SDRF-Net: Synergistic Dual-Resolution Fusion for Semi-Supervised Medical Image Segmentation

Xu Tang[1], Shiwei Zhou[2], Dengdi Sun[2*], Haifeng Zhao[1*]

[1]School of Computer Science and Technology, Anhui University, Hefei 230601, China

[2]School of Artificial Intelligence, Anhui University, Hefei 230601, China

***Abstract*—Semi-supervised learning (SSL) improves medical image segmentation by exploiting limited labeled and abundant unlabeled data. However, the imbalance between large background and sparse foreground regions hinders the learning of rare classes, while inevitable errors in pseudo-labels cause class confusion and reduce segmentation accuracy. To address these issues, we propose a novel Synergistic Dual-Resolution Fusion Network (SDRF-Net). Our framework introduces a grid-based foreground reorganization (GFR) module to reconstruct and diversify foreground regions, improving rare class recognition. A dual-segmentation head architecture (DSHA) then synergistically integrates global contextual information and local fine-grained details. Furthermore, an entropy-aware weighting (EAW) mechanism is employed to reduce noise during pseudo-label fusion by suppressing unreliable predictions. Extensive experiments on three public medical image datasets show that SDRF-Net effectively alleviates class imbalance and pseudo-label inaccuracy, achieving state-of-the-art performance.**

***Index Terms*—Semi-supervised learning, Medical image segmentation, Synergistic Dual-Resolution Fusion**

I. Introduction

Semantic segmentation assigns a category label to every pixel in an image, and while deep neural networks have significantly advanced medical image segmentation [1]–[4], their reliance on large-scale annotated datasets remains a major bottleneck due to high annotation costs [5]–[7]. Semi-supervised learning (SSL) addresses this by leveraging limited labeled data alongside abundant unlabeled data [8]–[10], reducing annotation dependency and improving generalization [11]–[14]. Therefore, it has become an important idea for medical image segmentation.

Pseudo-labeling is one of the widely adopted strategies in semi-supervised medical image segmentation (SSMIS), and the quality of the generated pseudo-labels is crucial for model learning. Since pseudo-labels inevitably contain errors, effectively mitigating these errors remains a central challenge in this field. Existing methods, such as BCP [15], which aligns distributions via region-based copy-paste, MCF [16], which employs a dual-network framework with dynamic pseudo-label generation, and multi-decoder or auxiliary-decoder approaches [17]–[19], have alleviated this problem to some extent. However, these methods generally overlook the impact of foreground–background distribution on pseudo-label quality. In particular, models often underperform on rare foreground classes, resulting in a higher rate of pseudo-label errors.

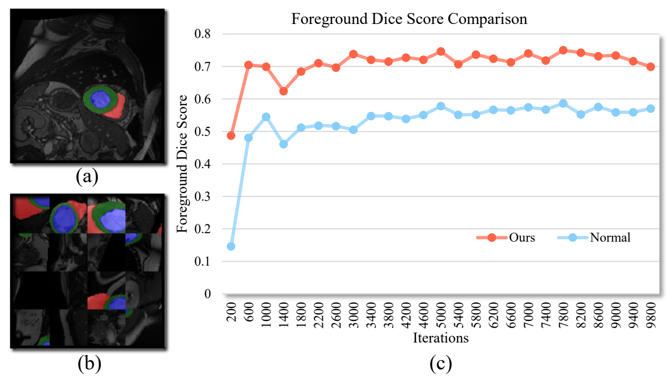

Fig. 1: The Dice Score progression for the foreground class during U-Net pre-training on ACDC. The orange and blue lines denote performance using the reconstructed and original datasets, respectively.

To address the aforementioned issues, as illustrated in Fig. 1(c), we enhance the model's multidimensional representation capabilities for rare foreground categories by reconstructing diverse foreground-background combinations. Simultaneously, to facilitate more effective learning of foreground features in both original and reconstructed data, we further introduce a Dual Segmentation Head Architecture (DSHA). Specifically, the Grid-based Foreground Reorganization (GFR) module divides the input image into N×N non-overlapping grids and systematically redistributes foreground regions across different grid cells. Subsequently, the reconstructed image is restored to its original size (Fig. 1 (b)), enhancing the learnability of foreground features through controlled spatial redistribution. The Dual Segmentation Head Architecture (DSHA) achieves feature complementarity through a primary segmentation head and an auxiliary segmentation head. The primary segmentation head employs large, deep convolutional kernels to extract global semantic features from raw data, while the auxiliary segmentation head uses small convolutional kernels to capture local, fine-grained features

* Corresponding authors: Dengdi Sun (sundengdi@163.com), Haifeng Zhao (senith@163.com)

This work was supported by the National Natural Science Foundation of China (62576006, 62472004, 62076005), the Anhui Provincial University Natural Science Research Major Projects (2025AHGXZK20027), the Key Natural Science Project of Anhui Provincial Education Department (2025AHGXZK30717), the Provincial Quality Project of Education in the New Era 2023 (Postgraduate Education 2023lhpysfjd009), and the High-Performance Computing Platform of Anhui University.

from the restructured data. Extracted features are dynamically integrated based on the entropy-aware weighting strategy (EAW), adaptively balancing the contributions of global and local information. The main contributions can be summarized as follows:

- We propose a Synergistic Dual-Resolution Fusion Network that reorganizes diverse foreground classes and enhances pseudo-label quality.
- A Grid-based Foreground Reorganization strategy is introduced to construct diverse foreground distributions.
- We design a Dual-Segmentation with Entropy-Aware Weighting to learn two perspectives, adaptively fusing predictions for more reliable pseudo-labels.

II. Related work

A. Data Augmentation Strategies

Data augmentation plays a crucial role in semi-supervised medical image segmentation by enhancing training diversity and improving model generalization. Conventional techniques, such as geometric transformations and intensity perturbations, provide limited appearance variations. To further enrich the training distribution, recent methods explore region-level and sample-mixing strategies. Mixup [20] and CutMix [21] combine image regions or pixel distributions to increase sample diversity, while subsequent works extend these ideas to semi-supervised settings by leveraging labeled data to guide unlabeled learning, such as GuidedMix-Net [22], BCP [15], and ABD [23]. However, most existing augmentations perturb only the input space without explicitly considering foreground structures. In contrast, we introduce a foreground-aware data augmentation strategy combined with a dual-segmentation head architecture, which explicitly separates the learning of original data and reorganized foreground regions, thereby improving generalization and robustness, especially for rare foreground classes.

B. Network Architectures

Encoder–decoder architectures such as U-Net [3] and V-Net [2] form the backbone of many semi-supervised medical image segmentation methods. To further reduce prediction errors, recent studies adopt dual or multiple networks for mutual supervision and consistency regularization, including CauSSL [24], ECS [25], and DWCL [19]. More recent approaches incorporate hybrid architectures or complex training schemes, such as M-CnT [26], MCF [16] and AD-MT [27], which introduce increased architectural and training complexity. In contrast, our dual-segmentation head architecture captures complementary global and foreground-specific information, which helps refine pseudo-label quality and improve robustness under class imbalance.

III. Prelimary

In SSMIS, we are given a small labeled dataset $\mathcal{D}^\ell = \{(\boldsymbol{x}_i^\ell, \boldsymbol{y}_i^\ell)\}_{i=1}^N$ and a large unlabeled dataset $\mathcal{D}^u = \{\boldsymbol{x}_i^u\}_{i=N+1}^{M+N}$, where $N \ll M$, with $\boldsymbol{x}_i^\ell$, $\boldsymbol{y}_i^\ell$, and $\boldsymbol{x}_i^u$ denoting labeled images, their ground truths, and unlabeled data, respectively. Our goal is to learn a robust segmentation model $f(\cdot)$ that assigns pixel-wise labels from K classes. The segmentator $\mathcal{F}(\theta, \cdot) = E(\cdot) \circ D \circ (\mathcal{H}^o + \mathcal{H}^p)$ consists of an encoder $E(\cdot)$, a decoder D, and segmentation heads $\mathcal{H}(\cdot)$.

IV. Proposed Method

Our approach is fundamentally centered around the Mean Teacher model, a semi-supervised learning framework where the student model is guided by a more stable teacher model. The overall framework is shown in Fig. 2.

A. Dual-Segmentation with Entropy-Aware Weighting

In this paper, the teacher and student models are respectively defined as: $\mathcal{F}_t(\theta_t, \boldsymbol{x}) = E_t(\boldsymbol{x}) \circ D_t \circ (\mathcal{H}_t^o + \mathcal{H}_t^p)$ and $\mathcal{F}_s(\theta_s, \boldsymbol{x}) = E_s(\boldsymbol{x}) \circ D_s \circ (\mathcal{H}_s^o + \mathcal{H}_s^p)$. Both of them use a U-Net backbone (e.g., V-Net for 3D data). To effectively learn foreground class features from the original and reorganized data, we design a Dual-Segmentation Head Architecture (DSHA) in both the teacher and student models above, namely $\mathcal{H}^o$ and $\mathcal{H}^p$ in the formula. Here, the primary segmentation head $\mathcal{H}^o$ extracts global semantic features, while the auxiliary segmentation head $\mathcal{H}^p$ focuses on capturing fine-grained local features.

To further balance the output of the dual segmentation head. We propose an Entropy-Aware Weighting (EAW) strategy to dynamically fuse predictions from the dual segmentation heads based on their uncertainty. Let $\hat{\mathrm{Y}}_1$ and $\hat{\mathrm{Y}}_2$ represent the outputs from the primary and auxiliary heads, respectively. Their probability distributions are obtained via softmax normalization:

$$\mathrm{P}_{\hat{\mathrm{Y}}_1} = \mathrm{softmax}(\hat{\mathrm{Y}}_1), \quad \mathrm{P}_{\hat{\mathrm{Y}}_2} = \mathrm{softmax}(\hat{\mathrm{Y}}_2). \tag{1}$$

The uncertainty of each prediction is measured using entropy:

$$\begin{aligned} \mathrm{E}_{\hat{\mathrm{Y}}_1} &= -\sum\nolimits_i \mathrm{P}_{\hat{\mathrm{Y}}_1} \log(\mathrm{P}_{\hat{\mathrm{Y}}_1} + \epsilon), \\ \mathrm{E}_{\hat{\mathrm{Y}}_2} &= -\sum\nolimits_i \mathrm{P}_{\hat{\mathrm{Y}}_2} \log(\mathrm{P}_{\hat{\mathrm{Y}}_2} + \epsilon), \end{aligned} \tag{2}$$

where ϵ is a small constant for numerical stability. The final fusion weights are computed as:

$$\alpha_G = \frac{1 - \frac{\mathrm{E}_{\hat{\mathrm{Y}}1}}{\mathrm{E}_{\hat{\mathrm{Y}}1} + \mathrm{E}_{\hat{\mathrm{Y}}2} + \epsilon}}{\sum_{k=1}^{2} \left(1 - \frac{\mathrm{E}_{\hat{\mathrm{Y}}k}}{\mathrm{E}_{\hat{\mathrm{Y}}_1} + \mathrm{E}_{\hat{\mathrm{Y}}_2} + \epsilon}\right)}, \quad \alpha_L = 1 - \alpha_G. \tag{3}$$

The fused output is then obtained by:

$$\mathrm{Y} = \alpha_G \cdot \hat{\mathrm{Y}}_1 + \alpha_L \cdot \hat{\mathrm{Y}}_2. \tag{4}$$

B. Grid-based Foreground Reorganization

To mitigate severe class imbalance in medical image segmentation, particularly the sparsity of foreground regions, we propose a Grid-based Foreground Reorganization (GFR) strategy. GFR systematically reorganizes foreground regions within an N×N non-overlapping grid structure to construct diverse foreground distributions and improve recognition of rare foreground features. The approach consists of two key components: Grid Partitioning and Reorganization, and Foreground-Background Fusion.

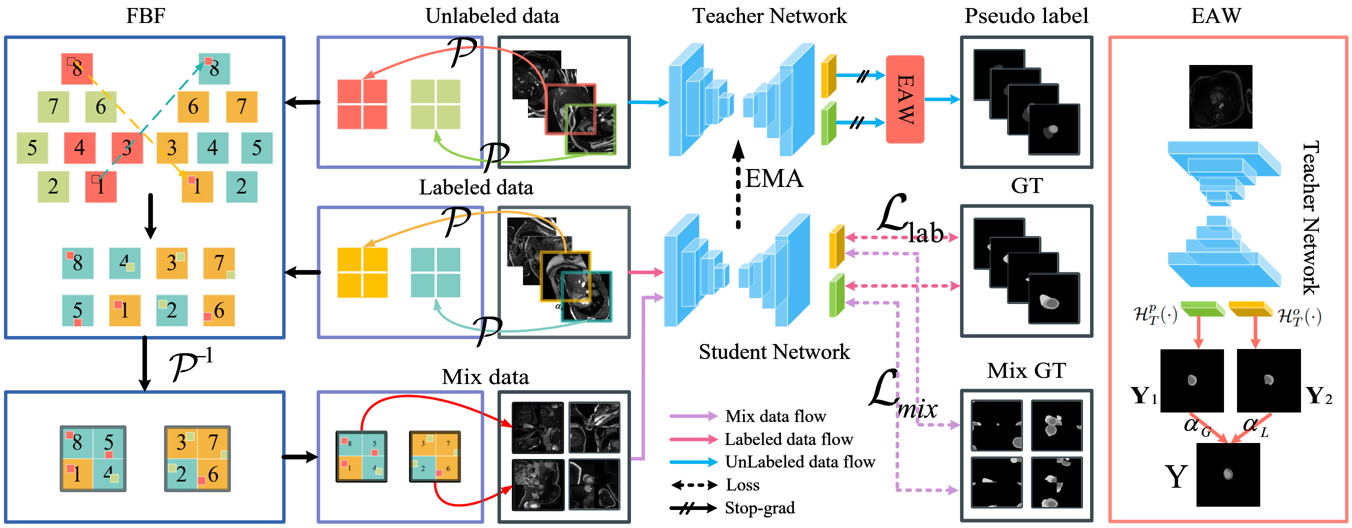

Fig. 2: The overall framework uses a 2×2 partitioning strategy on the ACDC dataset. Each sample is represented by a distinct colored block, with numerical labels indicating foreground proportion ranking and black rectangles showing extracted foreground features. Label processing is omitted for clarity.

TABLE I: Comparisons with SOTA on the ACDC dataset.

Method	Labeled	Unlabeled	Dice ↑	Jaccard ↑	95HD ↓	ASD ↓
U-Net	5%	0	47.83	37.01	31.16	12.62
U-Net	10%	0	79.41	68.11	9.35	2.70
U-Net	All	0	91.44	84.59	4.30	0.99
UA-MT	5%	95%	46.04	35.97	20.08	7.75
SASSNet			57.77	46.14	20.05	6.06
DTC			56.90	45.67	23.36	7.39
URPC			55.87	44.64	13.60	3.74
MC-Net			62.85	52.29	7.62	2.33
SS-Net			65.83	55.38	6.67	2.28
BCP			87.59	78.67	**1.90**	**0.67**
Ours			**87.95**	**79.19**	6.67	1.86
UA-MT	10%	90%	81.65	70.64	6.88	2.02
SASSNet			84.50	74.34	5.42	1.86
DTC			84.29	73.92	12.81	4.01
URPC			83.10	72.41	4.84	1.53
MC-Net			86.44	77.04	5.50	1.84
SS-Net			86.78	77.67	6.07	1.40
BCP			88.84	80.62	3.98	1.17
Ours			**90.46**	**83.06**	**1.58**	**0.57**

1) Grid Partitioning and Reorganization: Formally, the Grid Partitioning operation $\mathcal{P} : \mathbb{R}^{N\times H\times W\times C} \rightarrow \mathbb{R}^{N\times N\times \frac{H}{N}\times\frac{W}{N}\times C}$ divides the input image into $N \times N$ non-overlapping grid cells, where H, W, and C denote height, width, and number of channels, respectively. Its inverse operation $\mathcal{P}^{-1}$ reconstructs the image to its original dimensions. During self-training, we sample a batch of labeled data $\mathcal{D}^{\ell}_{\text{batch}} = \{(\boldsymbol{x}^{\ell}_i, \boldsymbol{y}^{\ell}_i)\}^{B}_{i=1}$ and unlabeled data $\mathcal{D}^{u}_{\text{batch}} = \{(\boldsymbol{x}^{u}_i, \hat{\boldsymbol{y}}^{u}_i)\}^{B}_{i=1}$, where B is the batch size and $\hat{\boldsymbol{y}}^u_i$ denotes the pseudo-label from the teacher model. Both batches undergo grid partitioning:

$$\mathcal{D}^{\ell}_{\text{grid}}, \mathcal{D}^{u}_{\text{grid}} = \mathcal{P}_{N\times N}(\mathcal{D}^{\ell}_{\text{batch}}, \mathcal{D}^{u}_{\text{batch}}). \tag{5}$$

yielding partitioned labeled data $\mathcal{D}^{\ell}_{\text{grid}} = \{(\boldsymbol{x}^{\ell}_{i,j}, \boldsymbol{y}^{\ell}_{i,j})\}^{B}_{i=1}{}^{N^2}_{,j=1}$ and unlabeled data $\mathcal{D}^{u}_{\text{grid}} = \{(\boldsymbol{x}^{u}_{i,j}, \hat{\boldsymbol{y}}^{u}_{i,j})\}^{B}_{i=1}{}^{N^2}_{,j=1}$, with $K = B \times N^2$ total segments per batch. The overall strategy is illustrated in Figure 2.

2) Foreground-Background Fusion: We introduce Foreground-Background Fusion (FBF) to enhance the utilization of foreground information from labeled data. For grid cell pairs $(\boldsymbol{x}^{\ell}_{i,j}, \boldsymbol{y}^{\ell}_{i,j})$ from $\mathcal{D}^{\ell}_{\text{grid}}$ and $(\boldsymbol{x}^{u}_{i,j}, \hat{\boldsymbol{y}}^{u}_{i,j})$ from $\mathcal{D}^{u}_{\text{grid}}$, we define foreground and background masks as:

$$\boldsymbol{M}^{\ell}_{c} = \mathbb{1}(\boldsymbol{y}^{\ell}_{i,j} = c), \quad \boldsymbol{M}^{u}_{0} = \mathbb{1}(\hat{\boldsymbol{y}}^{u}_{i,j} = 0), \tag{6}$$

where $c \in 1, 2, \ldots, C$ denotes foreground classes. The mixed data is generated by:

$$\boldsymbol{x}^{m}_{i,p} = \boldsymbol{x}^{u}_{i,p} \odot (1 - \boldsymbol{M}^{\ell}_{c} \wedge \boldsymbol{M}^{u}_{0}) + \boldsymbol{x}^{\ell}_{i,p} \odot (\boldsymbol{M}^{\ell}_{c} \wedge \boldsymbol{M}^{u}_{0}), \tag{7}$$

$$\boldsymbol{y}^{m}_{i,p} = \hat{\boldsymbol{y}}^{u}_{i,p} \odot (1 - \boldsymbol{M}^{\ell}_{c} \wedge \boldsymbol{M}^{u}_{0}) + \boldsymbol{y}^{\ell}_{i,p} \odot (\boldsymbol{M}^{\ell}_{c} \wedge \boldsymbol{M}^{u}_{0}), \tag{8}$$

where $\odot$ denotes element-wise multiplication and $\wedge$ denotes logical AND. Prior to fusion, labeled data is sorted in descending order by foreground proportion while unlabeled data is sorted in ascending order, ensuring optimal matching between high-foreground labeled cells and low-foreground unlabeled cells. The mixed dataset is formally defined as:

$$\begin{aligned}\mathcal{D}^{\text{mix}}_{\text{grid}} = \{(&\boldsymbol{x}^{\text{m}}_{i,j} = \boldsymbol{x}^{u}_{i,j} \odot \overline{\boldsymbol{M}}_{i,j} + \boldsymbol{x}^{\ell}_{i,j} \odot \boldsymbol{M}_{i,j},\\ &\boldsymbol{y}^{\text{m}}_{i,j} = \hat{\boldsymbol{y}}^{u}_{i,j} \odot \overline{\boldsymbol{M}}_{i,j} + \boldsymbol{y}^{\ell}_{i,j} \odot \boldsymbol{M}_{i,j})\}^{B}_{i=1}{}^{N^2}_{,j=1},\end{aligned} \tag{9}$$

where $\boldsymbol{M}_{i,j} = \mathbb{1}(\boldsymbol{y}^{\ell}_{i,j} = c) \wedge \mathbb{1}(\hat{\boldsymbol{y}}^{u}_{i,j} = 0)$ is the fusion mask, while $\overline{\boldsymbol{M}}_{i,j} = \mathbf{1} - \boldsymbol{M}_{i,j}$ is complement. Finally, the inverse operation $\mathcal{P}^{-1}$ reconstructs the mixed data to the original:

$$\mathcal{D}^{\text{mix}} = \mathcal{P}^{-1}_{N\times N}(\mathcal{D}^{\text{mix}}_{\text{grid}}) \tag{10}$$

yielding $\mathcal{D}_{\text{mix}} = \{(\boldsymbol{x}^{\text{m}}_i, \boldsymbol{y}^{\text{m}}_i)\}^{K}_{i=1}$, which integrates global and local features to enhance learning of both comprehensive relational and discriminative features.

C. Optimization Objective

The loss function optimizes the student model during self-training. For both original and mixed data, predictions from the dual segmentation heads are used to compute a combined loss of Dice and cross-entropy: $\mathcal{L}_{\text{seg}} = \mathcal{L}_{\text{dice}} + \mathcal{L}_{\text{ce}}$. Using

TABLE II: Comparisons with SOTA on the LA dataset.

Method	Labeled	Unlabeled	Dice ↑	Jaccard ↑	95HD ↓	ASD ↓
V-Net	5%	0	52.55	39.60	47.05	9.87
V-Net	10%	0	82.74	71.72	13.35	3.26
V-Net	All	0	91.47	84.36	5.48	1.51
UA-MT	5%	95%	82.26	70.98	13.71	3.82
SASSNet			81.60	69.63	16.16	3.58
DTC			81.25	69.33	14.90	3.99
URPC			82.48	71.35	14.65	3.65
MC-Net			83.59	72.36	14.07	2.70
SS-Net			86.33	76.15	9.97	2.31
BCP			88.02	78.72	7.90	2.15
Ours			**88.82**	**79.96**	**7.68**	**1.93**
UA-MT	10%	90%	87.79	78.39	8.68	2.12
SASSNet			87.54	78.05	9.84	2.59
DTC			87.51	78.17	8.23	2.36
URPC			86.92	77.03	11.13	2.28
MC-Net			87.62	78.25	10.03	1.82
SS-Net			88.55	79.62	7.49	1.90
BCP			89.62	81.31	6.81	**1.76**
Ours			**90.22**	**82.28**	**6.08**	1.83

TABLE III: Comparisons with SOTA on the PROMISE12 dataset.

Method	Labeled	Unlabeled	Dice ↑	ASD ↓
U-Net	20%	0	60.88	13.87
	100%	0	84.76	1.58
CCT	20%	80%	71.43	16.61
URPC			63.23	4.33
SS-Net			62.31	4.36
SLC-Net			68.31	4.69
SCP-Net			77.06	3.52
ABD			82.06	1.33
Ours			**83.73**	**1.13**
ABD	10%	90%	81.81	1.46
Ours			**82.44**	**1.26**

$\mathcal{L}_{\text{seg}}$, the losses for labeled data and mixed data are defined as follows, respectively.

$$\begin{aligned}\mathcal{L}_{\text{lab}} &= \mathcal{L}_{\text{seg}}(\mathcal{H}_s^o(\boldsymbol{x}_i^\ell), \boldsymbol{y}_i^\ell)) + \mathcal{L}_{\text{seg}}(\mathcal{H}_s^p(\boldsymbol{x}_i^\ell), \boldsymbol{y}_i^\ell)),\\ \mathcal{L}_{\text{mix}} &= \mathcal{L}_{\text{seg}}(\mathcal{H}_s^o(\boldsymbol{x}_i^{\text{m}}), \boldsymbol{y}_i^{\text{m}})) + \mathcal{L}_{\text{seg}}(\mathcal{H}_s^p(\boldsymbol{x}_i^{\text{m}}), \boldsymbol{y}_i^{\text{m}})).\end{aligned} \tag{11}$$

The total loss is a weighted sum:

$$\mathcal{L}_{\text{all}} = \alpha\mathcal{L}_{\text{lab}} + \beta\mathcal{L}_{\text{mix}}, \tag{12}$$

where α and β balance the labeled and mixed data losses.

V. Experiments

A. Datasets and Metrics

Our experiments are conducted on three publicly available datasets. The LA dataset from the Atrial Segmentation Challenge [28] comprises 100 3D GE-MRI scans with ground truth labels. All scans have an isotropic resolution of 0.625 $\times$ 0.625 $\times$ 0.625 mm^3. We use 80 scans for training and 20 for validation. The ACDC dataset [29] from the 2017 Automated Cardiac Diagnosis Challenge contains 100 cardiac MRI sequences (256 $\times$ 256) for four-class segmentation: *background*, *right ventricle*, *left ventricle*, and *myocardium*. Data is divided into 70 training, 10 validation, and 20 test cases. The PROMISE12 dataset [30] from the MICCAI 2012 prostate segmentation challenge contains 50 MRI scans from patients with various diseases, acquired across multiple centers. The data is split into 35 training and 15 test cases. We employ four standard evaluation metrics: Dice Score (%), Jaccard Score (%), 95% Hausdorff Distance (95HD, voxels), and Average Surface Distance (ASD, voxels).

TABLE IV: Ablation study of GFR and DSHA on the ACDC.

Method	Labeled	Unlabeled	Dice ↑	Jaccard ↑	95HD ↓	ASD ↓
U-Net+DSHA	5%	95%	73.93	63.80	9.98	2.36
U-Net+GFR			87.66	78.72	5.39	1.64
U-Net+GFR+DSHA			**87.95**	**79.19**	**6.67**	**1.86**
U-Net+DSHA	10%	90%	88.17	79.66	2.18	1.77
U-Net+GFR			89.36	81.47	7.55	1.96
U-Net+GFR+DSHA			**90.35**	**82.91**	**2.14**	**0.76**

TABLE V: Ablation study of EAW on ACDC dataset. α_G refers to the weights of the outputs from $\mathcal{H}_s^o$, and α_L refers to the weights of the outputs from $\mathcal{H}_s^p$.

α_G	α_L	Labeled	Unlabeled	Dice ↑	Jaccard ↑	95HD ↓	ASD ↓
0.7	0.3	5%	95%	85.10	75.77	5.99	1.68
0.6	0.4			85.88	76.30	5.34	1.84
0.5	0.5			87.74	78.83	7.33	2.08
Adaptive				**87.78**	**78.88**	**4.46**	**1.47**
0.7	0.3	10%	90%	89.92	82.20	2.95	1.08
0.6	0.4			90.16	82.62	2.31	0.72
0.5	0.5			89.71	81.92	2.15	0.82
Adaptive				**90.46**	**83.06**	**1.58**	**0.57**

B. Implementation Details

We use SGD with an initial learning rate of 10^{-2} and weight decay of 10^{-4} across all datasets. For the LA dataset, we employ a 3D V-Net with a batch size of 8, dividing each 112×112×80 volume into 8 patches of 56×56×40. For ACDC [15] and PROMISE12, we use a batch size of 24. ACDC slices (256×256) are divided into 16 patches of 64×64, while PROMISE12 slices use a 2×2 partition, yielding 4 patches of 128×128.

C. Comparison with Sate-of-the-Art Methods

We conduct comprehensive evaluations on three medical datasets: ACDC, LA, and PROMISE12. Quantitative comparisons with state-of-the-art methods [15], [23], [31]–[38] are summarized in Tables I, II, and III. Across all datasets and under various labeled data ratios (5%, 10%, and 20%), our method consistently demonstrates superior performance compared to existing approaches. Qualitative results (Fig. 3, 4) further confirm the effectiveness of our approach. In particular, compared to other methods that often exhibit blurry boundaries, incomplete segmentation, and artifacts, our method produces more accurate delineations and preserves finer structural details. These visual comparisons highlight our method's enhanced capability in handling complex anatomical variations and challenging foreground-background distributions across different medical imaging modalities.

D. Ablation Studies

Effects of different components Ablation experiments are conducted on the ACDC dataset to evaluate the contributions of key components in the proposed method, with a particular focus on the DSHA and GFR. As shown in Table IV, under the 5% labeled setting, when the DSHA is adopted, introducing the GFR strategy leads to a substantial performance improvement, resulting in a Dice score increase of 14.02% on the ACDC dataset. Moreover, under the 10% labeled setting, with DSHA in place, the incorporation of GFR further enhances the segmentation performance, yielding an additional Dice gain of

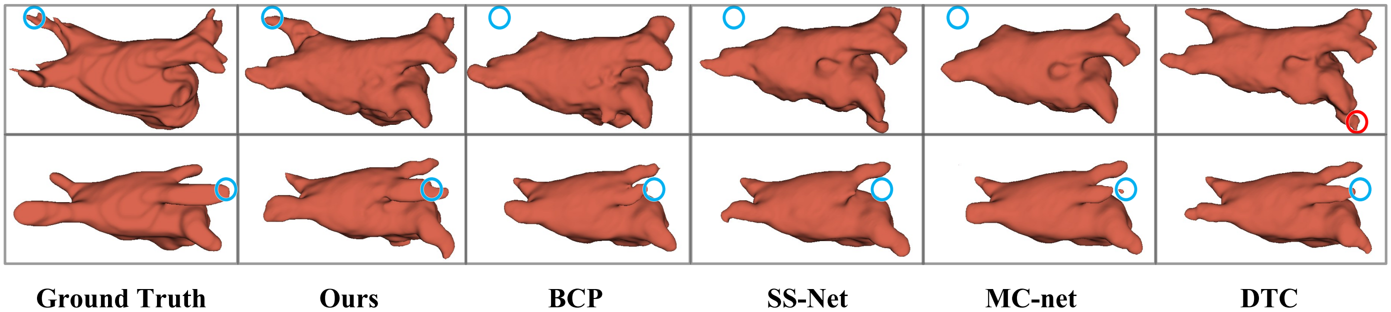

Fig. 3: The vision of LA dataset. Blue circles denote regions that are missing compared to the Ground Truth, whereas red circles highlight regions that are predicted in excess relative to the Ground Truth.

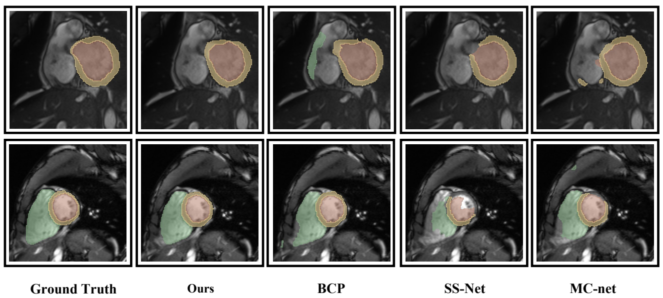

Fig. 4: The vision of the ACDC dataset.

TABLE VI: Ablation study of Grid cell Size (2 × 2 vs 4 × 4) on ACDC and PROMISE12 datasets with 10% labeled data

Dataset	Partition	Labeled	Unlabeled	Dice ↑	Jaccard ↑	95HD ↓	ASD ↓
ACDC	2×2	10%	90%	89.36	81.27	5.73	1.62
	4×4			**90.46**	**83.06**	**1.58**	**0.57**
PROMISE12	2×2	10%	90%	**82.44**	**70.29**	**3.01**	**1.26**
	4×4			81.74	69.41	15.79	4.53

TABLE VII: Ablation Study of Grid cell Size with 4 × 4 Partitioning on ACDC datasets.

Labeled	Unlabeled	$\mathcal{P}$	FBF	$\mathcal{P}^{-1}$	Dice ↑	Jaccard ↑	95HD ↓	ASD ↓
5%	95%		✓		69.19	56.55	21.38	6.27
		✓		✓	84.93	74.93	9.00	2.89
		✓	✓		85.48	75.58	7.76	2.34
		✓	✓	✓	**87.78**	**78.88**	**4.46**	**1.47**
10%	90%		✓		86.08	76.37	10.54	3.09
		✓		✓	89.24	81.20	4.75	1.31
		✓	✓		89.98	82.30	3.04	1.05
		✓	✓	✓	**90.35**	**82.91**	**2.14**	**0.76**

0.99%. Overall, the ablation results confirm the effectiveness of both the proposed GFR strategy and DSHA in enhancing segmentation performance.

Entropy-Aware Weighting To validate the effectiveness of our proposed EAW strategy for DSHA outputs, we conducted a comprehensive ablation study comparing two approaches: fixed weight and EAW. Our results show that EAW dynamically adjusts fusion weights by entropy levels, achieving superior Dice and Jaccard scores (see Table V).

Grid cell Size We conduct ablation studies on the grid cell size in our Grid-based Foreground Reorganization module. Specifically, we evaluate the impact of different grid configurations (N × N) where $N \in \{2, 4\}$ under 10% annotated data. As shown in Table VI, on the ACDC dataset, the 4 × 4 grid partition improved the Dice score by 1.10% and the Jaccard score by 1.79% compared to the 2 × 2 partition. In contrast, on the PROMISE12 dataset, the 2 × 2 partition performed better, increasing the Dice score by 0.70% and the Jaccard score by 0.88% over the 4 × 4 partition. Based on the experimental results, we think that the ACDC dataset contains more diverse foreground classes, which require finer-grained partitioning, whereas the PROMISE12 dataset involves more simple and single foreground classes, making it more suitable for coarser partitioning strategies.

Grid-based Foreground Reorganization In this section, we conduct a comprehensive ablation study on the three key components of the GFR module. Based on the results of the previous ablation study, a grid cell size of 4 × 4 is adopted for all experiments on the ACDC dataset. As demonstrated in Table VII, disabling the Grid Partitioning module and applying only FBF (Foreground-Background Fusion) to oringinal input data leads to the worst performance, which is consistently observed under both 5% and 10% annotation ratios. The experimental results demonstrate that incorporating the Grid Partitioning module leads to substantial performance improvements.

VI. Conclusion

This paper proposes a novel semi-supervised medical image segmentation framework, SDRF-Net, which targets the challenges of foreground–background imbalance and inaccurate pseudo-labels. It incorporates the GFR module and an DSHA. Specifically, GFR reconstructs local feature distributions to capture rare foreground patterns, while DSHA adaptively fuses global and local information to enhance prediction robustness. Through the synergistic effect of these two modules, our method effectively addresses the above issues and achieves SOTA performance on multiple benchmark datasets.

References

[1] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60–88, 2017.

[2] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 fourth international conference on 3D vision (3DV)*. Ieee, 2016, pp. 565–571.

[3] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*. Springer, 2015, pp. 234–241.

[4] L. Su, Z. Wang, Y. Shi, A. Li, and M. Wang, "Local augmentation based consistency learning for semi-supervised pathology image classification," *Computer Methods and Programs in Biomedicine*, vol. 232, p. 107446, 2023.

[5] Q. Dou, Q. Liu, P. A. Heng, and B. Glocker, "Unpaired multi-modal segmentation via knowledge distillation," *IEEE transactions on medical imaging*, vol. 39, no. 7, pp. 2415–2425, 2020.

[6] X. Li, H. Chen, X. Qi, Q. Dou, C.-W. Fu, and P.-A. Heng, "H-denseunet: hybrid densely connected unet for liver and tumor segmentation from ct volumes," *IEEE transactions on medical imaging*, vol. 37, no. 12, pp. 2663–2674, 2018.

[7] X. Luo, "SSL4MIS," https://github.com/HiLab-git/SSL4MIS, 2020.

[8] J. Su, Z. Luo, S. Lian, D. Lin, and S. Li, "Mutual learning with reliable pseudo label for semi-supervised medical image segmentation," *Medical Image Analysis*, vol. 94, p. 103111, 2024.

[9] M. Wajid, A. Iqbal, I. Malik, S. J. Hussain, and Y. Jan, "A semi-supervised approach for breast tumor segmentation using sparse transformer attention unet," *Pattern Recognition Letters*, vol. 187, pp. 63–72, 2025.

[10] J. Yuan, J. Ge, Z. Wang, and Y. Liu, "Semi-supervised semantic segmentation with mutual knowledge distillation," in *Proceedings of the 31st ACM International Conference on Multimedia*, 2023, pp. 5436–5444.

[11] S. Adiga, J. Dolz, and H. Lombaert, "Anatomically-aware uncertainty for semi-supervised image segmentation," *Medical Image Analysis*, vol. 91, p. 103011, 2024.

[12] W. Bai, O. Oktay, M. Sinclair, H. Suzuki, M. Rajchl, G. Tarroni, B. Glocker, A. King, P. M. Matthews, and D. Rueckert, "Semi-supervised learning for network-based cardiac mr image segmentation," in *Medical Image Computing and Computer-Assisted Intervention- MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, Proceedings*. Springer, 2017, pp. 253–260.

[13] X. Chen, Y. Yuan, G. Zeng, and J. Wang, "Semi-supervised semantic segmentation with cross pseudo supervision," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2021, pp. 2613–2622.

[14] X. Li, L. Yu, H. Chen, C.-W. Fu, L. Xing, and P.-A. Heng, "Transformation-consistent self-ensembling model for semisupervised medical image segmentation," *IEEE transactions on neural networks and learning systems*, vol. 32, no. 2, pp. 523–534, 2020.

[15] Y. Bai, D. Chen, Q. Li, W. Shen, and Y. Wang, "Bidirectional copy-paste for semi-supervised medical image segmentation," *CoRR*, vol. abs/2305.00673, 2023.

[16] Y. Wang, B. Xiao, X. Bi, W. Li, and X. Gao, "Mcf: Mutual correction framework for semi-supervised medical image segmentation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2023, pp. 15 651–15 660.

[17] Y. Chen, Y. Ma, X. Mei, L. Zhang, Z. Fu, and J. Ma, "Triple-task mutual consistency for semi-supervised 3d medical image segmentation," *Computers in Biology and Medicine*, vol. 175, p. 108506, 2024.

[18] Y. Wu, Z. Ge, D. Zhang, M. Xu, L. Zhang, Y. Xia, and J. Cai, "Mutual consistency learning for semi-supervised medical image segmentation," *Medical Image Analysis*, vol. 81, p. 102530, 2022.

[19] A. Das, C. Gautam, H. Cholakkal, P. Agrawal, F. Yang, R. Savitha, and Y. Liu, "Decoupled training for semi-supervised medical image segmentation with worst-case-aware learning," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2024, pp. 45–55.

[20] H. Zhang, "mixup: Beyond empirical risk minimization," *arXiv preprint arXiv:1710.09412*, 2017.

[21] S. Yun, D. Han, S. J. Oh, S. Chun, J. Choe, and Y. Yoo, "Cutmix: Regularization strategy to train strong classifiers with localizable features," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 6023–6032.

[22] P. Tu, Y. Huang, F. Zheng, Z. He, L. Cao, and L. Shao, "Guidedmixnet: Semi-supervised semantic segmentation by using labeled images as reference," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 2, 2022, pp. 2379–2387.

[23] H. Chi, J. Pang, B. Zhang, and W. Liu, "Adaptive bidirectional displacement for semi-supervised medical image segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 4070–4080.

[24] J. Miao, C. Chen, F. Liu, H. Wei, and P.-A. Heng, "Caussl: Causality-inspired semi-supervised learning for medical image segmentation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 21 426–21 437.

[25] R. Mendel, L. A. De Souza, D. Rauber, J. P. Papa, and C. Palm, "Semi-supervised segmentation based on error-correcting supervision," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXIX 16*. Springer, 2020, pp. 141–157.

[26] H. Huang, Y. Huang, S. Xie, L. Lin, R. Tong, Y.-W. Chen, Y. Li, and Y. Zheng, "Combinatorial cnn-transformer learning with manifold constraints for semi-supervised medical image segmentation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 3, 2024, pp. 2330–2338.

[27] Z. Zhao, Z. Wang, L. Wang, D. Yu, Y. Yuan, and L. Zhou, "Alternate diverse teaching for semi-supervised medical image segmentation," in *European Conference on Computer Vision*. Springer, 2025, pp. 227–243.

[28] Z. Xiong, Q. Xia, Z. Hu, N. Huang, C. Bian, Y. Zheng, S. Vesal, N. Ravikumar, A. Maier, X. Yang *et al.*, "A global benchmark of algorithms for segmenting the left atrium from late gadolinium-enhanced cardiac magnetic resonance imaging," *Medical image analysis*, vol. 67, p. 101832, 2021.

[29] O. Bernard, A. Lalande, C. Zotti, F. Cervenansky, X. Yang, P.-A. Heng, I. Cetin, K. Lekadir, O. Camara, M. A. G. Ballester *et al.*, "Deep learning techniques for automatic mri cardiac multi-structures segmentation and diagnosis: is the problem solved?" *IEEE Transactions on medical imaging*, vol. 37, no. 11, pp. 2514–2525, 2018.

[30] G. Litjens, R. Toth, W. Van De Ven, C. Hoeks, S. Kerkstra, B. Van Ginneken, G. Vincent, G. Guillard, N. Birbeck, J. Zhang *et al.*, "Evaluation of prostate segmentation algorithms for mri: the promise12 challenge," *Medical image analysis*, vol. 18, no. 2, pp. 359–373, 2014.

[31] L. Yu, S. Wang, X. Li, C.-W. Fu, and P.-A. Heng, "Uncertainty-aware self-ensembling model for semi-supervised 3d left atrium segmentation," in *Proceedings of the 22nd International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2019)*. Shenzhen, China: Springer, 2019, pp. 605–613.

[32] S. Li, C. Zhang, and X. He, "Shape-aware semi-supervised 3d semantic segmentation for medical images," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part I 23*. Springer, 2020, pp. 552–561.

[33] X. Luo, J. Chen, T. Song, and G. Wang, "Semi-supervised medical image segmentation through dual-task consistency," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 10, 2021, pp. 8801–8809.

[34] X. Luo, W. Liao, J. Chen, T. Song, Y. Chen, S. Zhang, N. Chen, G. Wang, and S. Zhang, "Efficient semi-supervised gross target volume of nasopharyngeal carcinoma segmentation via uncertainty rectified pyramid consistency," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*, 2021, pp. 318–329.

[35] Y. Wu, Z. Wu, Q. Wu, Z. Ge, and J. Cai, "Exploring smoothness and class-separation for semi-supervised medical image segmentation," in *International conference on medical image computing and computer-assisted intervention*. Springer, 2022, pp. 34–43.

[36] Y. Ouali, C. Hudelot, and M. Tami, "Semi-supervised semantic segmentation with cross-consistency training," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 12 674–12 684.

[37] J. Liu, C. Desrosiers, and Y. Zhou, "Semi-supervised medical image segmentation using cross-model pseudo-supervision with shape awareness and local context constraints," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2022, pp. 140–150.

[38] Z. Zhang, R. Ran, C. Tian, H. Zhou, X. Li, F. Yang, and Z. Jiao, "Self-aware and cross-sample prototypical learning for semi-supervised medical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2023, pp. 192–201.