2024 IEEE International Symposium on Biomedical Imaging (ISBI) | 979-8-3503-1333-8/24/$31.00 ©2024 IEEE | DOI: 10.1109/ISBI56570.2024.10635703

SEMI-ECNET: EDGE-CONSISTENCY BASED SEMI-SUPERVISED RETINAL VESSEL SEGMENTATION NETWORK

Yilun Qiu[1]*, Zhongxi Qiu*[1]*, Yan Hu*[1]*, Mingyang Bi*[1]*, Yubo Wang*[1]*, Jianwen Chen*[2]*, Yitian Zhao*[3]*, Heng Li*[1]*, Jiang Liu*[1]

1 Research Institute of Trustworthy Autonomous Systems and
Department of Computer Science and Engineering,
Southern University of Science and Technology, Shenzhen, Guangdong, China
2 Southern University of Science and Technology Hospital, Shenzhen, Guangdong, China
3 Institute of Biomedical Engineering, Ningbo Institute of Materials Technology and Engineering,
Chinese Academy of Sciences, Ningbo, Zhejiang, China

ABSTRACT

Blood vessel segmentation plays an important role in the diagnosis and treatment of retinal diseases. The performance of supervised deep-learning-based segmentation methods is dependent on the training labels, which brings a great burden to surgeons. Semi-supervised methods can solve the problem partly, but recently proposed algorithms hardly consider the complexity of the tree structures in retinal images, especially fine peripheral bronchi. Thus, we propose a novel edge-consistency based semi-supervised retinal vessel segmentation algorithm, named Semi-ECNet. Specifically, Semi-ECNet first generates two kinds of vessel maps, including an edge constraint map and a pixel-wise probability map in the model-prediction stage. Then for the loss-consistency stage, we adopt the Sobel operator and propose a novel loss strategy for the consistency constraints among these maps and the ground truth. Extensive experiments on a publicly available dataset demonstrate that our Semi-ECNet effectively leverages unlabeled data, and outperforms other state-of-the-art semi-supervised segmentation methods by introducing this innovative edge-consistency strategy.

***Index Terms*—** Blood vessel segmentation, Retinal image, Semi-supervised, Edge-Consistency

1. INTRODUCTION

Blood vessel segmentation is fundamental in the analysis of medical retinal fundus images [1, 2]. This task extracts valuable blood vessel information from medical retinal fundus images, offering insights into the overall systemic health state, which plays a crucial role in the diagnosis and treatment of ophthalmic diseases, such as diabetic retinopathy and macular degeneration [3]. Recently medical image segmentation witnessed remarkable advancements with the introduction of deep learning techniques employing supervised models with impressive performance. However, the effectiveness of supervised methods heavily depends on a substantial number of labeled fundus images, which is labor-intensive and demands extensive professional expertise, particularly in detecting tiny blood vessels.

A semi-supervised learning framework has emerged as a promising approach by leveraging a limited number of labeled data alongside a substantial amount of unlabeled data during training. There are two primary types of semi-supervised methods: pseudo-labeling based and consistency regularization based. Pseudo-labeling based methods involve generating reliable pseudo-labels for unlabeled data and incorporating them during the retraining phase. Lee et al. [4] initially investigated semi-supervised learning with pseudo-labeling, employing a fixed threshold to select high-confidence unlabeled samples for labeling. Subsequent studies by Cascante et al. [5] had further enhanced the quality of pseudo-labels through innovative improvements. Consistency regularization based methods aim to enhance generalization by enforcing prediction consistency in the presence of various perturbations. Mean Teacher (MT) [6] is a representative strategy that involves a student network updating parameters through gradient propagation and a teacher network updating parameters based on exponential moving average (EMA). Uncertainty Aware Mean Teacher (UA-MT) [7] utilizes uncertainty information to guide the student network in gradually learning from the meaningful and reliable targets provided by the teacher network. Both MT and UA-MT are data-level regularization methods that utilize the distribution or structural information of unlabeled data to constrain the learning process of the model. In contrast, Dual Task Consistency (DTC) [8] is a dual-task-consistency framework by explicitly incorporating task-level regularization. Task-level regularization focuses more on task-specific features

Yilun Qiu and Zhongxi Qiu are the co-first authors. Yan Hu is the corresponding author (huy3@sustech.edu.cn).

for model learning by utilizing the characteristics of the task itself or domain knowledge. This method enables the model to adapt more effectively to the specific requirements of the task and reduce overfitting to irrelevant features, rendering it a superior strategy compared to data-level regularization.

The effectiveness of the task-level regularization is proven in DTC where the consistency of the level set prediction and semantic segmentation is introduced to make the model learn from the unlabel data. However, the level set function is insufficient for the retinal vessel segmentation task due to the special structure of the vessel. We argue that the edge information is crucial for the vessel. Thus, a novel edge-consistency based semi-supervised retinal vessel segmentation network (Semi-ECNet) is introduced in this work. We enforce edge consistency by employing the edge consistency loss explicitly designed for unlabeled data in the loss-consistency stage. This loss is combined with the edge loss and the segmentation loss, collectively guiding the model learning process.

The contributions of our paper include:

1. We propose a novel semi-supervised blood vessel segmentation network based on an edge consistency strategy, named Semi-ECNet, which boosts the potential of edge information in retinal vessels.
2. We introduce a suitable edge extraction task and the edge consistency loss for the training of unlabeled data.
3. Extensive experiments based on a publicly available dataset prove that Semi-ECNet achieves comparable performance to fully supervised methods and other SOTA semi-supervised methods.

2. METHOD

Given a dataset $D = D^L \cup D^U$ where $D^L = (X^L, Y^L)$ containing labeled input images and labels with N^L samples and $D^U = (X^U)$ only containing unlabeled images with N^U samples, $N^U \gg N^L$. Our objective is to fully utilize unlabeled images to improve the segmentation accuracy of blood vessels based on our Semi-ECNet. The overall architecture of the model is illustrated in Fig.1. As we adopt an unlabeled dataset in Semi-ECNet, we divide the network into the model-prediction stage and the loss-consistency stage.

In the model-prediction stage, we try to provide two edge-related predictions. We first apply color-jitters techniques to the input medical retinal fundus images during the training process to enhance data diversity. This preprocessing step allows us to capture nuanced edge information, facilitating a more precise analysis of the retinal features. Then, these results are forwarded to the backbone network. The backbone network, commonly based on an encoder-decoder structure such as VNet [9], taking processed images as input, consists of two output channels: the edge constraint layer to generate the predicted edge constraint map with vascular edge information, and the pixel-wise prediction layer to generate the predicted pixel-wise probability map for pixel-level segmentation.

For the same input, the predictions of blood vessels are supposed to be the same, so we proposed the loss-consistency stage. In this stage, we design three distinct loss strategies to constrain the semi-supervised network, including the segmentation loss, the edge loss, and the edge consistency loss. The segmentation loss measures the difference between the predicted pixel-wise probability map and the ground truth labels. The edge loss quantifies the disparity between the predicted edge constraint map and the ground truth edge map. The edge consistency loss utilizes the output of unlabeled data by enforcing edge consistency between the predicted edge constraint map and the converted edge map from the predicted pixel-wise probability map. Labeled data is used to compute the segmentation loss, the edge loss, and the edge consistency loss. Unlabeled data, on the other hand, is only utilized in enhancing edge consistency by computing the edge consistency loss.

To convert the ground truth and predicted pixel-wise probability map to their corresponding edge maps, we introduce a novel Sobel operator [10] here to extract edge information from images to be processed:

$$\begin{cases} \hat{Y}_s = \Theta(Y) \\ \hat{Y}_s^L = \Theta(Y^L) \end{cases}$$

where $\hat{Y} = \hat{Y}^L \cup \hat{Y}^U$ denotes the predicted pixel-wise probability map obtained from all input images, Y^L represents the ground truth for labeled data, $\hat{Y}_s$ is the converted edge map, and $\hat{Y}_s^L$ is the ground truth edge map. Θ is the Sobel operator performing the edge extraction task.

We employ the widely used Dice loss [9] for the segmentation loss:

$$\mathcal{L}_s = Dice(Y^L, \hat{Y}^L) = 1 - \frac{2 \times \sum(Y^L \cdot \hat{Y}^L)}{\sum Y^L + \sum \hat{Y}^L}$$

where $Dice$ refers to the Dice loss. This loss serves as a guiding factor for the model, facilitating the optimization of predictions for labeled samples.

To enhance the performance of the edge constraint layer in producing a high-quality edge constraint map, we employ the Mean Square Error loss to calculate the edge loss.

$$\mathcal{L}_e = MSE(\hat{Y}_e^L, Y_s^L) = \frac{1}{N^L} \sum (\hat{Y}_e^L - Y_s^L)^2$$

where MSE represents the Mean Square Error loss and Y_e^L is the model output of labeled data obtained from the edge constraint layer.

To boost the potential of large-scale unlabeled data, we use the Mean Square Error loss to obtain the edge consistency loss between the predicted edge constraint map and the converted edge map:

$$\mathcal{L}_{ec} = MSE(\hat{Y}_e, \hat{Y}_s) = \frac{1}{N} \sum (\hat{Y}_e - \hat{Y}_s)^2$$

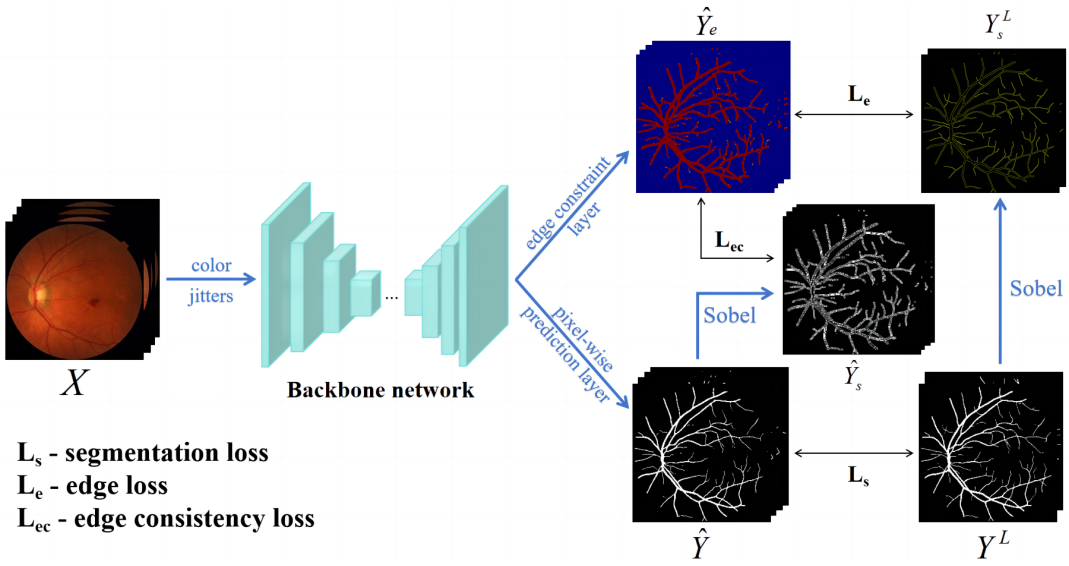

Fig. 1. Architecture of Semi-ECNet, including the model-prediction stage and the loss-consistency stage

where $N = N^L + N^U$, both $\hat{Y}_e$ and $\hat{Y}_s$ contain labeled and unlabeled data. This loss, denoted as $\mathcal{L}_{ec}$, facilitates the alignment between labeled and unlabeled data, promoting a coherent and consistent representation of edges, capturing both labeled and unlabeled data samples.

The overall loss can be expressed as follows:

$$\mathcal{L}_{total} = \mathcal{L}_s + \alpha\mathcal{L}_e + \beta\mathcal{L}_{ec}$$

where α and β define the weights of the edge loss and the edge consistency loss to balance these two losses, respectively. In our experiments, we empirically set the values of α and β to 0.5 each.

During the testing phase, only the pixel-wise probability map in the output of the backbone network is retained as the intermediate result. This result undergoes binary classification with a threshold of 0.5, and the final segmentation result is obtained accordingly.

3. EXPERIMENTS

3.1. Dataset and Data Pre-Processing

To evaluate our proposed model, we use the Fundus Image Dataset for Artificial Intelligence based Vessel Segmentation (FIVES) [11] as the dataset. This dataset contains 800 high-resolution multi-disease color fundus photographs with pixel-wise manual annotation, which is the largest retinal vessel segmentation dataset to date. We utilize 600 images for training, allocate 100 images for validating, and reserve another 100 images for testing. For all experiments, we randomly partition the training dataset into labeled and unlabeled data subsets. The best validation result is selected for testing purposes.

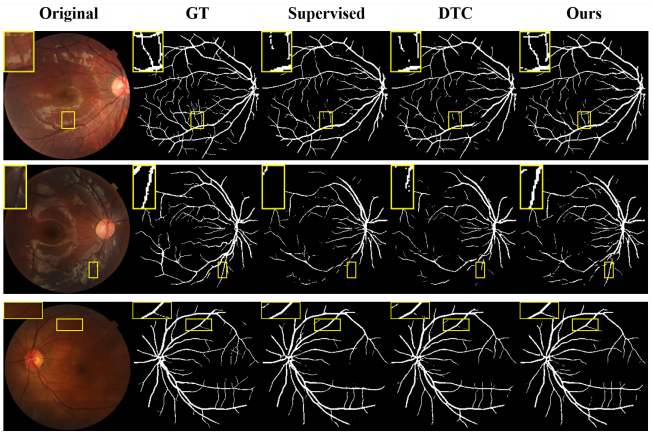

Fig. 2. Comparison of visual results on the FIVES dataset

We employ a zoom-in strategy to resize the images to an input size of 256 × 256 for the model. To further enhance the model's performance, we utilize the standard on-the-fly data augmentation methods [7] and incorporate color-jitters techniques. These techniques involve introducing random variations in brightness, hue, contrast, and saturation of input images during the training phase.

3.2. Implementation Details and Evaluation Metrics

Our model is implemented using the PyTorch library [12], and all experiments are conducted on a machine equipped with an NVIDIA Tesla A100 graphics card. For the model-prediction stage, we employ VNet as the backbone network in all experiments, as it exhibits outstanding performance on the current segmentation task. To generate the predicted constraint edge

Table 1. The comparison of different methods on the FIVES dataset in DSC, IoU, MCC, and BM

Label %	Methods	DSC	IoU	MCC	BM
100%	Supervised	0.7620	0.6235	0.7472	0.7424
	Ours	0.7640	0.6270	0.7502	0.7484
50%	Supervised	0.7535	0.6142	0.7398	0.7237
	Ours	0.7599	0.6214	0.7458	0.7555
10%	Supervised	0.7260	0.5790	0.7121	0.6790
	UA-MT	0.7456	0.6039	0.7308	0.7167
	DTC	0.7488	0.6074	0.7340	0.7224
	MC-Net-v1	0.7426	0.6007	0.7280	0.7219
	MC-Net-v2	0.7408	0.5988	0.7258	0.7337
	BCP	0.7376	0.5943	0.7261	0.6858
	MCF	0.7490	0.6076	0.7345	0.7210
	Ours	**0.7508**	**0.6088**	**0.7353**	**0.7398**
1%	Supervised	0.6473	0.4944	0.6315	0.6242
	Ours	0.7067	0.5574	0.6892	0.7124

map, we augment the original VNet network architecture by incorporating an additional regression layer. This layer comprises a convolution operation followed by a sigmoid function. We train the framework using the Adam optimizer for 6000 iterations. The initial learning rate (lr) is set to 0.01 and is decayed by a factor of 0.1 every 2500 iterations. In this study, a batch size of 24 is employed, comprising an even distribution of 12 labeled and 12 unlabeled images whenever the number of input images allows for such allocation. In cases where the number of input images is insufficient to maintain this balance, the batch size is adjusted to match the total number of input images.

The metrics to quantitatively evaluate the performance of our proposed semi-supervised model are the Dice Score Coefficient (DSC), the Intersection-over-Union (IoU), the Matthews Correlation Coefficient (MCC), and Bookmaker Informedness (BM).

3.3. Comparison Experiments

We conduct a comprehensive sequence of comparison experiments in semi-supervised learning to assess the feasibility of achieving our objective, including UA-MT [7], DTC [8], MC-Net-v1 [13], MC-Net-v2 [14], BCP [15], and MCF [16]. All these methods are evaluated based on the official parameter configurations. Additionally, we establish a baseline of a fully supervised approach. The comparison results are listed in Table 1 and the visualization results in Fig.2. As shown in Table 1, we test with different labeled data ratio settings, including 100%, 50%, 10%, and 1%. Our approach demonstrates a substantial enhancement in vessel segmentation results when compared to the fully supervised learning strategy. Notably, when only 10% of the data is labeled, our method outperforms all other state-of-the-art (SOTA) methods and the supervised framework across all four evaluation metrics. Thus, the superiority of our Semi-ECNet is proven based on the comparison results. Fig.2 visualizes the segmentation results of different methods of three cases using only 10% labeled data. We mark some results with yellow rectangles. The continuity of blood vessels segmented by our algorithm is better than that of the others. Moreover, our Semi-ECNet method exhibits enhancements in accurately segmenting fine vessels and vessels with low clarity, outperforming both the supervised learning method and DTC. These results indicate that our approach is more suited than six other state-of-the-art methods at accomplishing retinal fundus image vessel segmentation tasks.

Table 2. Ablation study of different task-level consistency

dual-task-consistency	color-jitters	edge-consistency	DSC
✓	×	×	0.7488
✓	✓	×	0.7478
×	×	✓	0.7493
×	✓	✓	**0.7508**

3.4. Ablation Studies

We conduct ablation experiments to study the impact of our proposed improvements built upon the DTC framework. The corresponding quantitative results are presented in Table 2, including dual-task-consistency, color-jitters, and edge-consistency strategies. Initially, we utilize the dual-task-consistency approach proposed within DTC without any modifications. Subsequently, we attempt to introduce color-jitters techniques to the input images. Nonetheless, this strategy proves to be ineffective within the DTC framework. Next, we evaluate our proposed edge-consistency strategy, which yields superior performance compared to the aforementioned methods. Finally, the fully implemented Semi-ECNet achieves additional optimization, improving the DSC to a value of 0.7508.

4. CONCLUSION

In this paper, we propose a novel semi-supervised blood vessel segmentation network, named Semi-ECNet, which is a task-level regularization based structure that applies edge consistency to utilize the importance of edge information. The edge consistency loss proposed in Semi-ECNet leverages both labeled and unlabeled data by promoting a coherent and consistent representation of edges. This strategy leads to a significant enhancement in the model's ability to segment fine vessels and vessels with low clarity. Extensive experiments demonstrate that Semi-ECNet outperforms the supervised learning method and other SOTA methods.

5. ACKNOWLEDGMENT

This work was supported in part by Guangdong College Students' Innovation and Entrepreneurship Training Program (2023S17), Guangdong Basic and Applied Basic Research Foundation(2022A1515010487), Shenzhen Stable Support Plan Program (20220815111736001) and Shenzhen Natural Science Fund (JCYJ20210324103800001, JCYJ20220530112609022).

6. COMPLIANCE WITH ETHICAL STANDARDS

This research study was conduced retrospectively using human subject data made available in open access. Ethical approval was not required as confirmed by the license attached with the open access data.

7. REFERENCES

[1] Zhongxi Qiu, Yan Hu, Xiaoshan Chen, Dan Zeng, Qingyong Hu, and Jiang Liu, "Rethinking dual-stream super-resolution semantic learning in medical image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–14, 2023.

[2] Tim Laibacher, Tillman Weyde, and Sepehr Jalali, "M2u-net: Effective and efficient retinal vessel segmentation for real-world applications," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 2019, pp. 0–0.

[3] P Saranya, S Prabakaran, Rahul Kumar, and Eshani Das, "Blood vessel segmentation in retinal fundus images for proliferative diabetic retinopathy screening using deep learning," *The visual computer*, pp. 1–16, 2022.

[4] Dong-Hyun Lee et al., "Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks," in *Workshop on challenges in representation learning, ICML*. Atlanta, 2013, vol. 3, p. 896.

[5] Paola Cascante-Bonilla, Fuwen Tan, Yanjun Qi, and Vicente Ordonez, "Curriculum labeling: Revisiting pseudo-labeling for semi-supervised learning," in *Proceedings of the AAAI conference on artificial intelligence*, 2021, vol. 35, pp. 6912–6920.

[6] Antti Tarvainen and Harri Valpola, "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results," *Advances in neural information processing systems*, vol. 30, 2017.

[7] Lequan Yu, Shujun Wang, Xiaomeng Li, Chi-Wing Fu, and Pheng-Ann Heng, "Uncertainty-aware self-ensembling model for semi-supervised 3d left atrium segmentation," in *MICCAI*, 2019.

[8] Xiangde Luo, Jieneng Chen, Tao Song, and Guotai Wang, "Semi-supervised medical image segmentation through dual-task consistency," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021, vol. 35, pp. 8801–8809.

[9] Fausto Milletari, Nassir Navab, and Seyed-Ahmad 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.

[10] O Rebecca Vincent, Olusegun Folorunso, et al., "A descriptive algorithm for sobel image edge detection," in *Proceedings of informing science & IT education conference (InSITE)*, 2009, vol. 40, pp. 97–107.

[11] Kai Jin, Xingru Huang, Jingxing Zhou, Yunxiang Li, Yan Yan, Yibao Sun, Qianni Zhang, Yaqi Wang, and Juan Ye, "Fives: A fundus image dataset for artificial intelligence based vessel segmentation," *Scientific Data*, vol. 9, no. 1, pp. 475, 2022.

[12] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al., "Pytorch: An imperative style, high-performance deep learning library," *Advances in neural information processing systems*, vol. 32, 2019.

[13] Yicheng Wu, Minfeng Xu, Zongyuan Ge, Jianfei Cai, and Lei Zhang, "Semi-supervised left atrium segmentation with mutual consistency training," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, Cham, 2021, vol. 12902, pp. 297–306.

[14] Yicheng Wu, Zongyuan Ge, Donghao Zhang, Minfeng Xu, Lei Zhang, Yong Xia, and Jianfei Cai, "Mutual consistency learning for semi-supervised medical image segmentation," *Medical Image Analysis*, vol. 81, pp. 102530, 2022.

[15] Yunhao Bai, Duowen Chen, Qingli Li, Wei Shen, and Yan Wang, "Bidirectional copy-paste for semi-supervised medical image segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 11514–11524.

[16] Yongchao Wang, Bin Xiao, Xiuli Bi, Weisheng Li, and Xinbo 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. 15651–15660.