A Simplified Deep Network Architecture on Optic Cup and Disc Segmentation

Guan-Ru Huang and Tien-Ruey Hsiang*
Department of Computer Science and Information Engineering
National Taiwan University of Science and Technology
Taipei, Taiwan
Email: m10615022@csie.ntust.edu.tw, *trhsiang@csie.ntust.edu.tw

***Abstract*—Glaucoma is caused by damaged optic nerves, and can lead to permanent vision loss. The cup-to-disk ratio (CDR) is a key criterion for glaucoma diagnosis, therefore an accurate automatic segmentation of the optic disc (OD) and optic cup (OC) in retinal fundus images has become a major research topic. However, with deeper deep learning models being used to complete a fundus segmentation, the segmentation results remain acceptable only on certain datasets. In this research, a lightweight deep-learning encoder–decoder architecture, which adopts polar coordinate transformation and histogram equalization to simplify the learning complexity, is proposed for the simultaneous segmentation of OD and OC. The proposed model employs an encoder–decoder architecture consisting of a feature extraction module and a multi-output module to both simplify the model and to adjust the result through different outputs and multilabel loss functions. Experiment results demonstrate that the proposed approach outperforms other deep network models on OD and OC segmentation over the datasets REFUGE, MESSIDOR, and RIM-ONE, which contain images captured from cameras with various specifications. In addition, better results on glaucoma screening through CDR calculation were obtained on the REFUGE dataset with the proposed method.**

***Index Terms*—Simplification, deep learning, optic disc segmentation, optic cup segmentation, glaucoma screening.**

I. Introduction

Studies have indicated that the number of glaucoma patients worldwide will increase to 111.8 million by 2040, and it is therefore important to guide the screening and treatment of glaucoma and the design of related public health strategies [1]. Because the initial symptoms of glaucoma are not obvious, patients often do not experience pain or vision loss. Glaucoma may be noticed only when the disease progresses to a significant loss of peripheral vision or complete blindness. Therefore, glaucoma is one of the main causes of irreversible blindness. Early detection and timely treatment are keys to preventing patients from vision loss, and effective diagnostic methods should therefore be developed. In terms of diagnosis, the cup-to-disk ratio (CDR) is usually used to determine the presence of glaucoma. Fig. 1 shows a retinal fundus image. The large circle in the middle is called the optic disc (OD). The small circle inside the OD is called the optic cup (OC). The CDR is ratio of the OC area to the OD area. In general, the CDR of a normal person is approximately 0.3 and does not exceed 0.6, whereas the CDR of a glaucoma patient exceeds 0.6. This also means that the higher the CDR is, the more severe the impairment in the optic nerve. An image segmentation task can be applied to the OC and OD in a fundus image to obtain the CDR and help assess the risk of glaucoma.

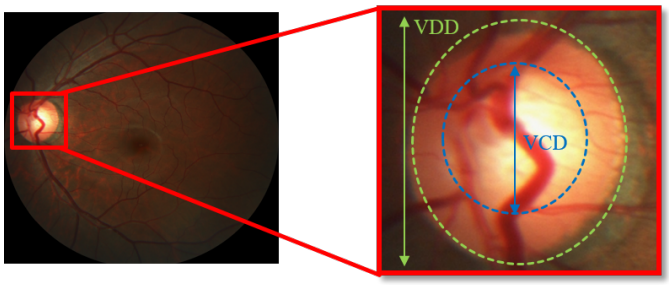


Fig. 1. Structure of a retinal fundus image: green ellipse, optic disc (OD); blue circle, optic cup (OC); and vertical CDR, ratio of vertical cup distance (VCD) to vertical disk distance (VDD).

Many studies have applied deep learning to fundus image segmentation. However, a single dataset or a dataset with too few images is used, making it difficult to verify the generalizability to different datasets, which is a major concern. The spatial resolution of retinal fundus images is usually quite high, and the input size of the model directly affects the model size. Therefore, many methods reduce the image size. Simplifying the deep network model becomes an important key to protecting a resized image from a distortion affecting the CDR determination. However, an underfitting easily occurs if the number of layers in the deep network is arbitrarily reduced to simplify the network. Therefore, improving the structure of each layer or block in a deep network can reduce the use of parameters and simplify the network. For example, MobileNet [2] and ShuffleNet t [3] apply object detection on mobile devices, and ESPNet [4] constructs an ESP module to achieve a lightweight segmentation network. A lightweight feature extraction module can thus be constructed to maintain the original size of the input image and avoid improper proportions.

In the past, manual feature extraction was mostly used for a fundus image diagnosis, but its insufficient discrimination made the results subject to pathological areas and low contrast. However, with the rapid development of deep learning, many computer vision applications achieve a good resolution. For example, convolutional neural networks (CNNs) perform

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

better in image classification (VGG16) [5] and segmentation (FCN)) [6]. For retinal images, Gulshan et al. demonstrated that deep learning systems have a high sensitivity and specificity for detecting inducible diabetic retinopathy [7]. Good results have also been achieved for fundus vascular segmentation [8]. Using retinal fundus images, a method has been proposed that reuses the features and reduces the required number of parameters to improve the segmentation, particularly the effect of OC segmentation [9].

Our goal is to achieve a higher generalizability in terms of glaucoma determination using ophthalmoscopes based on the premise of constructing a lightweight deep model. In this study, we propose a preprocessing method and a lightweight deep learning network. The main contributions are as follows:

1) For the preprocessing, we applied a polar coordinate transformation and histogram equalization to enhance the image features which effectively reduces the learning complexity.
2) For the network architecture, we use the encoder-decoder architecture of DeepLabV3+, a linear bottleneck of MobileNetV2, and an inverted residual block to reduce the parameters and achieve a light weight. A multi-output approach is employed to strengthen the OC and OD segmentation.
3) The output module used in the network separately segments the OD and OC, combining them for a clearer correlation for a multilabel segmentation. Through a preprocessing assessment, we tested the generalizability of the preprocessing on the REFUGE [10], MESSIDOR, and RIM-ONE [11] datasets. The results show that our method, U-Net, and M-Net exhibit significant improvements in the segmentation of OD and OC.
4) In the segmentation of the OD and OC, the CDR was calculated and applied to the REFUGE challenge dataset. Our method achieved a better effect with an AUC value of 0.94. In addition, the occupied parameters reached 6.71 MB, and the floating point operations (FLOPs) performed per second reached approximately 1.4 MB, thereby achieving quickness and a light weight under the effective segmentation levels.

The remainder of this paper is organized as follows: Section II reviews past fundus OD and OC segmentation techniques. Details of our method and its components are described in Section III. To test the effectiveness of our method, we planned and conducted experiments in two parts, as presented in Section IV. Some concluding remarks are finally given in Section V.

II. Related Studies

Prior studies have described numerous methods for segmenting OD and OC in retinal fundus images. The existing methods of automatic OD and OC segmentation can be roughly divided into four categories: shape-based and template matching, active contour and deformable based models, machine learning, and deep learning.

Shape-based and template matching models were the earliest to be applied to OC and OD segmentation. Such methods regard the OD as a circle [12] or ellipse [13], and use a Hough transformation or ellipse fitting to approximate the OD shape. However, the segmentation lacks stability owing to changes in image color, nonuniform intensity, prominent parts such as exudates present in the abnormality, and the influence of surrounding blood vessels inside the OD.

An active contour method (ACM) is widely applied to OC and OD segmentation. This method can convert the segmentation problem into an energy minimization problem, in which different energies are derived to reflect features such as the intensity, texture, and boundary smoothness of an image. [14] pioneered the use of ACM in the detection of contours based on image gradients . Since then, numerous researchers have applied different variants of ACM to OD and OC segmentation. For example, [15], proposed a gradient vector flow-based ACM to segment the OD.

Machine learning mainly relies on various artificial or image processing methods for feature extraction. The extracted features are trained as representatives of the image. A special preprocessing is usually required to assist in the feature display. Such methods are usually effective for a single dataset, and cannot be applied to various datasets. In [16], pixels and local neighborhood features extracted from the ROIs of the retinal fundus image are used to classify the OC and retina for each pixel through a support vector machine (SVM). In [17], an unsupervised method is used to segment the OC in the fundus image. Superpixel classification is regarded as a low-rank representation (LRR) problem, and the best parameters are adaptively selected to generate a graph of the final result.

Convolutional neural network (CNN) is an active research topic with wide uses in computer vision applications. Through CNNs, more complex features can be learned and extracted from the input images. The initial method focused on a single region for segmentation. For example, a full VGG16 CNN is used for OD segmentation [18], CNNs with large pixel blocks are applied [19], with classification and postprocessing used for each pixel block to achieve OC segmentation. However, with the development of segmentation technology, the use of an improved U-Net to segment the OD and OC was introduced [20], although OD and OC are segmented separately in order. In addition, a deep ensemble network for glaucoma detection was proposed [21]. Various network modules can be used to learn different levels of fundus images, such as the entire fundus image and the OD region. A segmentation-guided network is used to segment the OD to enhance the effect of glaucoma detection. In [22], a training architecture of M-Net based on polar coordinate transformation and multilabel concept is proposed. A multilabel concept was used to consider the simultaneous segmentation of OD and OC. A polar coordinate transformation was employed to process fundus images from another perspective. This method is also presently the best segmentation method available. Sun et al. [23] argued that a deep object detection network can simultaneously frame the boundary box and locate the OD and OC, and used a

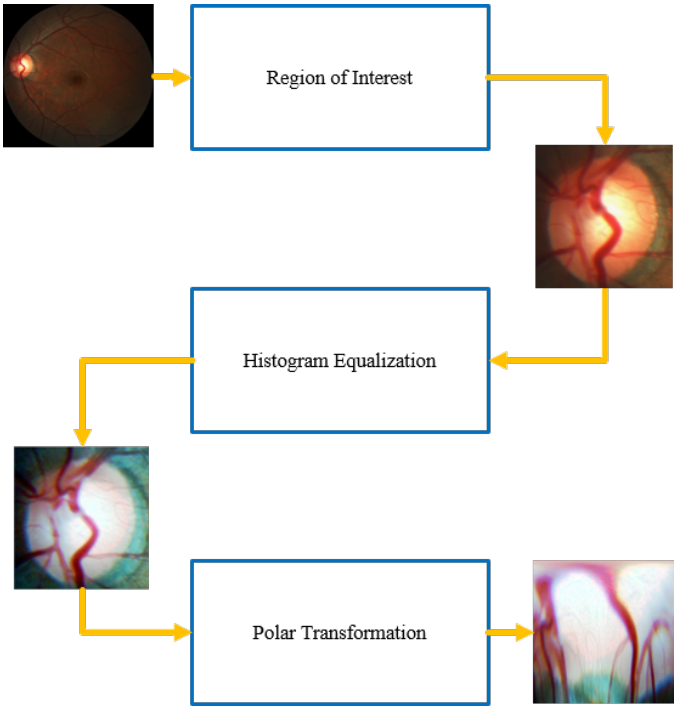


Fig. 2. Preprocessing flowchart. The original fundus image is cropped through the ROI (OC and OD), and is then processed using histogram equalization and a polar coordinate transformation to obtain the input image of the deep network model.

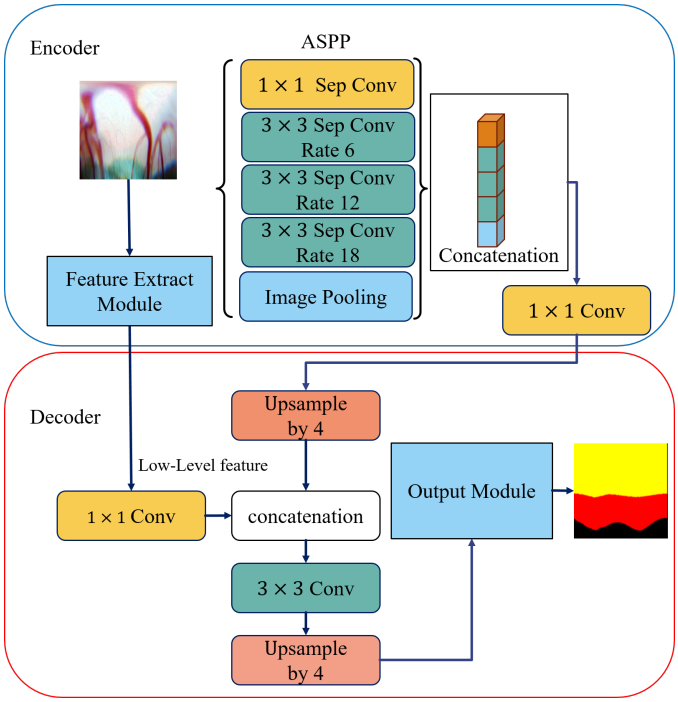


Fig. 3. Architecture of the deep learning network. With the encoder-decoder architecture as the backbone, the deep learning network includes feature extraction, atrous spatial pyramid pooling (ASPP), and multi-output modules, and generates a multilabel feature map that contains the OC and OD regions.

boundary box to estimate the vertical ellipse to achieve a rough segmentation. In [9], the DenseNet architecture is applied to reduce the parameters through reuse, and the generalizability of the model is tested on different datasets.

III. Proposed Method

In this section, we introduce the preprocessing of the fundus images and the architecture of the deep network applied. We employ ROI extraction, histogram equalization, and polar coordinate transformation for preprocessing, as shown in Fig. 2. In terms of deep network construction, we applied the encoder-decoder architecture of DeepLabV3+ and the lightweight concept of MobileNetV2 to construct a lightweight deep network for the segmenting of fundus images.

A. Image Preprocessing

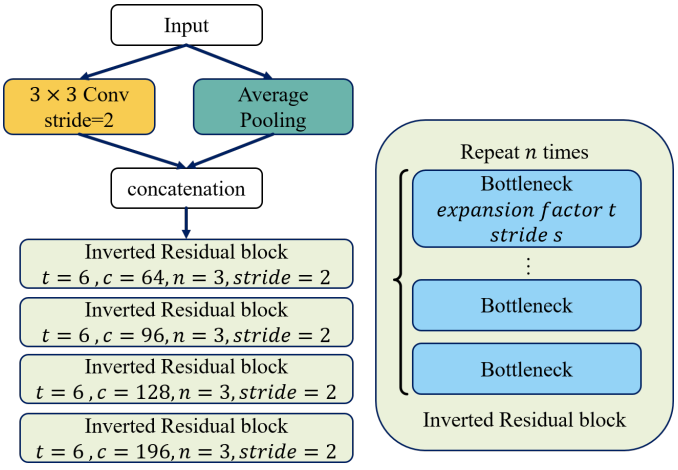
In general, the OC and OD regions in the fundus image account for a small proportion. The shapes of the OC and OD are close to a circle or ellipse, and the OC region does not exceed the OD region in size. Also, besides the slightly different colors in different segmented regions, the curvature of the blood vessels need to be considered to help judge the OC and OD. For the OC and OD regions of proportion, an existing automatic OC and OD positioning method is first adopted to briefly extract the OC and OD regions. Regarding the slightly different colors in different segmented regions, we applied histogram equalization to the three RGB channels in the image, and merged the three channels into a color image. In terms of approximating the shapes of the OC and OD, we used the center of the OC to establish a polar coordinate transformation.

B. Deep Network Architecture

A lightweight deep segmentation network is introduced in three steps in this section.

1) Encoder-decoder network architecture: In our study, we used the encoder-decoder architecture of DeepLabV3+ as the backbone of the deep network [24], as shown in Fig. 3. Based on the DeepLabV3+ architecture, we constructed a feature extraction module as the main body of the encoder, and then obtained the feature information of different scales through ASPP, and finally integrated all feature outputs with a 1×1 convolution. To make the model lightweight, ASPP and DeepLabV3+ in the encoder also uses a depthwise separable convolution instead of a standard convolution. In our case, the number of parameter calculations is significantly reduced, making the model lighter and faster, while still producing good results. The feature extraction module in the encoder uses the linear bottleneck of MobileNetV2 as the convolution method, which can mitigate the feature degradation problem of MobileNet and help our model produce better results.

The composition of the decoder is similar to that of DeepLabV3+. First, the features of the encoder are bilinearly upsampled by 4-fold. The feature extraction model is then concatenated to achieve the corresponding low-level features with the same spatial resolution. Finally, the decoder completes a simple module through the 4-fold bilinear upsampling



(A) Feature Extract Module (B) Inverted Residual block

Fig. 4. Feature extraction module constructed using inverted residual blocks. Fig. 4 (A) shows the overall structure of the feature extraction module. Four inverted residual blocks, shown in (B), are superimposed as feature extraction blocks. Here, n represents the number of repeat times, and c represents the output channel. The first bottleneck of each block uses stride s to decide whether to reduce the spatial resolution, and the remaining bottlenecks are set, such as stride = 1. The amplification factor t is based on the input of each block.

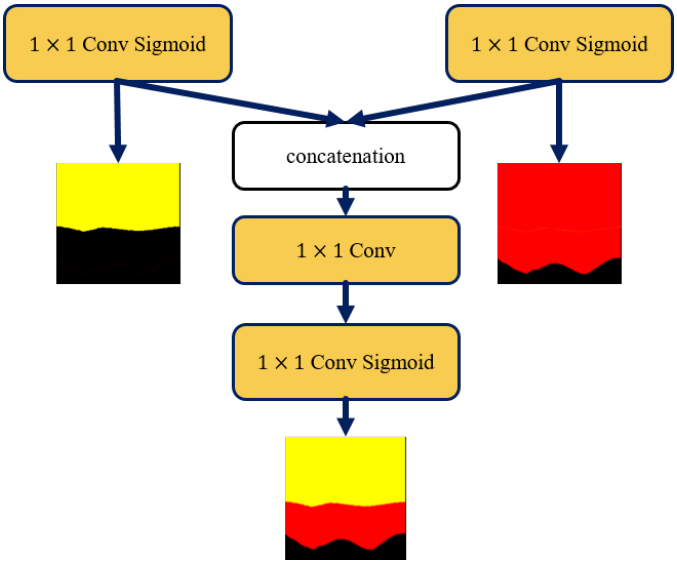


Fig. 5. Output module. Three different output results were obtained from 1×1 convolutions of three sigmoid activation functions. However, the left and right convolutions in the figure can generate the final result graph only after the features of OD and OC are merged through concatenations and 1×1 convolutions.

layer However, to achieve a better segmentation, we combined the original decoder and output module to supervise the learning in a multi-output manner and generate the OD and OC segmentation results.

2) Feature extraction module: In this module, to adapt to large-scale medical images, we used AveragePooling to reduce the image spatial resolution to the original 1/4 and reduce the dimensions of the two-stride convolution to extract the important features of the original image. A concatenation is then applied for the shrunken image and important features, reducing the burden from a large number of parameters and increasing the richness of the features. Two features of MobileNetV2, which possesses an inverted residual structure of a linear bottleneck, are then used to simplify the model size and stabilize the richness of the feature extraction. The use of the linear bottleneck component is improved based on the depthwise separable convolution, reducing the number of parameters. Finally, we stacked four inverted residual structures with linear bottlenecks to form the main blocks for extracting the feature information, as shown in Fig. 4 (A). The output of the first structure is used as a low-level feature to assist the decoder, which facilitates the OC and OD boundary segmentation.

3) Output module: Due to the OD region is contained in the OC region, two independent binary classifiers are applied for the multilabel classification of the OD and OC, which is closer to the real state and renders good results in M-Net. However, in addition to using the multilabel method as the main basis for segmentation, we added a separate segmentation of OD and OC to assist the supervision before multilabel segmentation, thus reducing the confusion between the OD and OC, as shown in Fig. 5. In addition, the output layers that separately segment the OD and OC only discriminate between regions and non-regions for each pixel value, and thus the OD and OC in the grayscale map of a single channel were used to regulate the final output segmentation. Finally, we used a Dice coefficient as the loss function. The loss weights of the OD and OC output layers were 0.3, whereas that of the output layer combining OD and OC was 0.4.

IV. Experiments

In the experiemnts, the input image size was fixed at 400×400. Because the experimental process mainly involved multilabel segmentation, the output of the deep learning network was a two-channel probability map of the OC and OD. Each pixel value represents a probability. A threshold value of 0.5 was used as the basis for dichotomization to generate a mask map. Segmentation results were generated using an ellipse fitting to facilitate the analysis of the CDR. To evaluate the segmentation performance of the methods, we compared the evaluation indicators such as a Dice coefficient (F-measurement), Jaccard (overlapping), accuracy, sensitivity, and specificity, as well as the actual segmentation graphs. We compared two well-known deep learning models applied to the fundus images, i.e., U-Net [25] and M-Net [22], using our method. First, the REFUGE training set was used for training, and the REFUGE verification set, RIM-ONE, and MESSIDOR were used for testing. The fundus images that had undergone a different preprocessing were used as the input. Finally, the existing methods used in RIM-ONE and MESSIDOR were compared.

For the evaluation of model, three types of preprocessed fundus images were used as the training inputs, including the

TABLE I
COMPARISON OF DIFFERNET METHODS UNDER DIFFERENT PREPROCESSING BASED ON THE REFUGE DATASET

Model	Optic Disc					Optic Cup				
	DC(F)	JC(O)	Acc	SEN	SPC	DC(F)	JC(O)	Acc	SEN	SPC
U-Net(Original)	0.887	0.803	0.817	0.324	0.994	0.822	0.719	0.797	0.622	0.993
U-Net(PT)	0.924	0.861	0.870	0.378	0.990	0.937	0.884	0.906	0.909	0.919
U-Net(HistPT)	**0.963**	**0.929**	**0.933**	**0.553**	**0.989**	**0.956**	**0.918**	**0.939**	**0.838**	**0.989**
M-Net(Original)	0.934	0.880	0.888	0.457	0.989	0.852	0.751	0.816	0.627	0.992
M-Net(PT)	0.954	0.915	0.922	0.556	0.994	0.825	0.716	0.789	0.603	0.995
M-Net(HistPT)	**0.964**	**0.932**	**0.936**	**0.526**	**0.995**	**0.967**	**0.937**	**0.953**	**0.936**	**0.960**
Proposed(Original)	0.948	0.902	0.908	0.437	0.997	0.926	0.866	0.901	0.769	0.983
Proposed(PT)	0.942	0.892	0.899	0.413	0.997	0.933	0.875	0.907	0.761	0.987
Proposed(HistPT)	**0.984**	**0.969**	**0.971**	**0.775**	**0.988**	**0.965**	**0.933**	**0.950**	**0.885**	**0.977**

TABLE II
COMPARISON OF DIFFERNET METHODS UNDER DIFFERENT PREPROCESSING BASED ON THE MESSIDOR DATASET

Model	Optic Disc					Optic Cup				
	DC(F)	JC(O)	Acc	SEN	SPC	DC(F)	JC(O)	Acc	SEN	SPC
U-Net(Original)	0.540	0.426	0.594	0.991	0.456	0.435	0.317	0.845	0.988	0.474
U-Net(PT)	0.552	0.435	0.847	0.914	0.804	0.290	0.191	0.921	0.970	0.307
U-Net(HistPT)	**0.843**	**0.751**	**0.942**	**0.988**	**0.788**	**0.537**	**0.391**	**0.931**	**0.997**	**0.410**
M-Net(Original)	0.566	0.433	0.691	0.993	0.443	0.277	0.202	0.955	0.963	0.504
M-Net(PT)	0.605	0.474	0.805	0.950	0.685	0.689	0.554	0.976	0.977	0.889
M-Net(HistPT)	**0.858**	**0.773**	**0.950**	**0.966**	**0.903**	**0.702**	**0.566**	**0.972**	**0.989**	**0.690**
Proposed(Original)	0.377	0.241	0.470	1	0.241	0.582	0.438	0.940	0.999	0.445
Proposed(PT)	0.341	0.212	0.382	1	0.212	0.566	0.428	0.931	0.997	0.446
Proposed(HistPT)	**0.886**	**0.805**	**0.963**	**0.984**	**0.872**	**0.717**	**0.580**	**0.972**	**0.991**	**0.688**

TABLE III
COMPARISON OF DIFFERNET METHODS UNDER DIFFERENT PREPROCESSING BASED ON THE DATASET

Model	Optic Disc					Optic Cup				
	DC(F)	JC(O)	Acc	SEN	SPC	DC(F)	JC(O)	Acc	SEN	SPC
U-Net(Original)	0.560	0.538	0.554	0.171	0.986	0.348	0.327	0.410	0.335	402
U-Net(PT)	0.976	0.954	0.954	0.504	0.969	0.881	0.814	0.827	0.713	0.910
U-Net(HistPT)	**0.949**	**0.903**	**0.906**	**0.308**	**0.998**	**0.951**	**0.910**	**0.921**	**0.670**	**0.994**
M-Net(Original)	0.877	0.846	0.849	0.249	0.906	0.06	0.04	0.168	0.141	0.490
M-Net(PT)	0.967	0.943	0.945	0.626	0.982	0.636	0.485	0.549	0.247	0.989
M-Net(HistPT)	**0.979**	**0.959**	**0.960**	**0.520**	**0.995**	**0.971**	**0.945**	**0.951**	**0.799**	**0.982**
Proposed(Original)	0.962	0.927	0.929	0.328	0.999	0.592	0.428	0.502	0.212	1
Proposed(PT)	0.973	0.948	0.950	0.407	0.997	0.399	0.249	0.3468	0.166	1
Proposed(HistPT)	**0.976**	**0.953**	**0.954**	**0.431**	**0.998**	**0.956**	**0.918**	**0.928**	**0.690**	**0.992**

original image (Original), polar transformed image (PT), and histogram equalized and polar transformed image (HistPT). Segmentation was applied using our method, U-Net, and M-Net. The segmentation results were compared, as shown in Tables I–III. Inspired by M-Net, we also used a polar coordinate transformation to balance the proportions of the OC, OD, and a background to avoid a model overfitting. However, one of our main contributions is to add histogram equalization to uniformize the pixel values in the image to obtain an image with a more prominent contour, which helps improve the extraction and segmentation of the model and achieve a better generalizability.

We observed from Tables I–III that the images processed through a polar coordinate transformation and histogram equalization obtained better segmentation results. The OC segmentation tested on the MESSIDOR and RIM-ONE datasets had a significant effect, indicating that a polar coordinate transformation can clearly describe the segmentation task, whereas histogram equalization can give the model a clearer outline and reduce the influence of chromatic aberrations. However, we found from Tables II and III that our method failed to easily complete the segmentation task when a polar coordinate transformation and histogram equalization were not employed. This indicates the need to clearly define the segmentation requirements while simplifying the deep network model. Fig. 6 shows an example of an actual segmentation of different models on different datasets. Each dataset was represented by two fundus images. The inputs of all deep models were preprocessed fundus images. The glaucoma and healthy fundus images were selected from the REFUGE and RIM-ONE datasets. On the MESSIDOR and REFUGE datasets, our method completely presented the OC segmentation compared to M-Net and U-Net, although our method was more likely to have a smaller OD segmentation range than that of the GT. In the test on the RIM-ONE dataset, we selected more challenging images. Our method performed better in the first image, whereas all methods failed in the segmentation of the second image. Therefore, in the future, we can try a different

TABLE IV
PERFORMANCE OF DIFFERENT MODELS ON DIFFERENT DATASETS

Model	Params	FLOPS
U-Net	10.13M	20297877
M-Net	8.54M	17081123
Proposed	6.71M	13510970

preprocessing or develop a strong deep network to strengthen the generalizability of the current model.

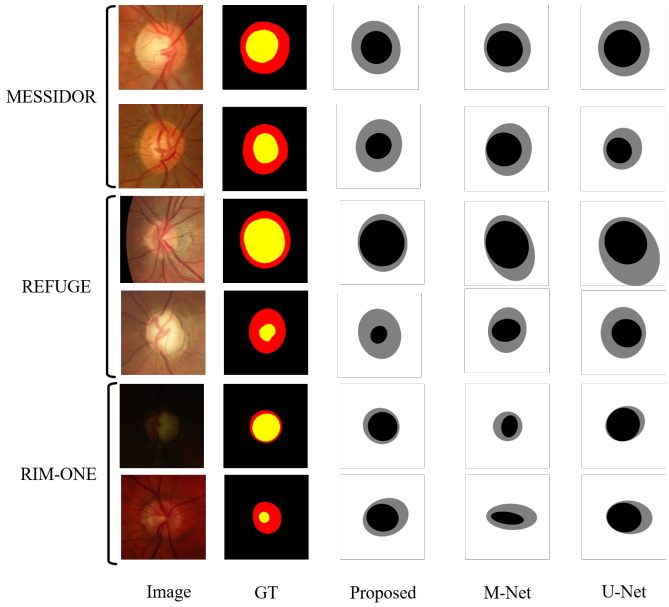


Fig. 6. Examples of OC segmentation. A fundus image, ground truth (GT), our network, M-Net, and U-Net are presented from left to right. Three different fundus image datasets are used. The yellow and red regions in GT are OD and OC, respectively. The results of OD and OC segmentation are expressed in black and gray, respectively.

We evaluated the size and speed of the models based on the number of parameters (Params) and the number of FLOPs applied per second. We observed from Table IV that U-Net took up the most space but had the worst segmentation effect. However, our model occupied the least amount of space and exhibited the fastest calculation speed. For the diagnosis of glaucoma, the diagnostic capability of our model was evaluated based on the actual CDR of the REFUGE dataset and information regarding the presence or absence of glaucoma. U-Net and M-Net were compared with our method. The images processed by U-Net and M-Net and images preprocessed by our method were used as inputs. First, we conducted an ellipse fitting for the segmented image generated by the model to facilitate our CDR calculation. We then trained a simple linear model based on the actual CDR and the information on the presence or absence of glaucoma, and judged whether glaucoma was present based on the calculated CDR.

Finally, the AUROC value of the CDR obtained by our method was 0.94, which was closer to the result of the real dataset. We observed from Fig. 7 that U-Net witnessed an increase of 0.18 in AUROC after preprocessing, and M-Net

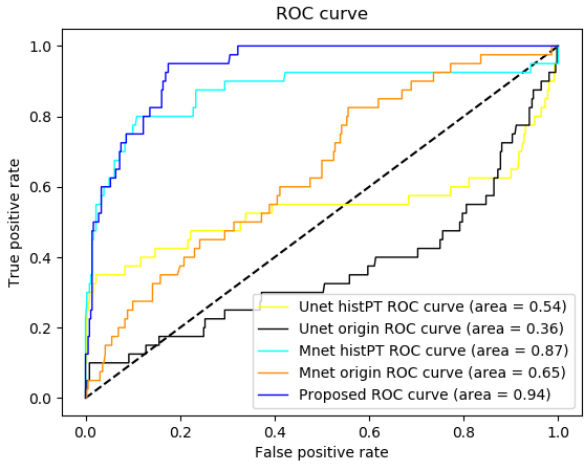


Fig. 7. AUROC evaluation for CDRs generated by different models on the REFUGE test set

witnessed an increase of 0.87. This confirms that this preprocessing method can improve the segmentation and diagnostic capability of other deep models.

V. CONCLUSION

In this study, we developed a deep learning method to simultaneously segment the OD and OC in a retinal fundus image. With the encoder-decoder architecture of DeepLabV3+ and the lightweight concept of MobileNetV2, we simplified the deep learning model to reduce the burden of high-resolution medical image input and enhance the features of the images through polar coordinate transformation and histogram equalization for better generalization. The results showed that our method performs well on different datasets in experiments. Our deep network architecture strucks a good balance between reducing the model parameter usage and improving the segmentation effect. Furthermore, our method can be applied to the REFUGE dataset through a CDR calculation for the glaucoma diagnosis.

REFERENCES

[1] Y.-C. Tham, X. Li, T. Y. Wong, H. A. Quigley, T. Aung, and C.-Y. Cheng, "Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis," *Ophthalmology*, vol. 121, no. 11, pp. 2081–2090, 2014.

[2] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.

[3] X. Zhang, X. Zhou, M. Lin, and J. Sun, "Shufflenet: An extremely efficient convolutional neural network for mobile devices," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 6848–6856.

[4] S. Mehta, M. Rastegari, A. Caspi, L. Shapiro, and H. Hajishirzi, "Espnet: Efficient spatial pyramid of dilated convolutions for semantic segmentation," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 552–568.

[5] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[6] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.

[7] V. Gulshan, L. Peng, M. Coram, M. C. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams, J. Cuadros *et al.*, "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *Jama*, vol. 316, no. 22, pp. 2402–2410, 2016.

[8] H. Fu, Y. Xu, D. W. K. Wong, and J. Liu, "Retinal vessel segmentation via deep learning network and fully-connected conditional random fields," in *2016 IEEE 13th international symposium on biomedical imaging (ISBI)*. IEEE, 2016, pp. 698–701.

[9] B. Al-Bander, B. Williams, W. Al-Nuaimy, M. Al-Taee, H. Pratt, and Y. Zheng, "Dense fully convolutional segmentation of the optic disc and cup in colour fundus for glaucoma diagnosis," *Symmetry*, vol. 10, no. 4, p. 87, 2018.

[10] J. I. Orlando, H. Fu, J. B. Breda, K. van Keer, D. R. Bathula, A. Diaz-Pinto, R. Fang, P.-A. Heng, J. Kim, J. Lee *et al.*, "Refuge challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs," *Medical image analysis*, vol. 59, p. 101570, 2020.

[11] F. Fumero, S. Alayón, J. L. Sanchez, J. Sigut, and M. Gonzalez-Hernandez, "Rim-one: An open retinal image database for optic nerve evaluation," in *2011 24th international symposium on computer-based medical systems (CBMS)*. IEEE, 2011, pp. 1–6.

[12] A. Aquino, M. E. Gegúndez-Arias, and D. Marín, "Detecting the optic disc boundary in digital fundus images using morphological, edge detection, and feature extraction techniques," *IEEE transactions on medical imaging*, vol. 29, no. 11, pp. 1860–1869, 2010.

[13] A. Giachetti, L. Ballerini, and E. Trucco, "Accurate and reliable segmentation of the optic disc in digital fundus images," *Journal of Medical Imaging*, vol. 1, no. 2, p. 024001, 2014.

[14] J. Lowell, A. Hunter, D. Steel, A. Basu, R. Ryder, E. Fletcher, and L. Kennedy, "Optic nerve head segmentation," *IEEE Transactions on medical Imaging*, vol. 23, no. 2, pp. 256–264, 2004.

[15] M. C. V. S. Mary, E. B. Rajsingh, J. K. K. Jacob, D. Anandhi, U. Amato, and S. E. Selvan, "An empirical study on optic disc segmentation using an active contour model," *Biomedical Signal Processing and Control*, vol. 18, pp. 19–29, 2015.

[16] D. W. K. Wong, J. Liu, N. M. Tan, F. Yin, B.-H. Lee, and T. Y. Wong, "Learning-based approach for the automatic detection of the optic disc in digital retinal fundus photographs," in *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*. IEEE, 2010, pp. 5355–5358.

[17] Y. Xu, L. Duan, S. Lin, X. Chen, D. W. K. Wong, T. Y. Wong, and J. Liu, "Optic cup segmentation for glaucoma detection using low-rank superpixel representation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2014, pp. 788–795.

[18] K.-K. Maninis, J. Pont-Tuset, P. Arbeláez, and L. Van Gool, "Deep retinal image understanding," in *International conference on medical image computing and computer-assisted intervention*. Springer, 2016, pp. 140–148.

[19] Y. Guo, B. Zou, Z. Chen, Q. He, Q. Liu, and R. Zhao, "Optic cup segmentation using large pixel patch based cnns," in *Proc. the Ophthalmic Medical Image Analysis Third International Workshop (OMIA 2016)*, 2016, pp. 129–136.

[20] A. Sevastopolsky, "Optic disc and cup segmentation methods for glaucoma detection with modification of u-net convolutional neural network," *Pattern Recognition and Image Analysis*, vol. 27, no. 3, pp. 618–624, 2017.

[21] H. Fu, J. Cheng, Y. Xu, C. Zhang, D. W. K. Wong, J. Liu, and X. Cao, "Disc-aware ensemble network for glaucoma screening from fundus image," *IEEE transactions on medical imaging*, vol. 37, no. 11, pp. 2493–2501, 2018.

[22] H. Fu, J. Cheng, Y. Xu, D. W. K. Wong, J. Liu, and X. Cao, "Joint optic disc and cup segmentation based on multi-label deep network and polar transformation," *IEEE transactions on medical imaging*, vol. 37, no. 7, pp. 1597–1605, 2018.

[23] X. Sun, Y. Xu, M. Tan, H. Fu, W. Zhao, T. You, and J. Liu, "Localizing optic disc and cup for glaucoma screening via deep object detection networks," in *Computational Pathology and Ophthalmic Medical Image Analysis*. Springer, 2018, pp. 236–244.

[24] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 801–818.

[25] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.