2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)
April 13-16, 2021, Nice, France

2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI) | 978-1-6654-1246-9/20/$31.00 ©2021 IEEE | DOI: 10.1109/ISBI48211.2021.9433917

CYCLE ADAPTIVE MULTI-TARGET WEIGHTING NETWORK FOR AUTOMATED DIABETIC RETINOPATHY SEGMENTATION

Lianyu Wang[1], Zhongyue Chen[1], Meng Wang[1], Tingting Wang[1], Weifang Zhu[1], Xinjian Chen[1,2]

[1]School of Electronic and Information Engineering, Soochow University, Suzhou, China
[2]State Key Laboratory of Radiation Medicine and Protection, Soochow University, Suzhou, China

ABSTRACT

Diabetic retinopathy (DR) is one of the most common microvascular complications of diabetes. Early and accurate screening of DR from fundus images is crucial for the ophthalmologist to make treatment plans. In recent years, many deep learning-based methods have been proposed for medical image segmentation. However, the DR lesions segmentation still meets great challenges.

In this work, we propose a novel cycle adaptive multi-target weighting network (CAMWNet) that mimics the biological vision system of the human brain. The network consists of two major parts: a novel adaptive multi-target weighting network (AMWNet) for DR lesions segmentation and a reverse data recovery network (RRN) to simulate the cycle perception in visual hierarchy. In addition, a novel joint loss function is designed to optimize the CAMWNet. Comprehensive experiments on Indian Diabetic Retinopathy Image Dataset (IDRiD) show that, CAMWNet achieves better performance than other state-of-the-art methods with accuracy, Dice similarity coefficients, sensitivity and specificity of 98.33%, 53.84%, 48.54% and 99.88%, respectively.

Index Terms—Diabetic retinopathy, neural network, retinal fundus images, multi-target segmentation.

1. INTRODUCTION

Diabetic retinopathy (DR) is a chronic eye disease depicted by the presence (see Fig. 1) of one or more retinal lesions like microaneurysms (MA), hemorrhages (HE), hard exudates (EX) and soft exudates (SE) [1]. DR is one of the main causes of blindness. However, more than 90% of blindness can be prevented by early diagnosis and treatment. Indeed, manual examinations take a long time and there are not enough experts to meet the growing demand for screening. Therefore, the development of the computer-aided automatic diagnostic system is crucial to the prevention and treatment of DR.

At present, DR lesions segmentation is faced with the problem of scarce data and inter-class difference. Playout et al. [2] proposed a supervised learning approach that training multi-task network to enhance the generalization ability of deeply segmentation networks. When the lesion regions are scattered and small-scaled, Yan et al. [3] integrating the decoder of a global-level UNet and a patch-level one to achieve better segmentation. Xue et al. [4] proposed a multi-task segmentation method based on a hybrid dynamic membrane system. After then, Sambyal et al. [5] proposed a modified UNet to

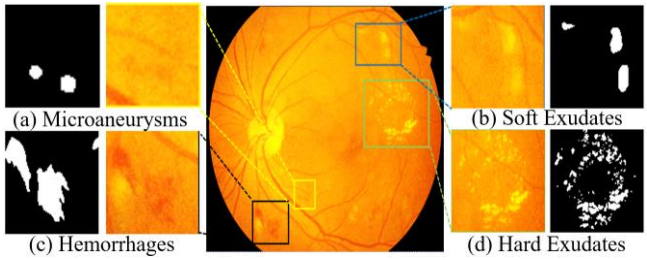

Fig. 1. Fundus photograph containing different retinal lesions associated with DR. (a) Microaneurysms (MA), (b) Soft Exudates (SE), (c) Hemorrhages (HE) and (d) Hard Exudates (EX).

precisely define the boundary of the region of interest and improve the segmentation performance of MA and EX. In addition, Guo et al. [6] proposed a top-k loss algorithm based on class-imbalance and loss-imbalance to alleviate misclassification. Although these methods have achieved promising results in the segmentation of certain types of lesions, it is still a challenge to segment MA, HE, EX, SE and optic disc (OD) simultaneously due to diverse shape, blurred boundary, unclear pathological characteristics of lesions.

Previous works have shown that task-optimized deep convolutional neural networks (CNN) are precise quantitative models of visual coding in primate visual cortex [7][8]. However, some structures of primate visual systems are not completely imitated by feed-forward CNN. By merging generative recurrent feedback, Huang et al. [9] enhance the consistency of the neural network. Their works show better adversarial robustness than the conventional feed forward CNN. Inspired by their works, we propose a novel cycle adaptive multi-target weighting network (CAMWNet). It mainly composed of a segmentation network and a reverse data recovery network (RRN). First, to alleviate the influence of background information in the fundus image, and to solve the problem of class-imbalance caused by the difference between retinopathy, a novel adaptive multi-target weighting network (AMWNet) is proposed as our segmentation network. To further improve the multi-target segmentation performance and the generality of the network, a U-shape network is designed and employed as our RRN to simulate feedback from the high-level visual hierarchy to the low-level. In addition, we develop a novel joint loss function to optimize CAMWNet. Finally, we conduct comprehensive experiments on Indian Diabetic Retinopathy Image Dataset (IDRiD) to evaluate the effectiveness of our proposed method. The experimental results show that, compared with other state-of-the-art methods, our proposed method has achieved outstanding performance in DR lesions segmentation task.

Authorized licensed use limited to: Soochow University. Downloaded on January 05,2022 at 03:03:26 UTC from IEEE Xplore. Restrictions apply.

2. METHOD

As shown in Fig. 2, the proposed CAMWNet consists of two major parts: an AMWNet for DR lesions segmentation and an RRN that aims to simulate feedback from the high-level visual hierarchy to the low-level. The fundus images are fed into AMWNet to predict the segmentation results and generate the target feature maps for recovering the data. Then, the target feature maps are fed into the RRN for restoration. Finally, a novel joint loss function is designed to optimize the network parameters end-to-end.

2.1. Adaptive Multi-Target Weighting Network (AMWNet)

It has been demonstrated that CNN can automatically learn rich feature information from the input images. However, due to the large inter-class differences in the DR lesions segmentation task, the class information of some targets will be weakened when decoded, which leads to the segmentation performance degradation. To solve these problems, we propose AMWNet for DR lesions segmentation. In this framework, weights are assigned adaptively to different targets by learning high-level global semantic information, so that the class imbalance problem can be solved. As shown in Fig. 2, AMWNet consists of three parts: encoder-path, decoder-path and our newly proposed adaptive multi-target weighting module (AMW). The encoder-path is used to extract rich semantic information and global features in the input image, and to down-sample the feature maps at different stages. The decoder-path aims to up-sample the feature maps with strong semantic information but weak resolution from higher-level stage. The features are fused with weak semantic but high-resolution information through skip-connection [10]. It has been demonstrated that the top features in the highest stage of CNN contain the strongest semantic information that is beneficial for classification. Therefore, to tackle the problem of class imbalance, AMW module is attached to the center stage to learn the semantic information and adaptively generate weights for different targets. The AMW module is constructed by three convolutional layers and two fully connection layers. In the AMW, the feature maps from top stage of encoder path $\mathbf{X}_H$ will be encoded to a weight vector $\mathbf{w} \in \mathbb{R}^{N\times 1}$ corresponding to different targets as follows:

$$\mathbf{w} = g_{AMW}(\mathbf{X}_H) \tag{1}$$

where $g_{AMW}(\cdot)$ represents AMW module, N denotes the targets numbers (except background). Then, the weight vector is multiplied by the feature maps $\mathbf{F} \in \mathbb{R}^{(N+1)\times H\times W}$ from the final layer of decoder path to obtain the final prediction result:

$$\mathbf{F}_c^w = \mathbf{F}_c \times w_c \quad c = 1,2,\dots,N \tag{2}$$

$$\mathbf{Y}'(x,y) = arg\max_c \mathbf{F}_c^w(x,y) \quad c = 0,1,\dots,N \tag{3}$$

where $\mathbf{F}_c$ and w_c are the feature map and weights correspond to the c-th target, respectively. $\mathbf{F}_c^w$ represents the weighted feature map corresponds to c-th target and (x,y) represents the spatial coordinates in the feature maps. $\mathbf{Y}'$ is the final prediction result.

2.2. Reverse Data Recovery Network (RRN)

It has been proved in biology that human brain cognition is a cyclic perception process, which can associate the target domain with the source domain of the object, and vice versa. These findings have also been introduced into the design of computer vision algorithms, and achieved promising results [7][8]. Inspired by their works, a perception mechanism is introduced into our model and applied to the task of DR lesions segmentation, as shown in Fig. 2.

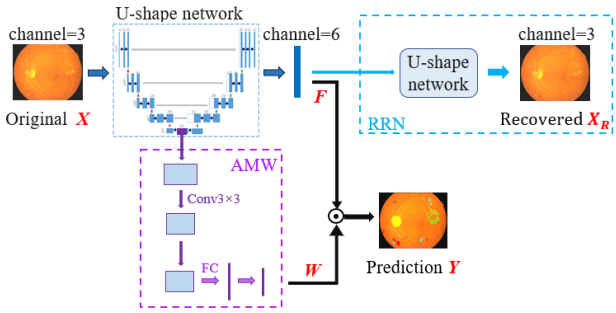

Fig. 2. An overview of the proposed cycle adaptive multi-target weighting network (CAMWNet).

The cycle perception system is constructed by introducing an RRN that aims to simulate feedback from the high-level visual hierarchy to the low-level. U-shape network is widely used in segmentation and data generation works [11] as its unique architecture. Therefore, we design a shallow U-shape network as our RRN for data recovery based on the features extracted by AMWNet:

$$\mathbf{X}_R = RRN(\mathbf{F}) \tag{4}$$

where $\mathbf{X}_R$ is the recovered data obtained by RRN and $\mathbf{F}$ is feature maps from the top of decoder-path. RRN is a part of the entire network and is jointly trained with AMWNet. With RRN, the semantic comprehension is further improved, which helps to improve the multi-task segmentation performance of DR lesions.

2.3. Loss Function

To further optimize the proposed model, we develop a novel joint loss for the proposed CAMWNet, as follows:

$$Loss_{total} = \alpha Loss_{AMW} + \beta Loss_{RRN} \tag{5}$$

where α and β are hyper-parameters. In this paper, α and β are set as 0.6 and 0.1, respectively. $Loss_{AMW}$ is mainly used to balance the inter-class difference and to improve segmentation performance, while $Loss_{RRN}$ aims to improve the segmentation robustness by cyclic sensing mechanism. They are defined as follows:

$$Loss_{AMW} = L_{BCE}(\mathbf{Y}', \mathbf{Y}) + L_{Dice}(\mathbf{Y}', \mathbf{Y}) \tag{6}$$

$$Loss_{RRN} = L_{MSE}(\mathbf{X}_R, \mathbf{X}) \tag{7}$$

where $\mathbf{Y}'$ and $\mathbf{Y}$ are the segmentation results and ground truth, respectively. $\mathbf{X}_R$ and $\mathbf{X}$ represents the recovered data obtained by RRN and the corresponding original fundus images, respectively.

3. EXPERIMENTAL RESULTS

3.1 Data description and implementation details

The dataset used in our experiments is the Indian Diabetic Retinopathy Image Dataset (IDRiD) [1], which is available for the segmentation and grading of the retinal image challenge hosted by ISBI (International Symposium on Biomedical Imaging) conference 2018. IDRiD contains 81 images with a resolution of 4288 × 2848. Pixel-level annotations of EX, HE, MA, OD and SE are provided. Unlike the IDRiD challenge, we performed joint segmentation on the lesion above. We performed 4-fold cross validation both in

Authorized licensed use limited to: Soochow University. Downloaded on January 05,2022 at 03:03:26 UTC from IEEE Xplore. Restrictions apply.

Table 1. DSC (%) of different networks on validation set. Bold numbers indicate the best performance.

Method	Mean DSC	HE DSC	EX DSC	MA DSC	SE DSC
UNet	50.46±1.80	51.53±3.17	71.21±3.26	**39.62±2.42**	39.50±1.63
CPFNet	34.44±4.37	37.08±4.50	61.22±4.50	8.45±5.72	31.01±9.60
CE-Net	47.65±3.04	51.06±4.00	62.85±3.22	30.30±1.90	46.39±7.18
AttUNet	42.11±4.49	49.97±4.93	70.38±1.83	32.22±18.71	15.86±16.36
CAMWNet	**53.84±2.36**	**52.73±5.44**	**71.41±2.48**	38.13±3.29	**53.11±3.08**

Table 2. More prediction results (%) of different networks on validation set. Bold numbers indicate the best performance.

Method	Mean DSC	ACC	SEN	SEP	JSC	PC	PCC
UNet	50.46±1.80	98.19±0.27	47.68±3.53	99.85±0.03	35.71±1.92	60.19±0.86	52.16±1.76
CPFNet	34.44±4.37	98.11±0.36	29.35±3.66	**99.92±0.02**	24.52±2.70	57.21±0.81	39.27±3.10
CE-Net	47.65±3.04	98.23±0.38	43.40±1.18	99.86±0.04	32.49±2.31	57.52±4.86	48.77±2.74
AttUNet	42.11±4.49	98.09±0.40	**50.75±3.57**	99.82±0.06	38.26±2.18	62.46±3.16	**54.90±1.94**
CAMWNet	**53.84±2.36**	**98.33±0.33**	48.54±3.94	99.88±0.02	**38.31±2.14**	**64.93±1.27**	54.80±2.11

ablation and contrast experiments. Except for the last batch with 21 images, the other batches contained 20 images that were not repeated. To ensure that the receptive field is sufficient and to reduce the computation, the images are resized to 512 × 512. This input size will not lose much information compared with 256 × 256.

The implementation of our proposed CAMWNet is based on the public platform Pytorch and NVIDIA Tesla K40 GPU with 12GB memory. The Adam optimizer with a learning rate of 0.0005 is adopted to optimize our model, and the batch size is set to 4. In addition, random vertical flipping was applied for data augmentation online. To be fair, all experiments adopt the same data preprocessing and training strategy. To verify the effectiveness of our proposed method, seven indicators, including Dice similarity coefficients (DSC), accuracy (ACC), sensitivity (SEN), specificity (SPE), Jaccard similarity coefficient (JSC), precision (PC), and Pearson's correlation coefficient (PCC) are used to evaluate the performance.

During the inference section, it takes only about 350ms and 0.89GB GPU memory to output the prediction for a 512×512 image, which indicates that it is easy to deploy the CAMWNet.

3.2 Results

We compared the proposed CAMWNet with other state-of-the-art methods, including UNet [10], CPFNet [12], CE-Net [13] and AttUNet [14], as shown in Table 1 and 2. As a non-lesion area, OD is excluded when evaluating the performance. The numerical results show the superiority of our proposed CAMWNet. The Mean DSC of CAMWNet is 53.84%, and is 3.38% higher than the baseline. Although some of AttUNet's performance is comparable to that of CAMWNet, it do badly in the segmentation of SE.

In addition, it can be seen from Table 1 that compared with other methods, the DSC of SE obtained by CAMWNet has been significantly improved. As relatively fewer samples, the problem of class imbalance is aggravated in the lesion SE, and our proposed method can still well meet this challenge.

To further demonstrate the effectiveness of the proposed method, the qualitative segmentation results are also given in Fig. 3. It can be seen that the CAMWNet is more accurate and has better robustness in the DR lesions segmentation task.

3.3. Ablation Study

To verify the validity of the proposed AMWNet and RRN, we also conduct ablation experiments. For convenience, we call the U-shape network as the baseline method. As shown in Table 3, AMWNet achieves substantial improvement over the baseline in terms of four targets, and results 53.43% Mean DSC. This outperforms the baseline with only a small computational increase. The addition of RRN also helps to improve the performance. Compared with the baseline, the Mean DSC increases 1.83%. Especially, our proposed AMW and RRN can be easily introduced into other encoder-decoder network, which is our near future work. Furthermore, the proposed CAMWNet achieves the highest Mean DSC, whose results further demonstrate the effectiveness of our proposed method.

4. CONCLUSION

In this work, we propose a novel cycle adaptive multi-target weighting network (CAMWNet) that mimics the biological vision system of the human brain. To tackle the problem of class imbalance and to further improve the segmentation performance, an adaptive multi-target weighting network is developed for DR lesions segmentation. Based on CAMWNet network architecture, a new joint loss function is proposed to optimize the network. The experiment results show that the proposed method achieves better performance than other state-of-the-art methods.

5. COMPLIANCE WITH ETHICAL STANDARDS

This research study was conducted retrospectively using human subject data made available in open access by https://ieee-dataport.org/open-access/indian-diabetic-retinopathy-image-dataset-idrid#files. Ethical approval was not required as confirmed by the license attached with the open access data.

6. ACKNOWLEDGEMENT

This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFA0701700, in part by the National Nature Science Foundation of China under Grant 61622114, and in part by the National Basic Research Program of China under Grant 2014CB748600.

Authorized licensed use limited to: Soochow University. Downloaded on January 05,2022 at 03:03:26 UTC from IEEE Xplore. Restrictions apply.

Table 3. Ablation study of multi-target segmentation on the IDRID.

Method	Mean DSC	HE DSC	EX DSC	MA DSC	SE DSC
baseline	50.46±1.80	51.53±3.17	71.21±3.26	**39.62±2.42**	39.50±1.63
AMWNet	53.43±1.42	**52.97±2.97**	**71.44±2.43**	39.17±1.61	50.14±2.16
Baseline + RRN	52.29±1.97	51.74±3.02	71.25±2.31	39.02±2.12	47.15±2.88
CAMWNet	**53.84±2.36**	52.73±5.44	71.41±2.48	38.13±3.29	**53.11±3.08**

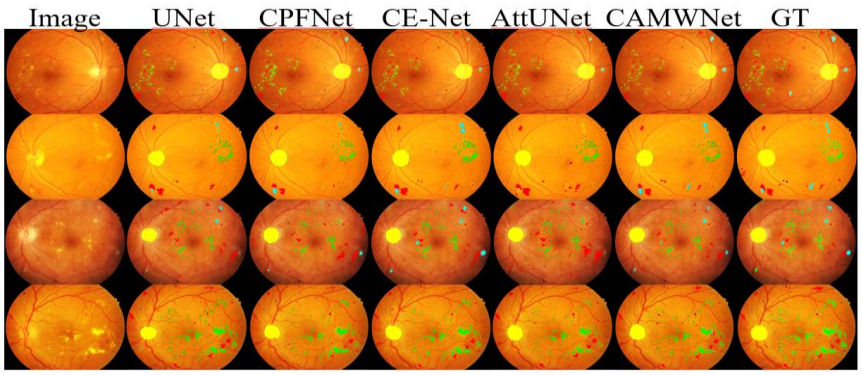

Fig. 3. Examples of segmentation. The red, green, blue, yellow, sky-blue regions represent hemorrhages (HE), hard exudates (EX), microaneurysms (MA), optic disc (OD) and soft exudates (SE) respectively.

7. REFERENCES

[1] P. Porwal, S. Pachade, R. Kamble, M. Kokare, G. Deshmukh, V. Sahasrabuddhe, and F. Meriaudeau, "Indian Diabetic Retinopathy Image Dataset (IDRiD): A Database for Diabetic Retinopathy Screening Research," *Data*, vol. 3, no. 3, 2018.

[2] C. Playout, R. Duval and F. Cheriet, "A Novel Weakly Supervised Multitask Architecture for Retinal Lesions Segmentation on Fundus Images," *IEEE Transactions on Medical Imaging*, vol. 38, no. 10, pp. 2434-2444, 2019.

[3] Z. Yan, X. Han, C. Wang, Y. Qiu, and S. Cui, "Learning Mutually Local-Global U-Nets for High-Resolution Retinal Lesion Segmentation in Fundus Images," *International Symposium on Biomedical Imaging*, pp. 597-600, 2019.

[4] J. Xue, S. Yan, J. Qu, F. Qi, C. Qiu, H. Zhang, M. Chen, T. Liu, D. Li, and X. Liu, "Deep Membrane Systems for Multitask Segmentation in Diabetic Retinopathy," *Knowledge-Based Systems*, vol. 183, Nov. 2019.

[5] N. Sambyal, P Saini, R Syal, and V Gupta, "Modified U-Net Architecture for Semantic Segmentation of Diabetic Retinopathy," *Biocybernetics and Biomedical Engineering*, vol. 40, no. 3, pp. 1094-1109, 2020.

[6] S. Guo, K. Wang, H. Kang, T. Liu, Y. Gao, and T. Li, "Bin loss for hard exudates segmentation in fundus images," *Neurocomputing*, vol. 392, pp. 314-324, Jun. 2020.

[7] U. Güçlü and M. A. J. v. Gerven, "Deep Neural Networks Reveal A Gradient in the Complexity of Neural Representations across the Ventral Stream," *The Journal of Neuroscience*, vol. 35, no. 27, pp. 10005–10014, 2004.

[8] A. Nayebi, D. Bear, J. Kubilius, K. Kar, S. Ganguli, D. Sussillo, and D. L. K. Yamins, "Task-driven convolutional recurrent models of the visual system,". *Advances in Neural Information Processing Systems 31*, pp. 5290–5301, 2018.

[9] Y. Huang, J. Gornet, S. Dai, Z. Yu, T. Nguyen, D. Y. Tsao, A. Anandkumar, "Neural Networks with Recurrent Generative Feedback," *arXiv preprint arXiv*:2007.09200, 2020.

[10] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *Medical Image Computing and Computer-Assisted Intervention*, vol. 9351, pp. 234–241, 2015.

[11] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," *Advances in Neural Information Processing Systems 27*, pp. 2672–2680, 2014.

[12] S. Feng, H. Zhao, F. Shi, X. Cheng, M. Wang, Y. Ma, D. Xiang, W. Zhu, and X. Chen, "CPFNet: Context Pyramid Fusion Network for Medical Image Segmentation," *IEEE Transactions on Medical Imaging*, vol. 39, no. 10, pp. 3008-3018, Oct. 2020

[13] Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, Y. Zhao, T. Zhang, s. Gao and J. Liu, "CE-Net: Context Encoder Network for 2D Medical Image Segmentation," *IEEE Transactions on Medical Imaging*, vol. 38, no. 10, pp. 2281-2292, Oct. 2019.

[14] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B.Glocker, D. Rueckert, "Attention U-Net: Learning Where to Look for the Pancreas," arXiv preprint arXiv:1804.03999, 2018.

Authorized licensed use limited to: Soochow University. Downloaded on January 05,2022 at 03:03:26 UTC from IEEE Xplore. Restrictions apply.