

Semi-supervised Segmentation through Rival Networks Collaboration with Saliency Map in Diabetic Retinopathy

Eunjin Kim¹, Gitaek Kwon¹, Jaeyoung Kim², and Hyunjin Park^{3,4}

¹ VUNO Inc., S. Korea

² Teamreboott Inc., S. Korea

³ Dept of Electrical and Computer Engineering, Sungkyunkwan University, S. Korea,

⁴ Center for Neuroscience Imaging Research, Institute for Basic Science, S. Korea
hyunjinp@skku.edu

Abstract. Automatic segmentation of diabetic retinopathy (DR) lesions in retinal images has a translational impact. However, collecting pixel-level annotations for supervised learning is labor-intensive. Thus, semi-supervised learning (SSL) methods tapping into the abundance of unlabeled images have been widely accepted. Still, a blind application of SSL is problematic due to the confirmation bias stemming from unreliable pseudo masks and class imbalance. To address these concerns, we propose a **Rival Networks Collaboration with Saliency Map (RiCo)** for multi-lesion segmentation in retinal images for DR. From two competing networks, we declare a victor network based on Dice coefficient onto which the defeated network is aligned when exploiting unlabeled images. Recognizing that this competition might overlook small lesions, we equip rival networks with distinct weight systems for imbalanced and under-performing classes. The victor network dynamically guides the defeated network by complementing its weaknesses and mimicking the victor's strengths. This process fosters effective collaborative growth through meaningful knowledge exchange. Furthermore, we incorporate a saliency map, highlighting color-striking structures, into consistency loss to significantly enhance alignment in structural and critical areas for retinal images. This approach improves reliability and stability by minimizing the influence of unreliable areas of the pseudo mask. A comprehensive comparison with state-of-the-art SSL methods demonstrates our method's superior performance on two datasets (IDRiD and e-ophtha). Our code is available at https://github.com/eunjinkim97/SSL_DRlesion.

Keywords: Semi-supervised learning · Retinal image segmentation · Mutual learning.

1 Introduction

Diabetic Retinopathy (DR) often leads to vision loss. Thus, segmentation of DR-related lesions, such as microaneurysm (MA), hemorrhage (HE), hard exudate (EX), and soft exudate (SE), can assist ophthalmologists in efficient diagnosis

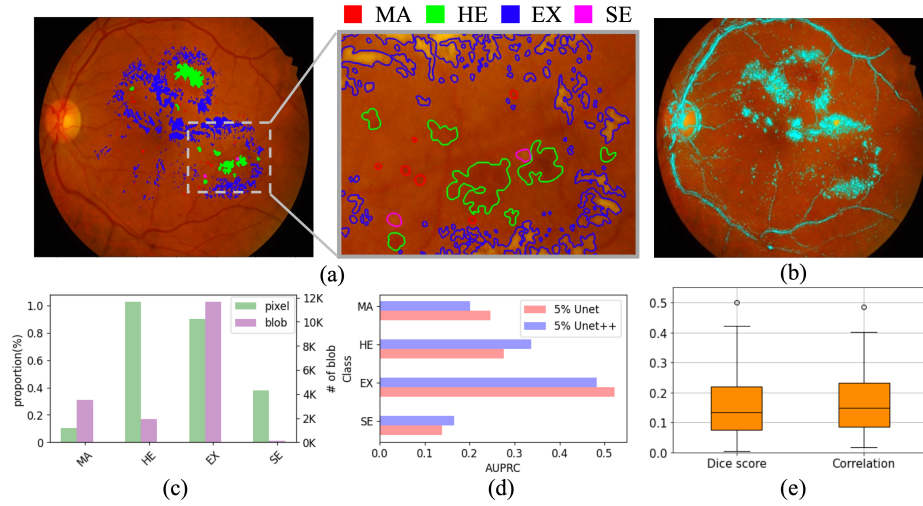


Fig. 1. Characteristics of a retinal image and DR-related lesions for four classes (MA, HE, EX, and SE) in red, green, blue, and pink, respectively. (b) Saliency map in skyblue on the retinal image. (c) Distribution of lesions in pixels and blobs. (d) Differences in performances between different architectures according to classes. (e) High Dice score and correlation between saliency maps and lesions.

or follow-up observations [15,21]. However, manual segmentation of heterogeneous scattered lesions is laborious, posing a challenge for automated methods that rely on large labeled datasets for training. To address this, semi-supervised learning (SSL) has been widely researched to utilize abundant unlabeled data effectively. Recently, many SSL approaches based on consistency regularization have garnered considerable interest, including adaptations of the Mean Teacher (MT) and the Cross Pseudo Supervision (CPS) [17,2].

The MT-based approach consists of a student model and teacher model updated by an exponential moving average from the student model, maintaining the consistency of the models' predictions from perturbed input images [3], especially in lower uncertainty regions [8,22]. However, these methods may underperform due to the risk of confirmation bias during training by accumulating errors from incorrect pseudo masks [1]. Besides that, many SSL studies, including the CPS, suggested co-training and model-level perturbations inducing invariant predictions [11,10]. While they can reduce confirmation bias with mutual learning, they still face obstacles with training bias towards the majority and easier classes [7,23].

Some SSL research on retinal images has suggested multi-task frameworks, investigating the relationship between the DR grade and lesions [18,24]. However, they did not fully address the heterogeneous distributions, such as imbalanced and unique morphological traits. Fig. 1c reveals general trends in pixel and blob count across classes, with HE lesions showing large areas but few blobs and MA

lesions presenting small areas with many blobs. Despite these patterns within each class, there are a variety of shapes and sizes as represented in Fig. 1a.

To confront the described issues, we propose Rival Networks Collaboration with Saliency Map, called **RiCo**, for effective multi-lesion semi-supervised segmentation. First, we select a victor network with a higher Dice score between two competing networks to obtain reliable pseudo masks. However, this may overlook small lesions due to the inherent bias of the Dice metric for larger lesions. Thus, we adopt a bifurcated rival network strategy: one for handling bias toward easy-to-learn classes (i.e., **difficulty-sensitive**) and another for class imbalance (i.e., **imbalance-sensitive**) with distinct network structures. This mechanism not only manages the biases but also captures diverse representations across various class distributions, as shown in Fig. 1d. Second, we introduce a victor-guided weighting strategy where the victor network directs the defeated network effectively to lessen the weakness of the defeated network when leveraging unlabeled images. If the victor network is the difficulty-sensitive one, we update the defeated network (i.e., **imbalance-sensitive**) weight with information derived from the difficulty-sensitive network. If the victor network is the imbalance-sensitive one, we update the defeated network (i.e., **difficulty-sensitive**) similarly. This approach prevents one-sided dominance and facilitates well-rounded mutual learning. Third, we employ a saliency map known to have high correlations with DR-lesions, as illustrated in Fig. 1b and e. We encourage the defeated network to match the victor’s pseudo masks in crucial areas highlighted by the saliency map to minimize errors in less reliable regions of the pseudo masks. It extends the concept of consistency for the low uncertainty area, exploring the lesion-related feature space in the pseudo masks. In conclusion, our method ensures balanced and superior performance across multi-classes by avoiding biases compared to leading SSL approaches on two datasets (IDRiD [13] and e-phtha [4]).

2 Preliminary

Recently, Wang et al. [19] suggested a dual-weight system for training difficulty-aware and distribution-aware networks. We briefly review them here since we build on their definition. The difficulty-aware weight is defined as $W_k^{diff} = w_{\gamma_{k,t}}(d_{k,t})^\alpha$. Here, $w_{\gamma_{k,t}}$ represents cumulative $1 - \gamma_{k,t}$, where $\gamma_{k,t}$ represents the Dice coefficient for k class at iteration t using the number accumulation iteration τ . The difficulty metric is defined as $d_{k,t} = \frac{du_{k,t} + \epsilon}{dl_{k,t} + \epsilon}$ with $du_{k,t} \approx \ln \frac{\gamma_{t,k}}{\gamma_{t-1,k}}$ for negative changes $\gamma_{t,k} - \gamma_{t-1,k} \leq 0$, $dl_{k,t} \approx \ln \frac{\gamma_{t,k}}{\gamma_{t-1,k}}$ for positive changes $\gamma_{t,k} - \gamma_{t-1,k} > 0$, and ϵ for smoothing factor. The distribution-aware weight is defined as $W_k^{dist} = \frac{\log(R_k)}{\max\{\log(R_i)\}_{i=1}^K}$, where $R_k = \frac{\max\{N_k\}_{i=1}^K}{N_k}$ and N_k represents the number of pixels representing k class. They are updated iteratively using an exponential moving average with a factor of β .

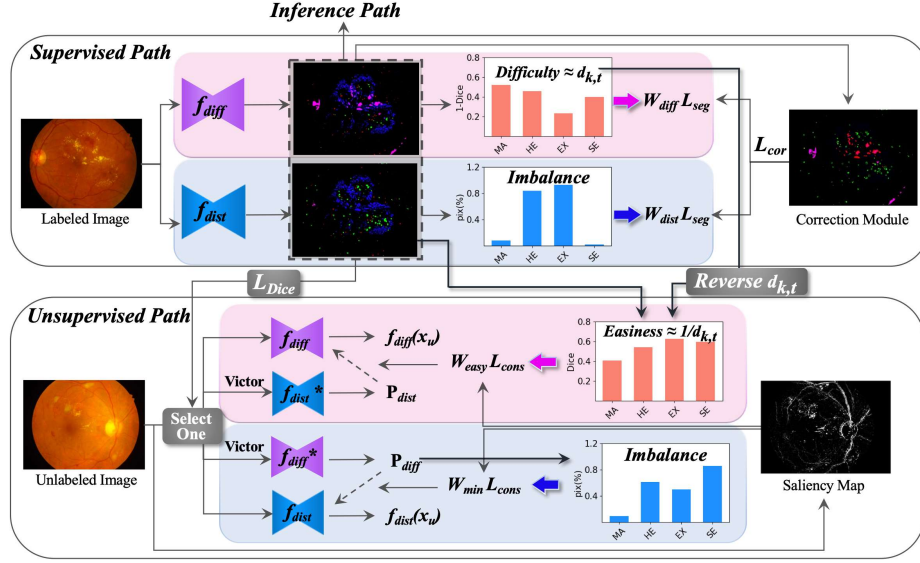


Fig. 2. Schematic of **RiCo**: Rival networks, f_{diff} and f_{dist} , possess distinct weights W_{diff} and W_{dist} for difficult and imbalanced with a correction module. The vector network determined by lower $\mathcal{L}_{Dice}$ generates P_V for $\mathcal{L}_{cons}$ with a saliency map for critical areas. If f_{dist} wins, f_{diff} is influenced by W_{easy} and if f_{diff} wins, f_{dist} is updated by W_{min} . An asterisk (*) as a superscript to f denotes a freezing network. The inference process is the mean of outputs of f_{diff} and f_{dist} .

3 Method

Let $\mathcal{B} = \mathcal{X} \cup \mathcal{U}$ be a batch of the training set, where $\mathcal{X} = \{(x_i, y_i) | y_i \in [0, 1]^{W \times H \times K}\}_{i=1}^{N_{\mathcal{X}}}$ is the $N_{\mathcal{X}}$ labeled set comprising pairs of images and masks for K classes and $\mathcal{U} = \{x_i\}_{i=1}^{N_{\mathcal{U}}}$ is the $N_{\mathcal{U}}$ unlabeled set with only images. For SSL tasks, the training loss is $\mathcal{L}_{total} = \mathcal{L}_{sup} + \lambda_{\mathcal{U}} \cdot \mathcal{L}_{\mathcal{U}}$, combining supervised loss $\mathcal{L}_{sup}$ and unsupervised loss $\mathcal{L}_{\mathcal{U}}$ weighted by $\lambda_{\mathcal{U}}$. Fig. 2 presents our strategy called **RiCo**.

3.1 Mutual Learning of Bifurcated Rival Networks

To explore the diverse representation of heterogeneous target distribution, we adopt a co-training framework from [20] with two networks of different structures, denoted as f_{diff} and f_{dist} : UNet [14] and UNet++ [25], respectively. They are independently forced to predict segmentation masks using $\mathcal{L}_{combo}$ which combines binary cross-entropy and Dice loss for $\mathcal{X}$ [16]. They compete to select a victor network based on the lower Dice loss $\mathcal{L}_{Dice}$, which then produces a pseudo mask P_V^k for $\mathcal{U}$. The defeated network learns to align with this P_V^k through consistency loss $\mathcal{L}_{cons}$.

This competition leads to size bias due to the inherent property of the Dice metric overlooking small and challenging classes. To tackle these issues of class imbalance and difficulty, we adopt the dual-weight system from [19], which [20] did not incorporate. $W_{diff} \in \mathcal{R}^K$ is assigned for the difficulty weight f_{diff} , diminishing weights to well-segmented classes while intensifying weights for hard classes. $W_{dist} \in \mathcal{R}^K$ for f_{dist} is iteratively tailored based on the class distribution of the predicted mask for $\mathcal{X}$. Thus, f_{diff} becomes the **difficulty-sensitive** network and f_{dist} is the **imbalance-sensitive** network. These rival networks with W_{diff}^k and W_{dist}^k are optimized using the following objective functions.

$$\mathcal{L}_{seg} = \frac{1}{N_{\mathcal{X}}K} \sum_{x_i \in \mathcal{X}} \sum_{k=1}^K [W_{diff}^k \mathcal{L}_{combo}(f_{diff}^k(x_i), y_i^k) + W_{dist}^k \mathcal{L}_{combo}(f_{dist}^k(x_i), y_i^k)]. \quad (1)$$

We implemented a correction module to utilize a discrepancy map D^k in [20] to improve mutual learning. We generate the pseudo masks from the predictions of two networks, P_{diff}^k and P_{dist}^k , by using a sharpening process with a temperature factor T . As follows in [20], each network is optimized with the correction loss $\mathcal{L}_{cor}$ using a discrepancy map D^k obtained by applying an exclusive OR operation on their pseudo masks in $\mathcal{X}$, as described below:

$$\mathcal{L}_{cor}^f = \frac{1}{N_{\mathcal{X}}K} \sum_{x_i \in \mathcal{X}} \sum_{k=1}^K \sum_{p,q} \frac{D_{p,q}^k \cdot [(f^k(x_{p,q}) - y_{p,q}^k)^2]}{\sum_{p,q} D_{p,q}^k}, \quad (2)$$

where indices p and q correspond to the x and y coordinates within the spatial dimensions. The supervised loss is defined as $\mathcal{L}_{sup} = \mathcal{L}_{seg} + \lambda_{cor} \mathcal{L}_{cor}$.

3.2 Victor-Guided Weighting Strategy for Consistency

To complement the weakness of the defeated network and leverage the strength of the victor network, we introduce a new victor-guided weight W_V^k , which can be either W_{easy}^k or W_{min}^k . Reversing the concept of the difficulty metric and weight W_{diff}^k , we propose an easiness metric $s_{k,t} = \frac{dl_{k,t} + \epsilon}{du_{k,t} + \epsilon}$ and a corresponding easiness weight $W_{easy}^k = v_{\psi_{k,t}}(s_{k,t})^\alpha$. Here, $v_{\psi_{k,t}}$ represents the cumulative $\psi_{k,t}$ at t^{th} iteration, where we define ψ as Dice score. $s_{k,t}$ finds out which classes the victor learns easily and $v_{\psi_{k,t}}$ informs which classes the victor has excelled in until t^{th} iteration. Therefore, W_{easy}^k spotlights the classes with higher Dice scores and a broader improvement range, encouraging the defeated network to resemble the victor's positive points by focusing on corresponding superior classes of P_V^k , which is the point [19] did not consider.

If f_{dist} (imbalance-sensitive) wins, the existing weights of f_{diff} (difficulty-sensitive) could be insufficient; thus, we revise the weights of f_{diff} to focus on easy classes specified by W_{easy}^k . That is, W_V^k is transformed into W_{easy}^k for the defeated one to be **easy-sensitive**, which complements the one-sided dominance of W_{diff}^k and evokes counterbalancing effects by leveraging the victor's strength.

If f_{diff} (difficulty-sensitive) wins, the existing weights of f_{dist} (imbalance-sensitive) could be insufficient; thus, we revise the weights of f_{dist} to focus on underrepresented classes predicted from f_{diff} . Here, W_V^k switches to W_{min}^k for the defeated one to be **minority-sensitive**, which emphasizes classes identified by f_{diff} as small classes. This step encourages f_{dist} to concentrate on reliable pseudo masks, enhancing W_{dist}^k in a similar stream. These updated weights are used in the unsupervised path through $\mathcal{L}_{cons}$. Our victor-guided dual-weight strategy for $\mathcal{U}$ constructs a cooperative competition to complement each other by reasonably leveraging unlabeled images. This strategy applies to the consistency loss $\mathcal{L}_{cons}$ and is described as follows:

$$W_V^k = \begin{cases} W_{easy}^k(\mathcal{L}_{Dice}(f_{dist}(x^l), y)), & \text{if } \bar{\mathcal{L}}_{Dice}(f_{diff}(x^l), y) \geq \bar{\mathcal{L}}_{Dice}(f_{dist}(x^l), y) \\ W_{min}^k(P_{diff}^u), & \text{if } \bar{\mathcal{L}}_{Dice}(f_{diff}(x^l), y) < \bar{\mathcal{L}}_{Dice}(f_{dist}(x^l), y) \end{cases} \quad (3)$$

, where $\bar{\mathcal{L}}_{Dice}$ denotes the average of $\mathcal{L}_{dice}$ over the classes $\mathcal{K}$.

3.3 Saliency Map-based Consistency Loss

We introduce a binarized saliency map $S(x^u)$ that captures color-striking features by employing a fine-grained static saliency method [12] and applying a threshold value th for binarization. $S(x^u)$ enables the defeated network f_D to pay attention to structural and eye-catching areas such as vessels, optic disc, and lesions when aligning with the victor’s pseudo mask P_V^k . By incorporating $S(x^u)$ into $\mathcal{L}_{cons}$, we boost consistency in crucial regions of the unlabeled images.

This approach is based on an insight that SSL methods with uncertainty ensure consistency in low-uncertainty areas derived from perturbed images [22]. It represents the significance of stable and reliable regions to avoid the wrong pseudo-mask regions. Expanding on this concept, we posit that $S(x^u)$ functions similarly to an uncertainty map, covering essential target areas and prior knowledge of retinal images. This procedure aims to minimize the cumulative confirmation bias and inherently explore the relationships among various lesion classes and structural features. $\mathcal{L}_{cons}$ becomes $\mathcal{L}_{\mathcal{U}}$ and is calculated as pixel-wise mean squared error using $S(x^u)$ as follows:

$$\mathcal{L}_{cons} = \frac{1}{N_{\mathcal{U}}K} \sum_{x_i \in \mathcal{X}_{\mathcal{U}}} \sum_{k=1}^K \sum_{p,q} \frac{S(x_{p,q}) \cdot [W_V^k(P_{V,p,q}^k - f_D^k(x_{p,q}))^2]}{\sum_{p,q} S(x_{p,q})}. \quad (4)$$

4 Experiments

Dataset and Implementation details Our proposed method, RiCo, is evaluated on two datasets, IDRiD [13] and e-optha [4] for multi-lesion segmentation. The IDRiD encompasses pixel-wise annotations for four types of DR lesions: MA, HE, EX, and SE. It is structured into a training set and a testing set, detailed

as follows: 54 and 27 pairs for MA and EX, 53 and 27 for HE, and 26 and 14 for SE, respectively. Note that the datasets have a class imbalance distribution, with less than 1% of lesions present in an image on average (see Fig. 1). We used the e-ophtha dataset comprising 47 images with Exudates, including EX and SE, and 148 images with MA. Following prior studies [5,6], we selected 21 images that exhibited both types of lesions (MA and Exudates) and split 11 for training and 10 for testing.

We implemented recent five SSL methods, UA-MT [22], CPS [2], CLD [9], DHC [19], and MCF [20] on our two datasets. We evaluated their performance across various labeled dataset configurations: 2%, 5%, 10%, and 20% for the IDRiD dataset and 10%, 20%, and 30% for the e-ophtha dataset. We empirically set λ_U as 0.1, λ_{cor} as 0.5, α at $\frac{1}{5}$, and the temperature factor T as 0.1 following [19,20]. To ensure fairness and showcase the efficacy of our approach, we conducted each experiment with three random seeds presenting the average and standard deviation of the results (i.e., average $\pm$ std). We assessed the performances using the area under the precision-recall curve (PRC) and the dice score coefficient (DSC). For details on training, please refer to the Supplementary material.

Table 1 shows the results from the IDRiD dataset to validate RiCo’s superior performance across multiple lesion classes with respect to four different

Table 1. Average DSC and PRC for four classes at 2%, 5%, 10%, and 20% labeled dataset proportions in the IDRiD dataset across three random seeds. **Bold** denotes the highest scores and underscores mark the second highest.

Prop.	2% (only 1 image)		5% (2 images)		10% (4 images)		20% (8 images)	
	Avg. DSC	Avg. PRC	Avg. DSC	Avg. PRC	Avg. DSC	Avg. PRC	Avg. DSC	Avg. PRC
Unet	24.89 $\pm$ 13.90	18.90 $\pm$ 11.36	37.05 $\pm$ 9.35	29.55 $\pm$ 8.83	50.60 $\pm$ 5.46	42.85 $\pm$ 6.92	54.05 $\pm$ 6.06	48.78 $\pm$ 6.67
Unet++	27.44 $\pm$ 15.89	20.36 $\pm$ 13.58	37.10 $\pm$ 11.57	29.65 $\pm$ 11.68	51.55 $\pm$ 3.69	44.02 $\pm$ 4.31	55.24 $\pm$ 2.31	49.94 $\pm$ 3.40
UA-MT	34.08 $\pm$ 10.40	24.64 $\pm$ 10.56	38.64 $\pm$ 13.00	32.57 $\pm$ 14.58	55.83 $\pm$ 5.00	50.29 $\pm$ 6.79	60.99 $\pm$ 3.56	58.34 $\pm$ 4.64
CPS	34.72 $\pm$ 9.50	26.90 $\pm$ 8.86	42.37 $\pm$ 10.05	35.07 $\pm$ 12.48	58.36 $\pm$ 4.60	54.94 $\pm$ 7.07	62.87 $\pm$ 1.54	61.97 $\pm$ 2.63
CLD	32.78 $\pm$ 8.94	25.33 $\pm$ 9.65	43.04 $\pm$ 9.28	35.28 $\pm$ 10.99	57.50 $\pm$ 5.06	54.48 $\pm$ 7.27	62.61 $\pm$ 2.26	61.99 $\pm$ 3.51
MCF	35.12 $\pm$ 10.44	28.04 $\pm$ 10.19	44.44 $\pm$ 8.75	38.72 $\pm$ 10.77	57.88 $\pm$ 3.49	55.54 $\pm$ 4.89	63.49 $\pm$ 3.23	63.02 $\pm$ 3.79
DHC	35.56 $\pm$ 12.6	27.82 $\pm$ 12.39	42.75 $\pm$ 9.8	36.01 $\pm$ 11.72	57.78 $\pm$ 4.92	54.37 $\pm$ 6.51	62.77 $\pm$ 2.13	61.38 $\pm$ 2.36
Ours	38.04$\pm$14.66	33.51$\pm$12.6	45.18$\pm$9.45	39.12$\pm$11.22	60.03$\pm$2.7	57.67$\pm$3.6	63.91$\pm$2.88	63.66$\pm$3.41

Table 2. Average DSC and PRC for MA and Exudates classes at 10%, 20%, and 30% labeled dataset proportions in the e-ophtha dataset across three random seeds.

Prop.	10% (only 1 image)		20% (2 images)		30% (3 images)	
	Avg. DSC	Avg. PRC	Avg. DSC	Avg. PRC	Avg. DSC	Avg. PRC
Unet	27.68 $\pm$ 0.38	19.05 $\pm$ 0.67	37.76 $\pm$ 7.36	28.53 $\pm$ 8.21	42.60 $\pm$ 7.15	34.89 $\pm$ 9.83
Unet++	27.38 $\pm$ 1.88	19.60 $\pm$ 3.59	39.49 $\pm$ 7.16	29.98 $\pm$ 8.44	43.42 $\pm$ 5.57	37.10 $\pm$ 8.02
UA-MT	36.58 $\pm$ 0.01	27.90 $\pm$ 0.97	43.82 $\pm$ 4.18	33.60 $\pm$ 4.69	45.67 $\pm$ 7.48	38.03 $\pm$ 10.62
CPS	35.35 $\pm$ 1.46	27.68 $\pm$ 0.02	43.29 $\pm$ 4.05	35.03 $\pm$ 5.35	45.94 $\pm$ 9.83	38.48 $\pm$ 13.42
CLD	32.72 $\pm$ 0.43	24.28 $\pm$ 0.56	40.41 $\pm$ 6.65	31.23 $\pm$ 8.32	40.95 $\pm$ 11.91	34.24 $\pm$ 14.66
MCF	39.59 $\pm$ 4.77	33.92 $\pm$ 6.44	44.36 $\pm$ 3.57	40.08 $\pm$ 6.38	46.64 $\pm$ 5.98	43.56 $\pm$ 9.83
DHC	35.74 $\pm$ 1.16	28.27 $\pm$ 2.79	44.24 $\pm$ 6.04	36.41 $\pm$ 7.18	44.78 $\pm$ 8.29	38.10 $\pm$ 10.92
Ours	40.08$\pm$4.22	34.66$\pm$6.15	44.74$\pm$4.68	40.48$\pm$8.24	47.99$\pm$6.14	44.75$\pm$10.62

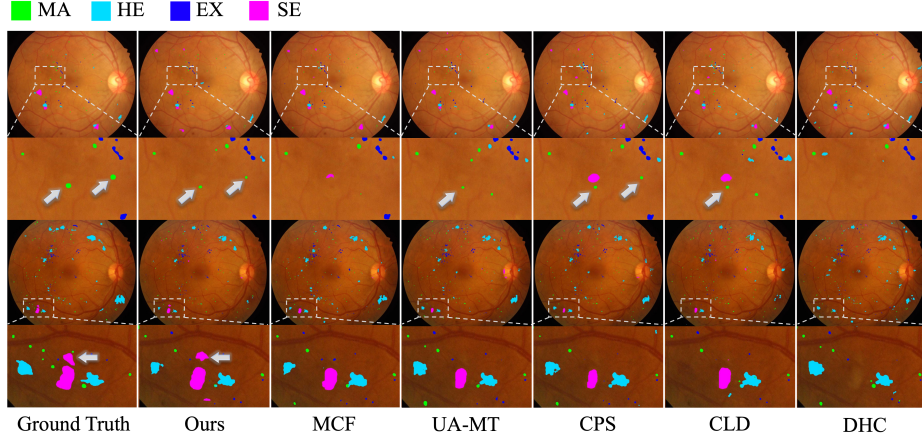


Fig. 3. Qualitative results at 5% labeled IDRiD dataset. The second-row zooms in on the first row’s images, and the fourth-row zooms in on the third row’s images. Gray arrows highlight subtle or challenging lesions.

proportions of labeled datasets. The average DSC and PRC of fully supervised learning are 62.93 ± 1.80 and 60.56 ± 2.7 for Unet and 62.93 ± 3.13 and 60.2 ± 4.5 for Unet++, respectively. Qualitative comparisons in Fig. 3 highlight RiCo’s proficiency in segmenting small and complex lesions. Table 2 represents the prominent results of ours in the e-optha dataset. The average DSC and PRC of fully supervised learning on e-optha are 50.03 ± 10.31 and 43.36 ± 13.2 for Unet and 51.78 ± 6.9 and 44.25 ± 9.95 for Unet++, respectively. Additional results per class are available in the Supplementary material.

To show the incremental impact of various components of our method, we performed an ablation study in Table 3. Here, the first row, the baseline method, represents the average results of predictions from two differently structured supervised networks. Using each component separately improved performance, but combining them resulted in complementary and balanced advances.

Table 3. Ablation Study on a 5% labeled dataset: “BR” for bifurcated rival networks with distinct weights, “VW” for the victor-guided weighting strategy, and “SC” for saliency-based consistency loss.

BR VW SC	Avg. DSC	Avg. PRC	MA	HE	EX	SE
	40.74 ± 8.61	33.79 ± 9.28	27.84 ± 6.94	35.71 ± 0.55	54.16 ± 8.61	17.46 ± 21.01
✓	44.56 ± 8.57	37.22 ± 10.79	31.99 ± 2.46	40.00 ± 0.13	52.94 ± 17.29	23.96 ± 23.29
✓	45.20 ± 9.11	38.14 ± 11.47	32.22 ± 2.93	42.14 ± 3.29	55.05 ± 16.92	23.13 ± 22.75
✓	44.20 ± 9.02	36.61 ± 11.87	30.74 ± 3.74	39.36 ± 1.89	53.32 ± 18.72	23.02 ± 23.12
✓	43.36 ± 7.60	38.47 ± 8.78	32.55 ± 0.63	47.47 ± 2.67	52.84 ± 9.66	21.02 ± 22.17
✓	45.18 ± 9.45	39.12 ± 11.22	32.49 ± 4.01	43.25 ± 2.81	55.41 ± 15.48	25.31 ± 22.56

5 Conclusion

We propose novel saliency map-enhanced rival networks with victor guidance to overcome confirmation and data biases simultaneously for multi-lesion segmentation with limited labeled data. We design a constructive competition where the defeated network learns the victor’s strengths, facilitating collaboration between two distinct rival networks.

Acknowledgments. This study was supported by National Research Foundation (NRF-2020M3E5D2A01084892), Institute for Basic Science (IBS-R015-D1), AI Graduate School Support Program (2019-0-00421), ICT Creative Consilience program (IITP-2024-2020-0-01821), and the Artificial Intelligence Innovation Hub program (RS-2021-II212068).

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

References

1. Arazo, E., Ortego, D., Albert, P., O’Connor, N.E., McGuinness, K.: Pseudo-labeling and confirmation bias in deep semi-supervised learning. In: 2020 International Joint Conference on Neural Networks (IJCNN). pp. 1–8. IEEE (2020)
2. Chen, X., Yuan, Y., Zeng, G., Wang, J.: Semi-supervised semantic segmentation with cross pseudo supervision. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2613–2622 (2021)
3. Cui, W., Liu, Y., Li, Y., Guo, M., Li, Y., Li, X., Wang, T., Zeng, X., Ye, C.: Semi-supervised brain lesion segmentation with an adapted mean teacher –model. In: Information Processing in Medical Imaging: 26th International Conference, IPMI 2019, Hong Kong, China, June 2–7, 2019, Proceedings 26. pp. 554–565. Springer (2019)
4. Decenciere, E., Cazuguel, G., Zhang, X., Thibault, G., Klein, J.C., Meyer, F., Marcotegui, B., Quellec, G., Lamard, M., Danno, R., et al.: Teleophtha: Machine learning and image processing methods for teleophthalmology. *Irbm* **34**(2), 196–203 (2013)
5. Guo, S., Li, T., Kang, H., Li, N., Zhang, Y., Wang, K.: L-seg: An end-to-end unified framework for multi-lesion segmentation of fundus images. *Neurocomputing* **349**, 52–63 (2019)
6. He, A., Wang, K., Li, T., Bo, W., Kang, H., Fu, H.: Progressive multiscale consistent network for multiclass fundus lesion segmentation. *IEEE transactions on medical imaging* **41**(11), 3146–3157 (2022)
7. He, R., Yang, J., Qi, X.: Re-distributing biased pseudo labels for semi-supervised semantic segmentation: A baseline investigation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 6930–6940 (2021)
8. Kendall, A., Gal, Y.: What uncertainties do we need in bayesian deep learning for computer vision? *Advances in neural information processing systems* **30** (2017)
9. Lin, Y., Yao, H., Li, Z., Zheng, G., Li, X.: Calibrating label distribution for class-imbalanced barely-supervised knee segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 109–118. Springer (2022)

10. Liu, J., Desrosiers, C., Zhou, Y.: 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. pp. 140–150. Springer (2022)
11. 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)
12. Montabone, S., Soto, A.: Human detection using a mobile platform and novel features derived from a visual saliency mechanism. *Image and Vision Computing* **28**(3), 391–402 (2010)
13. Porwal, P., Pachade, S., Kamble, R., Kokare, M., Deshmukh, G., Sahasrabudhe, V., Meriaudeau, F.: Indian diabetic retinopathy image dataset (idrid): a database for diabetic retinopathy screening research. *Data* **3**(3), 25 (2018)
14. 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)
15. Stitt, A.W., Curtis, T.M., Chen, M., Medina, R.J., McKay, G.J., Jenkins, A., Gardiner, T.A., Lyons, T.J., Hammes, H.P., Simo, R., et al.: The progress in understanding and treatment of diabetic retinopathy. *Progress in retinal and eye research* **51**, 156–186 (2016)
16. Taghanaki, S.A., Zheng, Y., Zhou, S.K., Georgescu, B., Sharma, P., Xu, D., Comaniciu, D., Hamarneh, G.: Combo loss: Handling input and output imbalance in multi-organ segmentation. *Computerized Medical Imaging and Graphics* **75**, 24–33 (2019)
17. Tarvainen, A., Valpola, H.: Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. *Advances in neural information processing systems* **30** (2017)
18. Ullah, Z., Usman, M., Latif, S., Khan, A., Gwak, J.: Ssm-d-unet: semi-supervised multi-task decoders network for diabetic retinopathy segmentation. *Scientific Reports* **13**(1), 9087 (2023)
19. Wang, H., Li, X.: DhC: Dual-debiased heterogeneous co-training framework for class-imbalanced semi-supervised medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 582–591. Springer (2023)
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. pp. 15651–15660 (2023)
21. Wu, H., Zhang, X., Geng, X., Dong, J., Zhou, G.: Computer aided quantification for retinal lesions in patients with moderate and severe non-proliferative diabetic retinopathy: a retrospective cohort study. *BMC ophthalmology* **14**, 1–5 (2014)
22. Yu, L., Wang, S., Li, X., Fu, C.W., Heng, P.A.: Uncertainty-aware self-ensembling model for semi-supervised 3d left atrium segmentation. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part II 22. pp. 605–613. Springer (2019)
23. Zhang, Y., Xiang, T., Hospedales, T.M., Lu, H.: Deep mutual learning. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 4320–4328 (2018)

24. Zhou, Y., He, X., Huang, L., Liu, L., Zhu, F., Cui, S., Shao, L.: Collaborative learning of semi-supervised segmentation and classification for medical images. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 2079–2088 (2019)
25. Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J.: Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. IEEE transactions on medical imaging **39**(6), 1856–1867 (2019)