence masks based on model predictions and feature correlations, enhancing the reliability of pseudo-labels. The CACL encourages intra-class compactness and inter-class separability by leveraging category-wise mean features, strengthening global feature discrimination. Experiments on ACDC and MSCMR datasets demonstrate superior performance over existing methods, highlighting the efficacy of integrating propagation strategies with contrastive learning. However, the current work is still limited to processing 2D image segmentation. Future work will extend this framework to 3D segmentation and explore its application in other medical imaging domains, enabling more efficient, scalable annotation frameworks for clinical use.

References

1. Hu, P., Liu, R., Zhang, S.: A cardiac mri segmentation method based on context cascade attention. In: 2024 6th International Conference on Communications, In formation System and Computer Engineering (CISCE). pp. 284289. IEEE (2024)
2. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classi cation with deep con volutional neural networks. Advances in neural information processing systems 25 (2012)
3. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
4. Webb, S., et al.: Deep learning for biology. Nature 554(7693), 555557 (2018)
5. Ammar, A., Bouattane, O., Youss , M.: Automatic cardiac cine mri segmentation and heart disease classi cation. Computerized Medical Imaging and Graphics 88, 101864 (2021)
6. Chen, Z., Sun, Q.: Weakly-supervised semantic segmentation with image-level labels: from traditional models to foundation models. arXiv preprint arX iv:2310.13026 (2023)
7. Ahn, J., Kwak, S.: Learning pixel-level semantic a nity with image-level super vision for weakly supervised semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 49814990 (2018)
8. Luo, X., Hu, M., Liao, W., Zhai, S., Song, T., Wang, G., Zhang, S.: Scribble supervised medical image segmentation via dual-branch network and dynamically mixed pseudo labels supervision. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 528538. Springer (2022)
9. Rajchl, M., Lee, M.C., Oktay, O., Kamnitsas, K., Passerat-Palmbach, J., Bai, W., Damodaram,M.,Rutherford, M.A., Hajnal, J.V., Kainz, B., et al.: Deepcut: Object segmentation from bounding box annotations using convolutional neural networks. IEEE transactions on medical imaging 36(2), 674683 (2016)
10. Wang, J., Xia, B.: Polar transformation based multiple instance learning assisting weakly supervised image segmentation with loose bounding box annotations. arXiv preprint arXiv:2203.06000 (2022)
11. Lin, D., Dai, J., Jia, J., He, K., Sun, J.: Scribblesup: Scribble-supervised convolu tional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 31593167 (2016)
12. Valvano, G., Leo, A., Tsaftaris, S.A.: Learning to segment from scribbles using multi-scale adversarial attention gates. IEEE Transactions on Medical Imaging 40(8), 1990 2001 (2021)
13. Tang, M., Djelouah, A., Perazzi, F., Boykov, Y., Schroers, C.: Normalized cut loss for weakly-supervised cnn segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 18181827 (2018)
14. Xu, Y., Quan, R., Xu, W., Huang, Y., Chen, X., Liu, F.: Advances in medical image segmentation: a comprehensive review of traditional, deep learning and hybrid approaches. Bioengineering 11(10), 1034 (2024)
15. Vernaza, P., Chandraker, M.: Learning random-walk label propagation for weakly supervised semantic segmentation. In: Proceedings of the IEEE conference on com puter vision and pattern recognition. pp. 71587166 (2017)
16. Kolesnikov, A., Lampert, C.H.: Seed, expand and constrain: Three principles for weakly-supervised image segmentation. In: Computer VisionECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 1114, 2016, Pro ceedings, Part IV 14. pp. 695711. Springer (2016)
17. Dai, J., He, K., Sun, J.: Boxsup: Exploiting bounding boxes to supervise convolu tional networks for semantic segmentation. In: Proceedings of the IEEE interna tional conference on computer vision. pp. 16351643 (2015)

18. Lee, J., Yi, J., Shin, C., Yoon, S.: Bbam: Bounding box attribution map for weakly supervised semantic and instance segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 26432652 (2021)
19. 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. 507522 (2018)
20. Lee, H., Jeong, W.K.: Scribble2label: Scribble-supervised cell segmentation via self-generating pseudo-labels with consistency. In: Medical Image Computing and Computer Assisted InterventionMICCAI 2020: 23rd International Conference, Li ma, Peru, October 48, 2020, Proceedings, Part I 23. pp. 1423. Springer (2020)
21. Zhang, Y., Ding, M., Bai, Y., Xu, M., Ghanem, B.: Beyond weakly supervised: Pseudo ground truths mining for missing bounding-boxes object detection. IEEE Transactions on Circuits and Systems for Video Technology 30(4), 983997 (2019)
22. Chaitanya, K., Erdil, E., Karani, N., Konukoglu, E.: Contrastive learning of glob al and local features for medical image segmentation with limited annotations. Advances in neural information processing systems 33, 1254612558 (2020)
23. Gidaris, S., Singh, P., Komodakis, N.: Unsupervised representation learning by predicting image rotations. arXiv preprint arXiv:1803.07728 (2018)
24. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A simple framework for con trastive learning of visual representations. In: International conference on machine learning. pp. 15971607. PmLR (2020)
25. He, K., Fan, H., Wu, Y., Xie, S., Girshick, R.: Momentum contrast for unsupervised visual representation learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 97299738 (2020)
26. Oord, A.v.d., Li, Y., Vinyals, O.: Representation learning with contrastive predic tive coding. arXiv preprint arXiv:1807.03748 (2018)
27. Bosnjak, M., Richemond, P.H., Tomasev, N., Strub, F., Walker, J.C., Hill, F., Buesing, L.H., Pascanu, R., Blundell, C., Mitrovic, J.: Semppl: Predicting pseudo labels for better contrastive representations. arXiv preprint arXiv:2301.05158 (2023)
28. Wang, Y., Wang, H., Shen, Y., Fei, J., Li, W., Jin, G., Wu, L., Zhao, R., Le, X.: Semi-supervised semantic segmentation using unreliable pseudo-labels. In: Pro ceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 42484257 (2022)
29. Yang, L., Qi, L., Feng, L., Zhang, W., Shi, Y.: Revisiting weak-to-strong consisten cy in semi-supervised semantic segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 72367246 (2023)
30. Wang, Y., Chen, H., Heng, Q., Hou, W., Fan, Y., Wu, Z., Wang, J., Savvides, M., Shinozaki, T., Raj, B., et al.: Freematch: Self-adaptive thresholding for semi supervised learning. arXiv preprint arXiv:2205.07246 (2022)
31. Bernard, O., Lalande, A., Zotti, C., Cervenansky, F., Yang, X., Heng, P.A., Cetin, I., Lekadir, K., Camara, O., Ballester, M.A.G., et al.: Deep learning techniques for automatic mri cardiac multi-structures segmentation and diagnosis: is the problem solved? IEEE transactions on medical imaging 37(11), 2514 2525 (2018)
32. Zhang, K., Zhuang, X.: Cyclemix: A holistic strategy for medical image segmen tation from scribble supervision. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1165611665 (2022)
33. 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. Pattern recognition 122, 108341 (2022)

34. Kim, B., Ye, J.C.: Mumford shah loss functional for image segmentation with deep learning. IEEE Transactions on Image Processing 29, 18561866 (2019)
35. Obukhov, A., Georgoulis, S., Dai, D., Van Gool, L.: Gated crf loss for weakly supervised semantic image segmentation. arXiv preprint arXiv:1906.04651 (2019)
36. Zhang, L., Li, W., Bi, K., Li, P., Zhang, L., Liu, H.: Fddseg: Unleashing the pow er of scribble annotation for cardiac mri images through feature decomposition distillation. IEEE Journal of Biomedical and Health Informatics (2024)
37. Grandvalet, Y., Bengio, Y.: Semi-supervised learning by entropy minimization. Advances in neural information processing systems 17 (2004Hu, P., Liu, R., Zhang, S. (2024, May). A Cardiac MRI Segmentation Method Based on Context Cascade Attention. In 2024 6th International Conference on Communications, Information System and Computer Engineering (CISCE) (pp. 284-289). IEEE.