EFFDNet: A Scribble-Supervised Medical Image Segmentation Method with Enhanced Foreground Feature Discrimination

Jinhua Liu[1,2], Shu Yun Tan[2,4], Xulei Yang[5], Yanwu Xu[6], and Si Yong Yeo[1,2,3*]

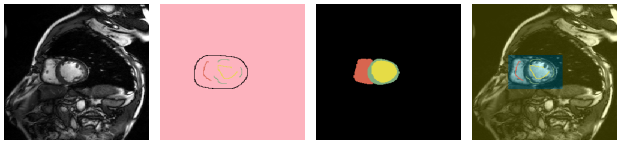
[1] MedVisAI Lab
[2] Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore
[3] Centre of Al in Medicine, Singapore
[4] National Healthcare Group Polyclinics, Singapore
[5] Institute for Infocomm Research (I2R), A*STAR, Singapore
[6] South China University of Technology, Guangdong, China

Abstract. Medical image segmentation, a critical task in medical image analysis, plays a key role in assisting clinical diagnostic workflows. However, traditional fully supervised learning methods for segmentation require large, high-quality annotations from expert physicians, which is resource-intensive and time-consuming. To mitigate this, scribble supervised segmentation approaches use simplified annotations to reduce annotation costs. Nevertheless, the simplistic nature of scribble annotations limits the model's ability to accurately distinguish foreground anatomical structures from the background and differentiate between various anatomical classes. This limitation results in low accuracy in capturing foreground morphology and hinders the model's generalization ability. To address this, we propose an Enhanced Foreground Feature Discrimination Network (EFFDNet) that better leverages semantic information in scribble annotations to improve the network's foreground discrimination ability. EFFDNet introduces an innovative Foreground-Background Separation Loss (FBSL), enhancing the model's ability to distinguish between foreground and background features, and improving the morphological accuracy of foreground anatomical region recognition. Additionally, we propose a new Foreground Augmentation with Diverse Context (FADC) strategy to further enhance the network's attention on the foreground and increase training sample diversity, mitigating overfitting and improving generalization. We validate our approach through systematic experiments on two publicly available datasets, demonstrating significant improvements over existing methods. The code is available at: https://github.com/Aurora-003-web/EFFDNet.

Keywords: Segmentation · Scribble supervised segmentation · Deep Learnng · Medical Imaging.

1 Introduction

Medical image segmentation [24, 26] is essential for clinical diagnosis and research, helping identify anatomical and pathological structures in medical im-



(a) (b) (c) (d)

Fig. 1. Illustration of scribble-supervised segmentation and its implied foreground-background semantics: (a) Image; (b) Scribble; (c) Segmentation; (d) Foreground-background semantics, where the blue area contains the foreground region and the yellow area contains the background region.

ages. The segmented outcomes can aid clinicians in decision-making and disease evaluation, advancing medical research. However, deep learning-based fully supervised segmentation typically requires large, accurately annotated datasets, often involving time-consuming and labor-intensive manual annotations by experts, especially with the rise of large-scale models [35]. Weakly Supervised Learning (WSL) addresses this issue by simplifying the annotation process with alternative labeling strategies such as scribbles [3, 17, 34, 16], bounding boxes [6, 21, 25, 32], point labels [33, 23, 4, 28, 31], and image-level labels [1, 12, 36, 9, 7, 8], thus reducing the complexity and labor of annotation. We specifically explore scribble supervision (as shown in Fig. 1 (b)), where rough sketches of the target organ or tissue is used to guide the network training, thus simplifying the annotation while still providing some anatomical details.

Recently, many methods for deep learning-based scribble-supervised segmentation have emerged. Among them, some methods treat scribble annotations as seed regions and propagate the information to the unannotated regions. For example, Lin et al. [17] proposed a graphical model optimization method to enhance segmentation by propagating scribble information using graphical model. Vernaza et al. [30] developed a Differentiable Random Walk-based [10] label propagation algorithm for better generalization. Can et al. [3] introduced an iterative training framework with Conditional Random Field (CRF) optimization to improve the performance. On the other hand, some methods treat scribble annotations as sparse annotations with partial pixel-level labels, leveraging them through methods such as partially cross-entropy (pCE) loss, and further developing various strategies to enhance segmentation performance. For example, Valvano et al. [29] presented Multi-Scale Adversarial Attention Gates, incorporating adversarial signals for precise localization. Zhang et al. [34] optimized segmentation with mix augmentation and cyclic consistency constraints. Luo et al. [20] introduced a dual-branch network and dynamic mixed pseudo-label supervision to address the issue of the model refusing to update. Li et al. [16] combined CNN and Transformer features and proposed an Attention-guided Class Activation Map (ACAM) branch to improve segmentation performance.

Although the aforementioned methods have been proposed, we argue that existing methods overlook the rich semantic information embedded in scrib-

bles, where experts often unknowingly incorporate prior knowledge of anatomical structures. Merely treating scribble annotations as initial seed regions or sparse pixel-level annotations, without exploiting the underlying rich information, often fails to provide sufficient details for networks to effectively separate foreground anatomical structures from the background, distinguish between different anatomical classes, or accurately identify the morphology of structures. Through observation, we find that scribbles inherently contain foreground-background semantics. Specifically, we find that regions with foreground-class scribbles tend to contain target anatomical structures, while regions without them tend to contain non-target areas, as shown in Fig. 1 (d). Exploiting this implicit foreground-background semantics can better guide network training and improve performance. In light of this observation, we introduce a novel loss function—**Foreground-Background Separation Loss (FBSL)**—which aims to maximize the separability of foreground and background features within the feature space, enhancing the network's ability to discriminate foreground anatomical structures. This further improves the morphological accuracy of the target areas and the clarity in distinguishing between different classes. Moreover, to further enhance the network's capability in recognizing foreground, we propose a novel **Foreground Augmentation with Diverse Context (FADC)** strategy. This strategy randomly substitutes foreground regions from different samples, increasing the sample diversity of foreground regions in various contexts (i.e., different backgrounds), which boosts the network's sensitivity to foreground features and improves the diversity of training samples. Additionally, by addressing overfitting—a common challenge in weakly supervised medical image segmentation (WSMIS)—FADC leads to a significant enhancement in performance.

In summary, by combining the two strategies, we innovatively propose a framework, named Enhanced Foreground Feature Discrimination Network (EFFD-Net). Specifically, the main contributions can be summarized as follows: (1) We introduce the FBSL to better utilize the foreground-background semantics contained in scribble annotations, thereby enhancing foreground discrimination in the feature space. (2) We propose the FADC mechanism to enhance the network's foreground sensitivity and model generalization by using a new foreground augmentation strategy. (3) We design a specialized network architecture optimized for medical image segmentation with scribble supervision, achieving near fully supervised performance while lowering annotation costs. Experiments on two medical datasets demonstrate its state-of-the-art performance.

2 Methodology

This paper proposes a WSMIS framework, which is illustrated in Fig. 2. First, we formalize the task for clarity in the following description. Specifically, we define a set consisting of $|\mathcal{D}|$ samples as $\mathcal{D} = \{(\mathcal{X}_i, \mathcal{Y}_i)\}_{i=1}^{|\mathcal{D}|}$. In this set, each sample consists of a pair of elements: the input $\mathcal{X}_i \in \mathbb{R}^{H \cdot W}$, which represents a two-dimensional (2D) slice image, and the corresponding annotation $\mathcal{Y}_i \in \{0,1\}^{H \cdot W \cdot C+1}$, which is a manually annotated scribble, covering C categories of

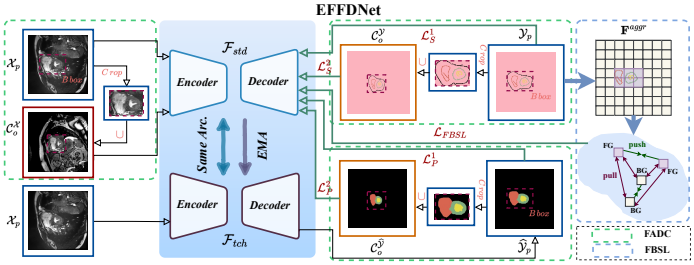


Fig. 2. The pipeline of the Enhanced Foreground Feature Discrimination Network (EFFDNet). Here, "FG" denotes the foreground, and "BG" denotes the background. The "push" brings tensors closer in feature space, while "pull" distances them. $\mathcal{L}_S^{1,2}$ and $\mathcal{L}_P^{1,2}$ denote the supervised and unsupervised losses for original and augmented data, respectively. $\mathcal{L}_{FBSL}$ is the Foreground-Background Separation Loss.

target objects to be segmented. Notably, the $(C+1)$-th class in the annotation is specifically used to identify pixel regions that are not annotated.

Basic Framework Our basic framework is inspired by the Mean Teacher [27], which is originally proposed for consistency regularization in semi-supervised learning. We reformulated it into a pseudo-label-based model to perform WSMIS.

It involves two networks: the student network $\mathcal{F}_{std}$ and the teacher network $\mathcal{F}_{tch}$. During training, the $\mathcal{F}_{std}$ is guided by scribble annotations $\mathcal{Y}_i$ and pseudo-labels $\widehat{\mathcal{Y}}_i$, which are generated by the $\mathcal{F}_{tch}$ as $\widehat{\mathcal{Y}}_i = \arg\max\left(\mathcal{F}_{tch}(\mathcal{X}_i;\Theta_{tch})\right)$. These pseudo-labels provide additional segmentation supervision to update the student network's weights Θ_{std} through gradient descent.

The $\mathcal{F}_{tch}$ is not updated via gradient descent. Instead, its weights Θ_{tch} are updated using an Exponential Moving Average (EMA) strategy. This helps stabilize the training process and improve the model's generalization. At each training step *step*, Θ_{tch} is smoothly updated based on $\Theta_{tch}^{step} = \alpha\Theta_{tch}^{step-1} + (1-\alpha)\Theta_{std}^{step}$. Among them, α is the EMA decay coefficient that controls the update rate.

In the basic framework, the loss function consists of two items. The first item is the scribble-supervised loss function $\mathcal{L}_S^1$, which can be expressed as follows:

$$\mathcal{L}_S^1 = -\frac{1}{N\cdot C\cdot \sum_{\Omega_{\mathcal{Y}_i^c}}}\sum_{i=1}^{N}\sum_{c=1}^{C}\sum_{(j,k)\in\Omega_{\mathcal{Y}_i^c}} \mathcal{Y}_i^c(j,k)\log\left(\mathcal{F}_{std}(\mathcal{X}_i;\Theta_{std})^c(j,k)\right). \quad (1)$$

Here, (j,k) represents the pixel index, while $\Omega_{\mathcal{Y}}$ denotes the set of pixels containing scribble annotations. N represents the sample size. The second item is the pseudo-label supervised loss, which can be formulated as follows:

$$\mathcal{L}_P^1 = -\frac{1}{N\cdot C\cdot H\cdot W}\sum_{i=1}^{N}\sum_{c=1}^{C}\sum_{j=0}^{H}\sum_{k=0}^{W}\widehat{\mathcal{Y}}_i^c(j,k)\log\left(\mathcal{F}_{std}(\mathcal{X}_i;\Theta_{std})^c(j,k)\right). \quad (2)$$

Foreground-Background Separation Loss To improve the network's foreground discrimination, we propose an innovative loss, $\mathcal{L}_{FBSL}$, as follows.

First, we extract the feature maps $\mathbf{F}$ from a mini-batch of data in the student network. These feature maps are obtained from the layer preceding the segmentation head. Next, we apply a local region aggregation operation to $\mathbf{F}$:

$$\mathbf{F}_i^{aggr}(r,\, s) \,=\, \frac{1}{\frac{H}{K}\cdot\frac{W}{K}} \sum_{m=0}^{\frac{H}{K}-1} \sum_{n=0}^{\frac{W}{K}-1} \mathbf{F}_i^{(r\cdot\frac{H}{K}+m,\, s\cdot\frac{W}{K}+n)}. \tag{3}$$

Here, $\frac{H}{K}\cdot\frac{W}{K}$ denotes the size of the local regions (i.e., $K\cdot K$ local regions are divided.), while (r, s) denotes the spatial index of the $\mathbf{F}_i^{aggr}$. Then, based on foreground-background semantics of the scribble annotations, we assigned corresponding region labels $\mathbf{R}$ to each local region of aggregation feature map $\mathbf{F}_i^{aggr}$:

$$\mathbf{R}_i(r,\, s) \,=\, \begin{cases} 1 & \text{if } \sum_{m=0}^{\frac{H}{K}-1} \sum_{n=0}^{\frac{W}{K}-1} \mathcal{Y}_i^{(r\cdot\frac{H}{K}+m,\, s\cdot\frac{W}{K}+n)} \mathbb{I}_C > 0, \\ 0 & \text{else.} \end{cases} \tag{4}$$

Here, the indicator function $\mathbb{I}_C(\cdot)$ selects only the scribbles belonging to the target $\{1, \ldots, C\}$ classes, excluding the unannotated $(C+1)$-th class. Inspired by contrastive losses such as InfoNCE loss [22] and SupConLoss [13], and leveraging the foreground-background semantics, we further design our $\mathcal{L}_{FBSL}$. The function aims to bring the foreground regions ($\mathbf{R}_i(r,\, s) = 1$) closer together, the background regions ($\mathbf{R}_i(r,\, s) = 0$) closer to each other, and simultaneously separate the foreground from the background, thereby enhancing the spatial distinction of features between them:

$$\mathcal{L}_{FBSL} = -\frac{\sum_{e=0}^{\frac{H}{K}N} \sum_{f=0}^{\frac{W}{K}N} \sum_{u=1}^{\mathcal{M}_{(e,f)}} \log \left[\frac{\exp \langle \mathbf{z}_{(e,f)}, \mathbf{z}^+_{(e,f,u)} \rangle / \tau}{\exp \langle \mathbf{z}_{(e,f)}, \mathbf{z}^+_{(e,f,u)} \rangle / \tau + \sum_{v=1}^{\mathcal{N}_{(e,f)}} \exp \langle \mathbf{z}_{(e,f)}, \mathbf{z}^-_{(e,f,v)} \rangle / \tau} \right]}{\frac{H}{K}N \cdot \frac{W}{K}N \cdot \mathcal{M}_{(e,f)}}. \tag{5}$$

Here, we define $\{\mathbf{z}\}$ as the set of feature tensors obtained by applying L2 normalization to the features at each location (r, s) in $\mathbf{F}^{aggr}$. This step enhances the training efficiency and stability. Each $\mathbf{z}_{(e,f)}$ serves as an anchor, with $\mathcal{M}_{(e,f)}$ being the number of positive samples and $\mathcal{N}_{(e,f)}$ the number of negative samples for each anchor. The set $\mathbf{z}^+_{(e,f)}$ denotes the positive, and $\mathbf{z}^-_{(e,f)}$ denotes the negative samples. The $\langle\cdot\rangle$ is the similarity measure function, where cosine similarity is used, and τ is the temperature coefficient. Positive and negative samples of the anchor are determined using the region labels $\mathbf{R}$: regions with the same label as the anchor are positive, and those with a different label are negative.

Finally, during the training, we introduce the $\mathcal{L}_{FBSL}$ on top of the base network: $\mathcal{L}^1 \,=\, \mathcal{L}^1_S + \lambda(\mathcal{L}^1_P + \delta\mathcal{L}_{FBSL})$. Here, λ and δ are the weighting coefficients.

Foreground Augmentation with Diverse Context Furthermore, we propose FADC to augment the dataset. For a given randomly selected sample $\mathcal{X}_p$, we select a sample $\mathcal{X}_o$ from the current batch and apply the following process:

$$\mathcal{C}_o^{\mathcal{X}} \,=\, \mathcal{X}_o \setminus \mathrm{B\,box}(\mathcal{X}_o) \cup \mathrm{C\,rop}(\mathcal{X}_p,\, \mathrm{B\,box}(\mathcal{X}_p)); \tag{6}$$

$$\mathcal{C}_o^{\mathcal{Y}} \,=\, \mathcal{Y}_o \setminus \mathrm{B\,box}(\mathcal{Y}_o) \cup \mathrm{C\,rop}(\mathcal{Y}_p,\, \mathrm{B\,box}(\mathcal{Y}_p)); \tag{7}$$

$$\mathcal{C}_o^{\widehat{\mathcal{Y}}} \,=\, \widehat{\mathcal{Y}}_o \setminus \mathrm{B\,box}(\widehat{\mathcal{Y}}_o) \cup \mathrm{C\,rop}(\widehat{\mathcal{Y}}_p,\, \mathrm{B\,box}(\widehat{\mathcal{Y}}_p)). \tag{8}$$

Here, the function $\mathrm{B\,box}(\cdot)$ locates the foreground region based on the scribble bounding box (i.e., the minimal bounding box enclosing the scribble annotated region.), and $\mathrm{C\,rop}(\cdot)$ is responsible for cropping out this region. The operation $\setminus$ removes the foreground to retain the background, while $\cup$ merges the cropped foreground with the background. After these steps, we obtain samples with foregrounds having diverse background contexts, denoted as $\{\mathcal{C}^{\mathcal{X}}\}$, along with the corresponding scribbles $\{\mathcal{C}^{\mathcal{Y}}\}$ and pseudo-labels $\{\mathcal{C}^{\widehat{\mathcal{Y}}}\}$. These processed data are then used to train the neural network by modifying $\mathcal{L}_S^1$ and $\mathcal{L}_P^1$, using $\{\mathcal{C}^{\mathcal{X}}\}$ as input and $\{\mathcal{C}^{\mathcal{Y}}\}$, $\{\mathcal{C}^{\widehat{\mathcal{Y}}}\}$ as supervision. This results in the new loss functions $\mathcal{L}_S^2$ and $\mathcal{L}_P^2$ for training. Then, the overall loss for learning from the samples obtained through FADC can be expressed as $\mathcal{L}^2$: $\mathcal{L}^2 = \mathcal{L}_S^2 + \lambda\mathcal{L}_P^2$. Finally, the total loss $\mathcal{L}$ of our EFFDNet is as shown below: $\mathcal{L} = \mathcal{L}^1 + \mathcal{L}^2$.

3 Experiments and Results

Datasets and Evaluation Metrics We used two publicly available benchmark datasets: the ACDC [2] and the NCI-ISBI [5]. The ACDC consists of 200 cine-MRI scans from 100 patients, with three segmentation classes (left ventricle (LV), right ventricle (RV), and myocardium (Myo)). The NCI-ISBI contains 80 MRI scans with two classes (central gland (CG) and peripheral zone (PZ)). Details on scribble-based annotations can be found in [19]. Due to the low cross-slice resolution of these datasets, we used a slice-wise 2D training strategy [29], resizing each slice to 256×256 pixels. In testing, all slices were reassembled to reconstruct 3D images for evaluation (i.e., the original 3D data is used to evaluate accuracy.). A five-fold cross-validation approach was applied to assess segmentation accuracy. The Dice Similarity Coefficient (DSC) (%) was used to measure the effectiveness of our method.

Implementation Details Our method is implemented using PyTorch and runs on NVIDIA GeForce RTX 4090 GPU. We employ U-Net [24, 19] as the backbone, which can be replaced with other advanced models by simply modifying the networks. To mitigate overfitting, we apply random rotations and flips for augmentation. Training is conducted using the SGD optimizer with an initial learning rate of $1e$-2 (adjusted via polynomial scheduling), a momentum of 0.9, and a weight decay of $1e$-4. The batch size is set to 12, and training runs for up to $60,000$ iterations to ensure model convergence. α is set to 0.99, K is set to 8, and λ and δ are set to 0.6 and 0.3, respectively.

Comparison with Other Methods To validate the effectiveness of our proposed method, we comprehensively compare it with multiple advanced WSL approaches. These comparative methods were obtained through our own execution. Quantitative results (mean and standard deviation) on the ACDC and NCI-ISBI datasets are summarized in Table 1, with the best and second-best results highlighted in red and blue, respectively. Our method achieves state-of-the-art DSC scores across all categories, with statistically significant improvements (* indicates that our method significantly outperforms other WSL methods ($p < 0.05$)). Specifically, on the ACDC dataset, it outperforms Scribformer by 0.74%, 1.55%,

Table 1. Comparison of methods on the ACDC and NCI-ISBI datasets. The reported values represent DSC metric. * indicates that our method significantly outperforms other WSL methods ($p < 0.05$). RV, Myo, and LV are three classes from the ACDC dataset, while PZ and CG are two classes from the NCI-ISBI dataset.

Type	Method	ACDC			NCI-ISBI	
		RV	Myo	LV	PZ	CG
WSL	pCE [17]	56.44(11.51)*	56.63(3.60)*	69.00(10.12)*	22.03(6.69)*	45.33(4.61)*
	RW [10]	81.54(4.29)*	71.02(4.08)*	84.75(3.37)*	72.72(5.12)	78.94(3.58)*
	USTM [18]	79.27(4.18)*	74.07(3.40)*	76.60(7.85)*	65.57(3.62)*	36.20(9.82)*
	S2L [15]	83.68(2.54)*	81.87(2.83)*	87.44(6.67)	69.79(4.57)*	55.66(4.82)*
	MLoss [14]	83.37(2.56)*	82.56(2.55)*	90.68(4.01)	70.87(4.17)*	81.39(1.58)*
	EntMin [11]	83.21(3.03)*	80.99(2.82)*	88.73(4.57)*	59.19(4.48)*	42.74(5.45)*
	DMPLS [20]	86.22(2.71)*	83.82(2.38)*	91.46(3.27)	69.14(4.18)*	56.43(7.94)*
	Scribformer [16]	86.24(3.11)	84.01(2.13)*	91.07(3.63)	69.37(2.49)*	74.20(3.54)*
	Ours	86.98(2.55)	85.56(2.59)	92.48(2.43)	72.77(4.18)	86.67(0.62)
FSL	FullSup [24]	89.49(1.90)	89.07(1.88)	93.95(2.76)	77.23(3.96)	87.90(0.73)

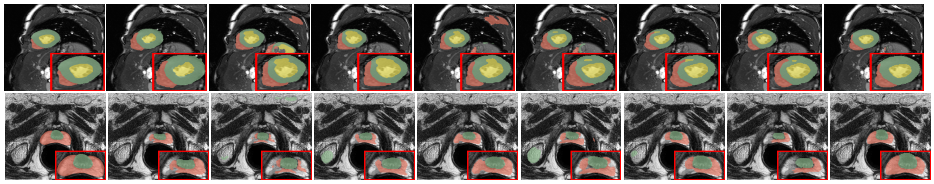


Fig. 3. Comparison visualization on ACDC (first row) and NCI-ISBI (second row).

and 1.41% in the RV, Myo, and LV categories, respectively. The advantage is more pronounced on the NCI-ISBI dataset, with improvements of 3.40% and 12.47% in the PZ and CG categories, respectively. Notably, our method, despite relying on weak scribble annotations, achieves performance comparable to fully supervised methods (FullSup), demonstrating its ability to reduce annotation costs while maintaining high segmentation accuracy. Fig. 3 show that our method better captures anatomical structures in foreground regions. Specifically, the first row exhibits reduced inter-class misclassification for RV (orange), Myo (green), and LV (yellow), along with finer boundary adherence. Meanwhile, the second row demonstrates more complete segmentation and more precise delineation for PZ (orange) and CG (green). And the results also show that our method mitigates background over-segmentation. The above analyses confirm our model's superior ability to enhance foreground discrimination, thereby improving segmentation accuracy metrics and morphological accuracy.

Ablation Study Table 2 shows the results of ablation study, demonstrating that both proposed strategies significantly enhance performance. On the ACDC, our $\mathcal{L}_{FBSL}$ improves DSC by 2.77%, 4.44%, and 4.46% for RV, Myo, and LV, respectively. On the NCI-ISBI, the gains are even higher, reaching 6.48% and 19.98% for PZ and CG. The FADC also contributes notable improvements of

Table 2. The ablation experiments conducted on $\mathcal{L}_{FBSL}$ and FADC

Method	ACDC			NCI-ISBI	
	RV	Myo	LV	PZ	CG
Baseline	83.13(3.92)	80.16(2.37)	87.71(4.39)	65.91(9.17)	65.92(15.11)
w/. $\mathcal{L}_{FBSL}$	85.90(2.84)	84.60(1.61)	92.17(3.40)	72.39(4.24)	85.90(0.28)
w/. FADC	85.88(3.12)	82.55(2.97)	89.49(3.48)	71.15(7.28)	82.18(1.66)
Ours	86.54(2.47)	85.67(2.71)	92.15(2.70)	73.00(3.95)	86.37(0.62)

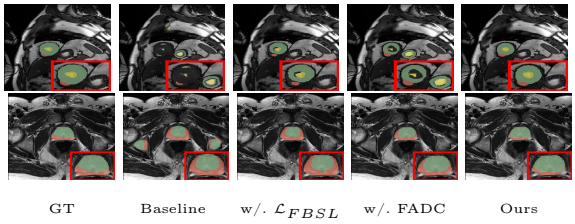


Fig. 4. Ablation visualization on ACDC (first row) and NCI-ISBI (second row).

2.75%, 2.39%, 1.78%, 5.24%, and 16.26% across these categories. Combining both strategies achieves the best overall performance, with average improvements of 4.45% and 13.77%. These results highlight the contributions of our method, and Fig. 4 qualitatively illustrates the visualization improvements. First, the $\mathcal{L}_{FBSL}$ loss enhances foreground-background separation, reducing background over-segmentation (e.g., incorrect extra segmentations of Myo and LV) and improving the accuracy of segmentation morphology. Then, the FADC further strengthens foreground discrimination and generalization by introducing images of diverse augmented foreground, reducing inter-class confusion. For example, in the first row, FADC improves the morphological accuracy of RV, Myo, and LV, while in the second row, it reduces misclassification between PZ and CG. Together, these strategies collectively prevent background over-segmentation, reduce inter-class confusion, and maintain morphological accuracy.

In addition, we analyzed the sensitivity of the hyperparameters δ and K in the $\mathcal{L}_{FBSL}$, as shown in Fig. 5. Experimental results indicate that the model's performance is relatively insensitive to variations in δ and K. Consequently, we select $\delta = 0.3$ and $K = 8$ as the optimal value for the best average accuracy.

4 Conclusion

In this paper, we propose EFFDNet, which enhances foreground discrimination ability. By analyzing the foreground-background semantics in scribbles, we introduce a novel loss function, FBSL, to improve the network's ability to distinguish between foreground and background regions in feature space, addressing misegmentation while preserving morphology accuracy of segmentations. Additionally, we design a new mechanism, FADC, which enhances the network's sensitivity of

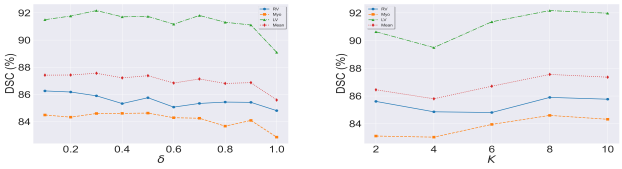


Fig. 5. Sensitivity analysis of hyperparameters δ and K.

foreground regions and mitigates overfitting. Experiments on two medical image datasets demonstrate the effectiveness of our EFFDNet and proposed strategies, resulting in notable improvements.

Although our network reduces misclassification across categories by enhancing foreground discrimination, it does not explicitly impose constraints on misclassification within the foreground region. In future work, we will focus on improving the network's ability to address this issue by designing more specific loss functions or adopting pseudo-label correction strategies.

Acknowledgments. This project is supported by the Ministry of Education, Singapore, under its Academic Research Fund Tier 1 RG25/24 and Academic Research Fund Tier 1 RS16/23. This project is also supported by Lee Kong Chian School of Medicine - Ministry of Education Start-Up Grant.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Ahn, J., Kwak, S.: Learning pixel-level semantic affinity with image-level supervision for weakly supervised semantic segmentation. In: CVPR 2018. pp. 4981–4990 (2018)
2. Bernard, O., Lalande, A., Zotti, C., Cervenansky, F., Yang, X., Heng, P., Cetin, I., Lekadir, K., Camara, O., Ballester, M.Á.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)
3. 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: MICCAI 2018 Workshop. pp. 236–244 (2018)
4. Chamanzar, A., Nie, Y.: Weakly supervised multi-task learning for cell detection and segmentation. In: ISBI 2020. pp. 513–516 (2020)
5. Clark, K.W., Vendt, B.A., Smith, K.E., Freymann, J.B., Kirby, J.S., Koppel, P., Moore, S.M., Phillips, S.R., Maffitt, D.R., Pringle, M., Tarbox, L., Prior, F.W.: The cancer imaging archive (TCIA): maintaining and operating a public information repository. Journal of digital imaging **26**(6), 1045–1057 (2013)
6. Dai, J., He, K., Sun, J.: Boxsup: Exploiting bounding boxes to supervise convolutional networks for semantic segmentation. In: ICCV 2015. pp. 1635–1643 (2015)
7. Fan, J., Lv, T., Di, Y., Li, L., Pan, X.: Pathmamba: Weakly supervised state space model for multi-class segmentation of pathology images. In: MICCAI 2024. vol. 15008, pp. 500–509 (2024)

8. Feng, S., Chen, J., Liu, Z., Liu, W., Wang, Z., Lan, R., Pan, X.: Mining gold from the sand: Weakly supervised histological tissue segmentation with activation relocalization and mutual learning. In: MICCAI 2024. vol. 15008, pp. 414–423 (2024)
9. Fu, J., Wang, G., Lu, T., Yue, Q., Vercauteren, T., Ourselin, S., Zhang, S.: UM-CAM: uncertainty-weighted multi-resolution class activation maps for weakly-supervised segmentation. Pattern Recognition **160**, 111204 (2025)
10. Grady, L.J.: Random walks for image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence **28**(11), 1768–1783 (2006)
11. Grandvalet, Y., Bengio, Y.: Semi-supervised learning by entropy minimization. In: NeurIPS 2004. pp. 529–536 (2004)
12. Huang, Z., Wang, X., Wang, J., Liu, W., Wang, J.: Weakly-supervised semantic segmentation network with deep seeded region growing. In: CVPR 2018. pp. 7014–7023 (2018)
13. Khosla, P., Teterwak, P., Wang, C., Sarna, A., Tian, Y., Isola, P., Maschinot, A., Liu, C., Krishnan, D.: Supervised contrastive learning. Advances in neural information processing systems **33**, 18661–18673 (2020)
14. Kim, B., Ye, J.C.: Mumford-shah loss functional for image segmentation with deep learning. IEEE Transactions on Image Processing **29**, 1856–1866 (2020)
15. Lee, H., Jeong, W.: Scribble2label: Scribble-supervised cell segmentation via self-generating pseudo-labels with consistency. In: MICCAI 2020. pp. 14–23 (2020)
16. Li, Z., Zheng, Y., Shan, D., Yang, S., Li, Q., Wang, B., Zhang, Y., Hong, Q., Shen, D.: Scribformer: Transformer makes CNN work better for scribble-based medical image segmentation. IEEE Transactions on Medical Imaging **43**(6), 2254–2265 (2024)
17. Lin, D., Dai, J., Jia, J., He, K., Sun, J.: Scribblesup: Scribble-supervised convolutional networks for semantic segmentation. In: CVPR 2016. pp. 3159–3167 (2016)
18. 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)
19. Luo, X.: WSL4MIS. https://github.com/Luoxd1996/WSL4MIS (2021)
20. 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: MICCAI 2022. pp. 528–538 (2022)
21. Oh, Y., Kim, B., Ham, B.: Background-aware pooling and noise-aware loss for weakly-supervised semantic segmentation. In: CVPR 2021. pp. 6913–6922 (2021)
22. van den Oord, A., Li, Y., Vinyals, O.: Representation learning with contrastive predictive coding. CoRR **abs/1807.03748** (2018)
23. Qu, H., Wu, P., Huang, Q., Yi, J., Yan, Z., Li, K., Riedlinger, G.M., De, S., Zhang, S., Metaxas, D.N.: Weakly supervised deep nuclei segmentation using partial points annotation in histopathology images. IEEE Transactions on Image Processing **39**(11), 3655–3666 (2020)
24. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: MICCAI 2015. pp. 234–241 (2015)
25. Song, C., Huang, Y., Ouyang, W., Wang, L.: Box-driven class-wise region masking and filling rate guided loss for weakly supervised semantic segmentation. In: CVPR 2019. pp. 3136–3145 (2019)
26. Sun, C., Guo, S., Zhang, H., Li, J., Chen, M., Ma, S., Jin, L., Liu, X., Li, X., Qian, X.: Automatic segmentation of liver tumors from multiphase contrast-enhanced CT images based on fcns. Artificial Intelligence in Medicine **83**, 58–66 (2017)

27. Tarvainen, A., Valpola, H.: Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In: NeurIPS 2017. pp. 1195–1204 (2017)
28. Tian, K., Zhang, J., Shen, H., Yan, K., Dong, P., Yao, J., Che, S., Luo, P., Han, X.: Weakly-supervised nucleus segmentation based on point annotations: A coarse-to-fine self-stimulated learning strategy. In: MICCAI 2020. pp. 299–308 (2020)
29. 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)
30. Vernaza, P., Chandraker, M.: Learning random-walk label propagation for weakly-supervised semantic segmentation. In: CVPR 2017. pp. 2953–2961 (2017)
31. Wang, Z., Zhang, Y., Wang, Y., Cai, L., Zhang, Y.: Dynamic pseudo label optimization in point-supervised nuclei segmentation. In: MICCAI 2024. vol. 15008, pp. 220–230 (2024)
32. Wei, J., Hu, Y., Cui, S., Zhou, S.K., Li, Z.: Weakpolyp: You only look bounding box for polyp segmentation. In: MICCAI 2023. vol. 14222, pp. 757–766 (2023)
33. Yoo, I., Yoo, D., Paeng, K.: Pseudoedgenet: Nuclei segmentation only with point annotations. In: MICCAI 2019. pp. 731–739 (2019)
34. Zhang, K., Zhuang, X.: Cyclemix: A holistic strategy for medical image segmentation from scribble supervision. In: CVPR 2022. pp. 11646–11655 (2022)
35. Zhang, Z., Yu, Y., Chen, Y., Yang, X., Yeo, S.Y.: Medunifier: Unifying vision-and-language pre-training on medical data with vision generation task using discrete visual representations. In: CVPR 2025. pp. 29744–29755 (2025)
36. Zhou, Y., Wu, Y., Wang, Z., Wei, B., Lai, M., Shou, J., Fan, Y., Xu, Y.: Cyclic learning: Bridging image-level labels and nuclei instance segmentation. IEEE Transactions on Medical Imaging **42**(10), 3104–3116 (2023)