Synergy-Guided Regional Supervision of Pseudo Labels for Semi-Supervised Medical Image Segmentation

Tao Wang[1], Xinlin Zhang[1], Yuanbin Chen[1], Yuanbo Zhou[1], Longxuan Zhao[1], Tao Tan[2], and Tong Tong[1]([✉])

[1] College of physics and information engineering, Fuzhou University, Fujian China
ttraveltong@gmail.com
[2] Faculty of Applied Science, Macao Polytechnic University, Macao, China.

Abstract. Semi-supervised learning has received considerable attention for its potential to leverage abundant unlabeled data to enhance model robustness. Despite the widespread adoption of pseudo labeling in semi-supervised learning, existing methods often suffer from noise contamination, which can undermine the robustness of the model. To tackle this challenge, we introduce a novel Synergy-Guided Regional Supervision of Pseudo Labels (SGRS-Net) framework. Built upon the mean teacher network, we employ a Mix Augmentation module to enhance the unlabeled data. By evaluating the synergy before and after augmentation, we strategically partition the pseudo labels into distinct regions. Additionally, we introduce a Region Loss Evaluation module to assess the loss across each delineated area. Extensive experiments conducted on the LA, Pancreas-CT and BraTS2019 dataset have demonstrated superior performance over current state-of-the-art techniques, underscoring the efficiency and practicality of our framework. The code is available at https://github.com/ortonwang/SGRS-Net.

Keywords: Medical image segmentation · Semi-supervised learning · Pseudo label.

1 Introduction

With the rapid development of deep learning, fully supervised learning segmentation methods, such as the U-Net [17], and UNeXt [19] have achieved success. However, these methods often depend on a large amount of precisely annotated training data, which can be both expensive and time-consuming to obtain. Such a dependency constrains the scalability and applicability of medical image segmentation methods. In recent years, semi-supervised learning has attracted widespread attention for its potential to utilize fewer labeled data alongside a wealth of unlabeled data, thereby enhancing the model's generalization capability. Numerous semi-supervised medical image analysis methods have been introduced, including pseudo labels [2], deep co-training [27] [25], deep adversarial learning [8], the mean teacher and its extensions [18][9] among others. These

methods effectively leverage both labeled and unlabeled data to develop powerful models.

Pseudo labels learning is a widely adopted method. This approach initiates by training the model on labeled data, then generating pseudo labels for unlabeled data using predicted probability maps. Subsequently, these pseudo labels are integrated with the labeled data for further training of the model, aimed at enhancing the accuracy and generalization ability of the image [7]. Building on this approach, Shanfu Lu et al. [10] have introduced two auxiliary decoders into a network to generate pseudo labels from these auxiliary decoders. Moreover, Luo et al. [12] have developed the Cross Teaching framework, where the prediction of one network serves as the pseudo label to directly guide another network in an end-to-end manner.

While these methods effectively enhance the performance of models, the inherent inaccuracy of pseudo labels, when compared to ground truth, remains a challenge. Indeed, some pseudo labels are of high quality and comparable to ground truth, yet others may bolster by noise. As iterative training progresses, this noise has the potential to adversely affect the robustness of the models. Although directly applying noise-robust loss functions can mitigate the effects of noise to some extent, in the case of high-quality pseudo labels, endeavoring to optimize against unnecessary noise suppression may lead to overfitting.

To effectively utilize pseudo labels, we introduce the Synergy-Guided Regional Supervision (SGRS-Net). During training, a Pseudo Label Generation Module is employed to generate pseudo label. Additionally, we introduce a Mix Augmentation Module which utilize annotated data to augment unlabeled data. Following this, our Synergy Evaluation (SE) module allows us to partition pseudo labels into different regions and then apply Regional Loss Evaluation module to evaluate the corresponding loss for each region. This strategic approach is designed to maximize the use of pseudo label while minimizing the impact of potential noise.

In summary, our contributions are multifaceted: (1) We introduce a Synergy-Guided Regional Supervision (SGRS-Net) framework for semi-supervised medical image segmentation. (2) This framework introduces a Pseudo Label Generation Module, and a Mix Augmentation (MA) module to significantly enhance the diversity of the unlabeled dataset. (3) We introduce a Synergy Evoluation (SE) module and a Regional Loss Evaluation (RLE) module, designed to mitigate the impact of noise while fully capitalizing on the supervisory signal provided by pseudo label. (4) Experimental results on the LA, Pancreas-CT, and BraTS2019 datasets demonstrate significant improvements over previous state-of-the-art (SOTA) methods, underscoring our framework's remarkable efficacy in scenarios with a limited number of labeled images.

2   Method

This study introduces the SGRS-Net for semi-supervised medical image segmentation. For clarity, the notions and notations are summarized in Table 1.
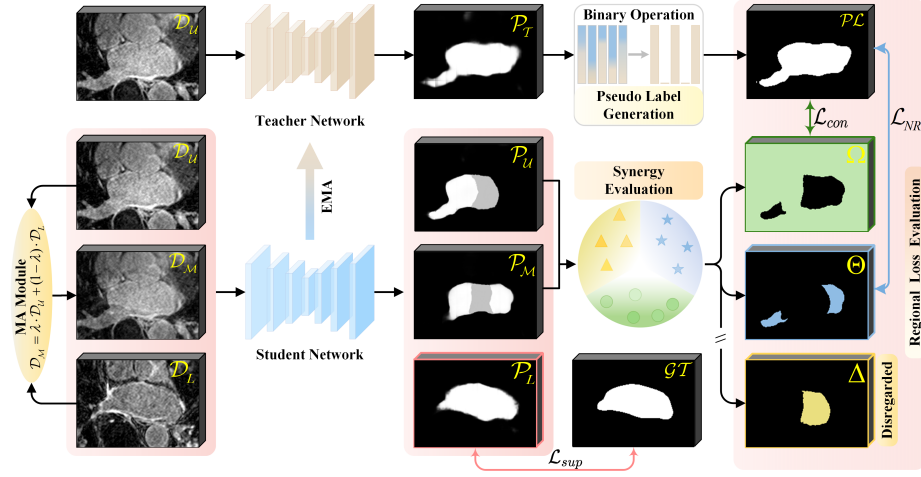
Fig. 1: Overview of our proposed SGRS-Net.

Notations	Descriptions
$\mathcal{D}_L,\mathcal{GT}$	Labeled Image and correspond Ground Truth
$\mathcal{D}_\mathcal{U},\mathcal{PL}$	Unlabeled Image and correspond Pseudo Label
$\mathcal{D}_\mathcal{M}$	Image mixed from $\mathcal{D}_\mathcal{U}$ and $\mathcal{D}_L$
θ_S,θ_T	Parameters of the student network and the teacher network
Δ,Ω,Θ	Regions regard as disregarded, consistent, and inconsistent, respectively

Table 1: Summary of mathematical notions and corresponding notations.

Overall architecture design: Fig. 1 illustrates the framework of the SGRS-Net. Throughout the training, the teacher network implements parameter updates from the student network through exponential moving averages (EMA). We begin by assessing unlabeled data with the teacher network, obtaining $\mathcal{PL}$. Subsequently, the Mix Augmentation (MA) module is employed to enhance $\mathcal{D}_\mathcal{U}$, resulting in $\mathcal{D}_\mathcal{M}$. Upon processing by the student network on $(\mathcal{D}_L, \mathcal{D}_\mathcal{U}, \mathcal{D}_\mathcal{M})$, we derive the corresponding predictions $(\mathcal{P}_L, \mathcal{P}_\mathcal{U}, \mathcal{P}_\mathcal{M})$. After that, the Synergy Evaluation (SE) module is employed to evaluate the synergy between $\mathcal{P}_\mathcal{U}$ and $\mathcal{P}_\mathcal{M}$, and partition the $\mathcal{P}_L$ into three regions: Δ, Ω, and Θ. Finally, the Regional Loss Evaluation (RLE) module is used to evaluate the losses. Detailed explanations of each module will be provided in the subsequent sections.

Pseudo Label Generation Module: To obtain pseudo label for $\mathcal{D}_\mathcal{U}$, this study adopts a mean teacher framework. The teacher network is utilized to evaluate the $\mathcal{D}_\mathcal{U}$, i.e., $\mathcal{P}_\mathcal{T} = f_{\theta_T}(\mathcal{D}_U)$, and through a binarization process, the corresponding $\mathcal{PL}$ is generated, i.e., $\mathcal{PL} = \mathrm{ArgMax}(\mathcal{P}_\mathcal{T})$.

Mix Augmentation Module: To enhance the diversity of the dataset, methods such as CutMix [24], which alters local regions, and PolypMix [5], which targets lesion-specific areas, have shown promising results. However, these techniques may disrupt global pixel-level consistency when evaluating synergistic

effects. Therefore, we adopt the Mix-Up [26] strategy for data augmentation. For a pair of samples $\mathcal{D}_{\mathcal{U}}$ and $\mathcal{D}_L$, this process can be formulated as follows:

$$\lambda = max[Beta(\alpha, \alpha), 1 - Beta(\alpha, \alpha)] \tag{1}$$

$$\mathcal{D}_{\mathcal{M}} = \lambda \cdot \mathcal{D}_{\mathcal{U}} + (1 - \lambda) \cdot \mathcal{D}_L \tag{2}$$

where α is a randomly generated hyperparameter. It is noteworthy that our method primarily focuses on $\mathcal{D}_{\mathcal{U}}$, incorporating $\mathcal{D}_L$ to increase the diversity of $\mathcal{D}_{\mathcal{U}}$, without the need for additional annotation. Since the coefficient λ is mostly close to 1, the noise artifacts introduced by Mix-Up are minimal and have negligible impact. Therefore, we use the $\mathcal{PL}$ corresponding to the $\mathcal{D}_{\mathcal{U}}$ as the pseudo label for $\mathcal{D}_M$ in the subsequent steps.

Synergy Evaluation Module: To improve the utilization of valuable information in pseudo label while mitigating the impact of noise, we introduce a SE module to divide the pseudo label into three parts: Δ, Θ, and Ω, based on their synergy. Specifically, we begin by assessing the information entropy of $\mathcal{P}_{\mathcal{U}}$ and $\mathcal{P}_{\mathcal{M}}$. As higher entropy signals greater uncertainty, we introduce a threshold, τ. Regions exhibiting entropy above τ are considered unreliable and are thus categorized as Δ. The procedures for this categorization are formulated as follows:

$$\mathcal{P}_{\mathcal{U}}, \ \mathcal{P}_{\mathcal{M}} = \text{SoftMax}(f_{\theta_S}(\mathcal{D}_{\mathcal{U}})), \ \text{SoftMax}(f_{\theta_S}(\mathcal{D}_{\mathcal{M}})) \tag{3}$$

$$\text{Ent}(p) = -\sum_{j \in C} p_j \cdot \log p_j \tag{4}$$

$$\Delta = (\text{Ent}(\mathcal{P}_{\mathcal{U}}) > \tau) \cap (\text{Ent}(\mathcal{P}_{\mathcal{M}}) > \tau) \tag{5}$$

Where C represents the number of categories, and p_j denotes the probability of pixels for the j-th category. For regions that are clearly distinguishable with consistent results and inconsistent categories are identified as consistent and inconsistent predictions, labeled as Ω and Θ, respectively. The specific operations are formulated as follows:

$$\mathcal{A}_{\mathcal{U}}, \ \mathcal{A}_{\mathcal{M}} = \text{ArgMax}(\mathcal{P}_{\mathcal{U}}), \ \text{ArgMax}(\mathcal{P}_{\mathcal{M}}) \tag{6}$$

$$\Omega = (\mathcal{A}_{\mathcal{U}} \odot \mathcal{A}_{\mathcal{M}}) \wedge (\neg\Delta), \qquad \Theta = (\mathcal{A}_{\mathcal{U}} \oplus \mathcal{A}_{\mathcal{M}}) \wedge (\neg\Delta) \tag{7}$$

Regional Loss Evaluation Module: After synergy assessment, the predictions are divided into three segments: Δ, Ω, Θ. Subsequently, we conduct a regional loss evaluation for each category. Initially, for regions Δ, where a higher level of uncertainty is noted, the $\mathcal{PL}$ within this area is considered more susceptible to noise. Therefore, this region is excluded from the evaluation process.

Learning from $\mathcal{GT}$: During training, the $\mathcal{GT}$ serves as the gold standard for $\mathcal{D}_L$, and the loss is evaluated using $\mathcal{L}_{sup}$. Based on the Cross Entropy Loss and Dice Loss, the specific function is defined as follows:

$$\mathcal{L}_{pce}^{\mathbb{U}}(p, y) = -\sum_{i \in \mathbb{U}} \sum_{j \in C} y[i, j] \cdot \log p[i, j] \tag{8}$$

$$\mathcal{L}_{pdc}^{\mathbb{U}}(p,y) = 1 - \frac{2 \times \sum\limits_{i \in \mathbb{U}} \sum\limits_{j \in C} p[i,j] \cdot y[i,j]}{\sum\limits_{i \in \mathbb{U}} \sum\limits_{j \in C} p[i,j]^2 + \sum\limits_{i \in \mathbb{U}} \sum\limits_{j \in C} y[i,j]^2} \tag{9}$$

$$\mathcal{L}_{sup} = \mathcal{L}_{pce}^{\mathbb{U}}(\mathcal{P}_L, \mathcal{GT}) + \mathcal{L}_{pdc}^{\mathbb{U}}(\mathcal{P}_L, \mathcal{GT}) \tag{10}$$

where $\mathbb{U}$ represents all the pixels within $\mathcal{GT}$. The $p[i,j]$ and $y[i,j]$ denote the predicted probability and ground truth for the pixel i with class j, respectively.
Learning from Ω: For regions identified as Ω, characterized by lower entropy and consistent predictions, we introduce a local consistency supervision loss $\mathcal{L}_{con}$, which learns from unlabeled data by minimizing the difference between the prediction and the $\mathcal{PL}$ within Ω. This is implemented using Cross Entropy Loss and Dice Loss. Based on equation 8 and 9, the $\mathcal{L}_{con}$ is formulated as:

$$\mathcal{L}_c^{\Omega}(p,y) = \mathcal{L}_{pce}^{\Omega}(p,y) + \mathcal{L}_{pdc}^{\Omega}(p,y) \tag{11}$$

$$\mathcal{L}_{con} = \mathcal{L}_c^{\Omega}(\mathcal{P}_{\mathcal{U}}, \mathcal{PL}) + \mathcal{L}_c^{\Omega}(\mathcal{P}_{\mathcal{M}}, \mathcal{PL}) \tag{12}$$

Learning from Θ: For regions within Θ, where lower entropy is observed, indicating a degree of stability, yet discrepancies exist in the predicted categories. This observation suggests that the corresponding $\mathcal{PL}$ within this region may contain noise. To address this issue, we propose a regional noise robust loss function, denoted as $\mathcal{L}_{NR}$, designed to mitigate the influence of noise within Θ. The formulation of $\mathcal{L}_{NR}$ is presented as follows:

$$\mathcal{L}_{psce}^{\Theta}(p,y) = -\sum_{i \in \Theta} \sum_{j \in C} [(1-\varepsilon) \cdot y[i,j] \cdot \log(p[i,j]) + \frac{\varepsilon}{C} \log(p[i,j])] \tag{13}$$

$$\mathcal{L}_{psdc}^{\Theta}(p,y) = 1 - \frac{2 \times \sum\limits_{i \in \Theta} \sum\limits_{j \in C} p[i,j] \cdot y[i,j]}{\sum\limits_{i \in \Theta} \sum\limits_{j \in C} p[i,j]^2 + \sum\limits_{i \in \Theta} \sum\limits_{j \in C} y[i,j]^2 + \eta} \tag{14}$$

$$\mathcal{L}_s^{\Theta}(p,y) = \mathcal{L}_{psce}^{\Theta}(p,y) + \mathcal{L}_{psdc}^{\Theta}(p,y) \tag{15}$$

$$\mathcal{L}_{NR} = \mathcal{L}_s^{\Omega}(\mathcal{P}_{\mathcal{U}}, \mathcal{PL}) + \mathcal{L}_s^{\Omega}(\mathcal{P}_{\mathcal{M}}, \mathcal{PL}) \tag{16}$$

where ε and η are the smoothing parameter. The $p[i,j]$ and $y[i,j]$ denote the predicted probability and ground truth for the pixel i with class j, respectively.

2.1 Total Loss Funciton

The proposed SGRS-Net framework aims to learn from both labeled and unlabeled data, the total loss is thus articulated as:

$$\mathcal{L} = \mathcal{L}_{sup} + \lambda \cdot (\mathcal{L}_{con} + \mathcal{L}_{NR}) \tag{17}$$

where $\mathcal{L}_{sup}$, $\mathcal{L}_{con}$ and $\mathcal{L}_{NR}$ are defined in equations 10, 12 and 16, respectively. We integrate $\lambda(t)$, a widely utilized time-dependent Gaussian warming-up function [6], to modulate the balance between losses at different training stages, defined as:

$$\lambda(t) = \begin{cases} e^{(-5(1-\frac{t}{t_{warm}})^2)} & t < t_{warm} \\ 1 & t \geq t_{warm} \end{cases} \tag{18}$$

where t represents the current training step, and t_{warm} denotes the maximum warm training step.

3   Experiments and Results

3.1   Experimental Details

Dataset: The SGRS-Net was evaluated on three datasets: the**LA dataset** [23] with 100 3D GE-MRIs (80 samples for training and 20 for testing following [21]), the **Pancreas-CT dataset** [3] with 82 CT volumes (62 samples for training and 20 for testing following [21]), and the **BraTS2019 dataset** [14] with 335 MRIs (250 samples for training, 25 samples for validation and 60 for testing).
Implementation Details: Our framework was developed using PyTorch and executed on an Nvidia RTX 3090 GPU. During training, the patch size was set to $112 \times 112 \times 80$ for LA dataset and $96 \times 96 \times 96$ for Pancreas-CT and BraTS2019 dataset. For testing, the strides is $18 \times 18 \times 4$ for LA dataset and $16 \times 16 \times 16$ for Pancreas-CT and BraTS2019 dataset. We selected the V-Net [15] as the backbone to ensure a fair comparison. The SGRS-Net framework was trained for $15k$ iterations for LA and Pancreas-CT dataset and $60k$ for BraTS2019 dataset, employing the SGD optimizer with a momentum 0.9 and weight decay set to 1e-4. The initial batch size, learning rate, τ, and ε were set to 4, 1e-2, 0.296, 20, respectively. Four commonly used metrics are employed to assess the segmentation results: Dice, Jaccard (Jac), the average surface distance (ASD), and the 95% Hausdorff Distance (HD).

3.2   Result on LA, Pancreas-CT and BraTS2019 dataset

Table 2 details the performance of our model against eight recent semi-supervised learning methods across the three dataset, alongside the fully supervised V-Net model at various ratios to delineate the lower and upper bounds of performance. Remarkably, our model achieved the best results across nearly all four metrics. Particularly impressive is that on the LA and Pancreas-CT dataset, our model achieved Dice scores of 89.70% and 77.20% with 5% labeled data, representing a 2.36% and 13.55% improvement over the previously best-performing algorithm with the same amount of labeled data. Even more striking, with just 5% labeled data, the Dice score of our model exceeds that of comparative algorithms with 10% labeled data. With 10% labeled data, our approach attained Dice scores of 90.76% and 80.55%, approaching the upper bounds.

Figure 2 provides a visual comparison of segmentation results from 2D and 3D perspectives alongside the corresponding ground truth (GT) with 5% labeled data. It clearly demonstrates that our SGRS-Net excels in segmentation accuracy, particularly in 2D views, where our outcomes are notably closer to the GT than those of competing methods. These findings underscore the efficacy of our framework in leveraging unlabeled data to improve performance.

Table 2: Quantitative comparison with eight SOTA methods on the LA and Pancreas-CT dataset.

Method	Scans used		LA				Pancreas-CT				BraTS2019			
	Lb	Unlb	Dice↑	Jac↑	HD↓	ASD↓	Dice	Jac	HD	ASD	Dice	Jac	HD	ASD
V-Net	5%	0	52.55	39.60	47.05	9.87	29.32	19.61	43.67	15.42	74.28	64.42	13.44	2.60
V-Net	10%	0	78.57	66.96	21.20	6.07	54.94	40.87	47.48	17.43	78.67	68.75	10.44	2.23
V-Net	100%	0	91.62	84.60	5.40	1.64	83.76	72.48	4.46	1.07	88.58	80.34	6.19	1.36
DTC [11]			81.25	69.33	14.90	3.99	47.57	33.41	44.17	15.31	74.21	64.89	13.54	3.16
URPC [13]			82.48	71.35	14.65	3.65	45.94	34.14	48.80	23.03	78.74	68.20	17.43	4.51
SS-Net [22]			86.33	76.15	9.97	2.31	41.39	27.65	52.12	19.37	78.03	68.11	13.70	2.76
MC-Net+ [21]			82.07	70.38	20.49	5.72	32.45	21.22	58.57	24.84	78.69	68.38	16.44	4.49
CAML [4]	5%	95%	87.34	77.65	9.76	2.49	35.18	23.63	43.58	20.39	79.34	69.64	**11.02**	2.36
Co-BioNet [16]			84.30	74.67	8.33	2.38	52.82	39.20	29.46	6.16	73.27	63.12	14.37	**1.90**
BCP [1]			88.02	78.72	7.90	2.15	45.08	34.72	39.39	11.23	79.30	68.89	12.00	1.94
MCF [20]			83.34	72.20	16.65	5.46	63.65	49.72	18.06	3.97	71.82	62.60	14.66	3.96
Ours			**89.70**	**81.40**	**6.68**	**1.75**	**77.20**	**63.45**	**12.34**	**3.81**	**80.62**	**70.21**	14.58	3.64
DTC [11]			87.51	78.17	8.23	2.36	66.58	51.79	15.46	4.16	82.74	72.74	11.76	3.24
URPC [13]			85.01	74.36	15.37	3.96	73.53	59.44	22.57	7.85	84.16	74.29	11.01	2.63
SS-Net [22]			88.43	79.43	7.95	2.55	73.44	58.82	12.56	2.91	82.00	71.82	10.68	**1.82**
MC-Net+ [21]			88.96	80.25	7.93	1.86	70.00	55.66	16.03	3.87	79.63	70.10	12.28	2.45
CAML [4]	10%	90%	89.62	81.28	8.76	1.85	71.65	56.85	14.87	2.49	81.58	72.31	10.30	1.94
Co-BioNet [16]			89.20	80.68	6.44	1.90	77.89	64.79	8.81	**1.39**	75.22	65.32	13.56	1.94
BCP [1]			89.62	81.31	6.81	**1.76**	73.49	58.60	16.65	2.22	84.35	75.01	10.95	2.60
MCF [20]			87.67	78.42	9.16	2.79	65.39	51.43	13.85	2.39	79.28	69.08	12.43	3.58
Ours			**90.76**	**83.13**	**6.08**	1.87	**80.55**	**67.88**	**6.00**	2.50	**85.67**	**76.05**	**9.33**	2.76

3.3 Ablation study

Effect of the components: The ablation experiments are conducted on the LA dataset. We first conduct an ablation study to evaluate the effects of the MA module and the SE&RLE module. We use the pseudo labels generated by the Mean-teacher framework as the baseline. The results are detailed in Table 3. These results underscore that the integration of the MA module leads to a significant improvement. Moreover, integrating the SE&RLE module leads to further improvements. With 5% labeled data, the Dice score progresses from 88.61% to 89.76%. This demonstrates the effective use of pseudo labels by the SEM&RLE modules and underscores their crucial role in enhancing model accuracy.

Effect of the loss functions for distinct regions in $\mathcal{Y}$: In this paper, we used $\mathcal{L}_{con}$ and $\mathcal{L}_{NR}$ to evaluate losses within Ω and Θ, while excluding the evaluation of regions within Δ. To validate the rationale behind the RLE module, we conducted ablation studies on the loss functions for each region, as shown in
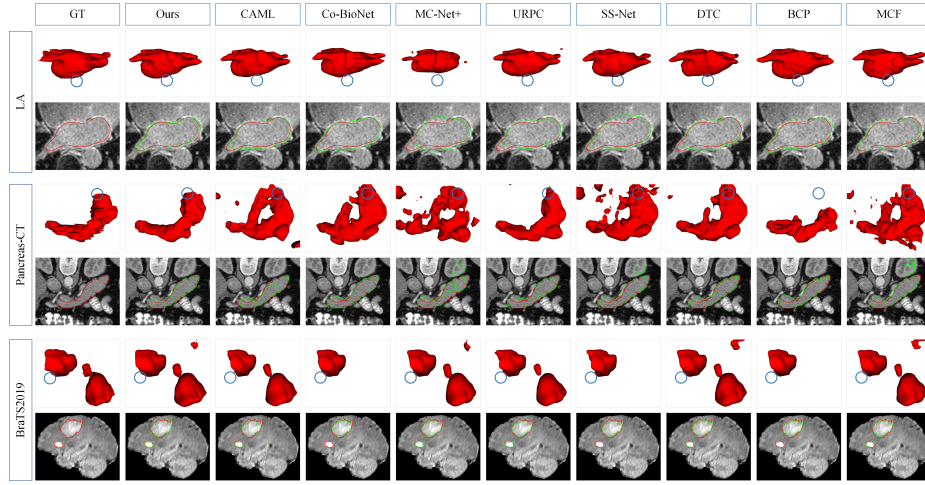
Fig. 2: Visualization of the segmentations results from different methods. The red lines denote the boundary of GT and the green lines denote the boundary of predictions.

Table 3: Effect of the components (upper) and effect of the loss function for different regions (lower).

Components			Scans used		Metrics			
Base	MA	SE&RLE	Lb	Unlb	Dice↑	Jac↑	HD↓	ASD↓
✓					87.45	77.95	8.95	**2.09**
✓	✓		4(5%)	76(95%)	88.61	79.65	11.09	2.84
✓	✓	✓			**89.72**	**81.42**	**8.46**	2.42
Regions			Scans used		Metrics			
Ω	Θ	Δ	Lb	Unlb	Dice↑	Jac↑	HD↓	ASD↓
$\mathcal{L}_{con}$	$\mathcal{L}_{con}$				89.04	80.32	7.45	1.97
$\mathcal{L}_{NR}$	$\mathcal{L}_{con}$				88.14	79.02	12.95	2.81
$\mathcal{L}_{NR}$	$\mathcal{L}_{NR}$		4(5%)	76(95%)	88.76	79.96	7.39	1.96
$\mathcal{L}_{NR}$	$\mathcal{L}_{NR}$	$\mathcal{L}_{NR}$			88.61	79.67	8.52	2.21
$\mathcal{L}_{con}$	$\mathcal{L}_{con}$	$\mathcal{L}_{con}$			88.61	79.65	11.09	2.84
$\mathcal{L}_{con}$	$\mathcal{L}_{NR}$				**89.70**	**81.40**	**6.68**	**1.75**

Table 3. The results indicate that excluding the Δ region improves the model's robustness. Traditional methods that used $\mathcal{L}_{con}$ for the entire $\mathcal{Y}$ achieved a Dice score of 88.61%. In contrast, our method achieved better performance with a Dice score of 89.70%, underscoring the effectiveness of the RLE module.

4   Conclusion

Semi-supervised learning has garnered widespread attention for its ability to utilize abundant unlabeled data to enhance robustness. Despite the popularity of pseudo labels strategy in semi-supervised learning, the effectiveness of existing methods can be diminished by noise contamination, which undermines model robustness. In response, we introduce SGRS-Net, which partitions the pseudo labels into distinct regions and evaluates the loss within each identified region using the proposed region-based loss evaluation module. Extensive experiments conducted on the LA, Pancreas-CT, and BraTS2019 dataset demonstrate superior performance over current state-of-the-art techniques, underscoring the efficiency and practicality of our framework.

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

References

1. Bai, Y., Chen, D., Li, Q., Shen, W., Wang, Y.: Bidirectional Copy-Paste for Semi-Supervised Medical Image Segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 11514–11524 (June 2023)
2. Chaitanya, K., Erdil, E., Karani, N., Konukoglu, E.: Local contrastive loss with pseudo-label based self-training for semi-supervised medical image segmentation. Medical Image Analysis **87**, 102792 (2023)
3. Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., Moore, S., Phillips, S., Maffitt, D., Pringle, M., et al.: The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository. Journal of Digital Imaging **26**, 1045–1057 (2013)
4. Gao, S., Zhang, Z., Ma, J., Li, Z., Zhang, S.: Correlation-aware mutual learning for semi-supervised medical image segmentation. In: Greenspan, H., Madabhushi, A., Mousavi, P., Salcudean, S., Duncan, J., Syeda-Mahmood, T., Taylor, R. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2023. pp. 98–108. Springer Nature Switzerland, Cham (2023)
5. Jia, X., Shen, Y., Yang, J., Song, R., Zhang, W., Meng, M.Q.H., Liao, J.C., Xing, L.: PolypMixNet: Enhancing semi-supervised polyp segmentation with polyp-aware augmentation. Computers in Biology and Medicine **170**, 108006 (2024)
6. Laine, S., Aila, T.: Temporal Ensembling for Semi-Supervised Learning. arXiv preprint arXiv:1610.02242 (2016)
7. Lee, D.H., et al.: Pseudo-Label: The Simple and Efficient Semi-Supervised Learning Method for Deep Neural Networks. In: Workshop on Challenges in Representation Learning, ICML. vol. 3, p. 896 (2013)

8. Lei, T., Zhang, D., Du, X., Wang, X., Wan, Y., Nandi, A.K.: Semi-Supervised Medical Image Segmentation Using Adversarial Consistency Learning and Dynamic Convolution Network. IEEE Transactions on Medical Imaging **42**(5), 1265–1277 (2023)

9. Li, X., Yu, L., Chen, H., Fu, C.W., Xing, L., Heng, P.A.: Transformation-Consistent Self-Ensembling Model for Semisupervised Medical Image Segmentation. IEEE Transactions on Neural Networks and Learning Systems **32**(2), 523–534 (2020)

10. Lu, S., Zhang, Z., Yan, Z., Wang, Y., Cheng, T., Zhou, R., Yang, G.: Mutually aided uncertainty incorporated dual consistency regularization with pseudo label for semi-supervised medical image segmentation. Neurocomputing **548**, 126411 (2023)

11. Luo, X., Chen, J., Song, T., Wang, G.: Semi-supervised Medical Image Segmentation through Dual-task Consistency. In: Proceedings of the AAAI conference on artificial intelligence. vol. 35, pp. 8801–8809 (2021)

12. Luo, X., Hu, M., Song, T., Wang, G., Zhang, S.: Semi-Supervised Medical Image Segmentation via Cross Teaching between CNN and Transformer. In: Konukoglu, E., Menze, B., Venkataraman, A., Baumgartner, C., Dou, Q., Albarqouni, S. (eds.) Proceedings of The 5th International Conference on Medical Imaging with Deep Learning. Proceedings of Machine Learning Research, vol. 172, pp. 820–833. PMLR (06–08 Jul 2022)

13. Luo, X., Wang, G., Liao, W., Chen, J., Song, T., Chen, Y., Zhang, S., Metaxas, D.N., Zhang, S.: Semi-supervised medical image segmentation via uncertainty rectified pyramid consistency. Medical Image Analysis **80**, 102517 (2022)

14. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al.: The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). IEEE Transactions on Medical Imaging **34**(10), 1993–2024 (2014)

15. Milletari, F., Navab, N., Ahmadi, S.A.: V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. In: 2016 Fourth International Conference on 3D Vision (3DV). pp. 565–571 (2016)

16. Peiris, H., Hayat, M., Chen, Z., Egan, G., Harandi, M.: Uncertainty-guided dual-views for semi-supervised volumetric medical image segmentation. Nature Machine Intelligence **5**(7), 724–738 (2023)

17. Ronneberger, O., Fischer, P., Brox, T.: 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. pp. 234–241. Springer (2015)

18. Tarvainen, A., Valpola, H.: Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (eds.) Advances in Neural Information Processing Systems. vol. 30. Curran Associates, Inc. (2017)

19. Valanarasu, Jeya Maria Jose and Patel, Vishal M.: UNeXt: MLP-Based Rapid Medical Image Segmentation Network. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2022. pp. 23–33. Springer Nature Switzerland, Cham (2022)

20. Wang, Y., Xiao, B., Bi, X., Li, W., Gao, X.: MCF: Mutual Correction Framework for Semi-Supervised Medical Image Segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 15651–15660 (2023)

21. Wu, Y., Ge, Z., Zhang, D., Xu, M., Zhang, L., Xia, Y., Cai, J.: Mutual consistency learning for semi-supervised medical image segmentation. Medical Image Analysis **81**, 102530 (2022)
22. Wu, Y., Wu, Z., Wu, Q., Ge, Z., Cai, J.: Exploring smoothness and class-separation for semi-supervised medical image segmentation. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2022. pp. 34–43. Springer Nature Switzerland, Cham (2022)
23. Xiong, Z., Xia, Q., Hu, Z., Huang, N., Bian, C., Zheng, Y., Vesal, S., Ravikumar, N., Maier, A., Yang, X., et al.: A global benchmark of algorithms for segmenting the left atrium from late gadolinium-enhanced cardiac magnetic resonance imaging. Medical Image Analysis **67**, 101832 (2021)
24. Yun, S., Han, D., Oh, S.J., Chun, S., Choe, J., Yoo, Y.: CutMix: Regularization Strategy to Train Strong Classifiers With Localizable Features. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (October 2019)
25. Zeng, X., Xiong, S., Xu, J., Du, G., Rong, Y.: Uncertainty Co-estimator for Improving Semi-Supervised Medical Image Segmentation. IEEE Transactions on Medical Imaging (2025)
26. Zhang, H., Cisse, M., Dauphin, Y.N., Lopez-Paz, D.: mixup: Beyond Empirical Risk Minimization. arXiv preprint arXiv:1710.09412 (2017)
27. Zheng, X., Fu, C., Xie, H., Chen, J., Wang, X., Sham, C.W.: Uncertainty-aware deep co-training for semi-supervised medical image segmentation. Computers in Biology and Medicine **149**, 106051 (2022)