


## Article

# An End-to-End Real-Time Lightweight Network for the Joint Segmentation of Optic Disc and Optic Cup on Fundus Images

Zhijie Liu <sup>1,2</sup>, Yuanqiong Chen <sup>1,3</sup>, Xiaohua Xiang <sup>4</sup>, Zhan Li <sup>5</sup> , Bolin Liao <sup>1</sup> and Jianfeng Li <sup>1,2,\*</sup><sup>1</sup> School of Computer Science and Engineering, Jishou University, Jishou 416000, China<sup>2</sup> School of Communication and Electronic Engineering, Jishou University, Jishou 416000, China<sup>3</sup> School of Computer Science and Engineering, Central South University, Changsha 410083, China<sup>4</sup> Department of Computer Science, Xiangxi National Vocational and Technical College, Jishou 416000, China<sup>5</sup> Department of Computer Science, Swansea University, Swansea SA1 8EN, UK

\* Correspondence: ljf@jsu.edu.cn

**Abstract:** Glaucoma is the second-most-blinding eye disease in the world and accurate segmentation of the optic disc (OD) and optic cup (OC) is essential for the diagnosis of glaucoma. To solve the problems of poor real-time performance, high algorithm complexity, and large memory consumption of fundus segmentation algorithms, a lightweight segmentation algorithm, GlauNet, based on convolutional neural networks, is proposed. The algorithm designs an efficient feature-extraction network and proposes a multiscale boundary fusion (MBF) module, which greatly improves the segmentation efficiency of the algorithm while ensuring segmentation accuracy. Experiments show that the algorithm achieves Dice scores of 0.9701/0.8959, 0.9650/0.8621, and 0.9594/0.8795 on three publicly available datasets—Drishti-GS, RIM-ONE-r3, and REFUGE-train—for both the optic disc and the optic cup. The number of model parameters is only 0.8 M, and it only takes 13 ms to infer an  $800 \times 800$  fundus image on a GTX 3070 GPU.



**Citation:** Liu, Z.; Chen, Y.; Xiang, X.; Li, Z.; Liao, B.; Li, J. An End-to-End Real-Time Lightweight Network for the Joint Segmentation of Optic Disc and Optic Cup on Fundus Images. *Mathematics* **2022**, *10*, 4288. <https://doi.org/10.3390/math10224288>

Academic Editor: Jakub Nalepa

Received: 11 October 2022

Accepted: 14 November 2022

Published: 16 November 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** convolutional neural network; optic disc and cup segmentation; glaucoma screening; medical auxiliary diagnosis

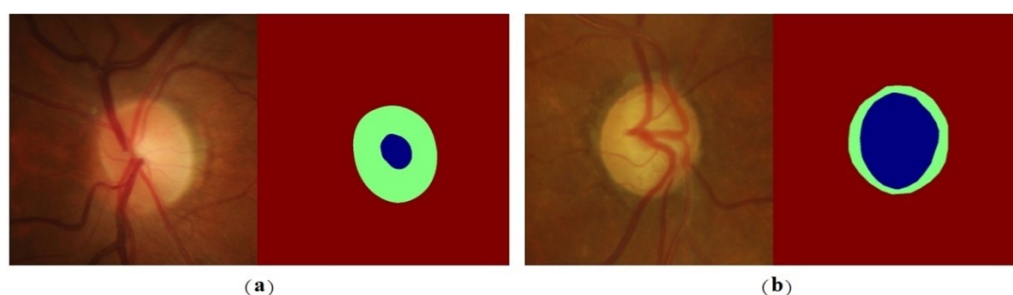
**MSC:** 68T07

## 1. Introduction

Glaucoma is a chronic eye disease that causes irreversible damage to vision. Patients with glaucoma suffer damage to the optic nerve due to increased intraocular pressure (IOP) caused by an imbalance between fluid production and drainage in the eye. The vertical cup-to-disk ratio (CDR) is one of the commonly used indicators for clinical screening of glaucoma. Usually, a CDR greater than 0.65 [1] is diagnosed as glaucoma. Figure 1 is a sample map of normal eyes and glaucoma fundus and the corresponding annotation map of the OD and OC. Accurate segmentation of the OD and the OC is essential for accurate CDR acquisition. At present, the clinical diagnosis of glaucoma is mainly made by ophthalmologists through manual diagnosis, which is somewhat subjective—results vary greatly between doctors and are inefficient. With the rapid development of information technology, rapid advances in medically assisted diagnostic techniques have been made, making large-scale glaucoma screening possible.

Semantic segmentation is a fundamental task of computer vision. With the great success of deep learning in the field of computer vision, algorithms based on deep learning have improved in efficiency and accuracy compared with traditional machine learning algorithms, providing new ideas for the development of medical image-assisted diagnosis technology. In recent years, segmentation algorithms for OD and OC based on deep learning have emerged one after another. The following are some excellent algorithms based on convolutional neural networks: The authors of [2] proposed an attention U-Net

fundus-image-segmentation algorithm based on transfer learning. The algorithm proposes an attention gate module which is used to focus on the target area. For the acquisition of the pretraining weights of the algorithm, first, model training is performed on the DRIONS-DB dataset to obtain a set of pretraining weights; these are trained on the Drishti-GS dataset to further modify the pretraining weights. Finally, the fundus images are segmented using the trained attention U-Net model combined with transfer learning. The authors of [3] proposed a segmentation network called BGA-Net, which is combined with adversarial learning to obtain a set of optimal model weights through alternate training to better segment OD and OC. The authors of [4] proposed an unsupervised domain-adaptation network called BEAL, which suppresses the geometric structure of the boundary while generating more realistic boundaries through adversarial learning. This method effectively reduces the formation of sawtooth on the segmentation boundary and improves the segmentation accuracy of OD and OC. The authors of [5] proposed a two-stage approach that first locates the OD and then jointly segments the OD and OC according to the region of interest. The method uses depthwise-separable convolution to improve the segmentation efficiency of the network model and adds a multiscale image pyramid to improve the accuracy and robustness of OD and OC segmentation methods. In [6], an unsupervised adaptive segmentation method for OD and OC is proposed. The method uses the image synthesis mechanism of GAN for feature alignment of the output image, while an edge-attention module (EAM) is introduced to enhance the representation of boundary information. The method outperforms other methods in the unsupervised approach, while the method is more advantageous on small datasets. The above deep-learning-based methods have achieved excellent segmentation performance in OD and OC segmentation tasks, but also have the following problems: (1) Feature-extraction networks that use classical classification networks as segmentation models (such as VGG [7], ResNet [8], DeepLab [9–12]). While such networks have good semantic segmentation performance, this is accompanied by tens of millions of network parameters. (2) The algorithm design is often complex, the computational complexity is high, and the reasoning time is long; so, it cannot meet the needs of large-scale glaucoma screening. (3) The algorithm itself has high requirements for computing power and the memory of the device, making it difficult to deploy and apply on mobile devices.



**Figure 1.** Example of fundus images. (a) Normal fundus and corresponding annotation map; (b) glaucoma fundus and corresponding annotation map. Green represents optic disc; blue represents optic cup.

In response to the above problems, this paper is devoted to designing a lightweight and efficient OD and OC segmentation algorithm, balancing the performance of the algorithm, model size, and inference speed so that the algorithm can meet the requirements of mobile devices. Real-time semantic segmentation has now become an important topic in edge computing and a large number of excellent algorithms have been proposed. Inspired by real-time segmentation networks [13–16], this algorithm designs a simple and lightweight feature-extraction network using a small amount of ordinary convolution, deep separable (DS) convolution, and asymmetric convolution (AC) for the extraction of spatial detail and contextual information, and a multiscale boundary fusion (MBF) [17–19] module to capture the OD and OC boundaries. In summary, the main contributions of this paper are as follows:

- 1 An end-to-end lightweight and efficient network model for OD and OC segmentation, GlauNet, is proposed, which significantly reduces the number of model parameters and computational complexity, and achieves competitive OD and OC segmentation results without the need for pretrained weights.
- 2 A multiscale boundary fusion (MBF) module is designed according to the characteristics of fundus images, including a multiscale feature fusion (MFF) branch and a boundary feature auxiliary (BFA) branch. This module fuses the multiscale feature map obtained by the MFF branch and the boundary feature map obtained by the BFA branch, which improves the segmentation accuracy of the optic disc and the optic cup and the robustness of the segmentation algorithm.
- 3 GlauNet is committed to the deployment and application of mobile devices. The model parameters are only 0.8M, and it only takes 13 ms to assess an  $800 \times 800$  fundus image on a GTX 3070 GPU.

The rest of this paper is organized as follows: Section 2 presents the segmentation model; Section 3 presents the experimental details and experimental results; Section 4 presents the ablation experiments; Section 5 discusses the algorithm and experimental results; Section 6 presents the summary of the paper.

## 2. Methods

The GlauNet segmentation model is mainly composed of three modules: the spatial-detail information-extraction module (A); the context-information-extraction module (B); and the decoding head module (C). The network structure is shown in Figure 2.

### 2.1. Overall Architecture

#### 2.1.1. Spatial-Detail Information-Extraction Module

The spatial-detail information-extraction module consists of three standard convolutional layers and an MBF module, where the blue dotted box in module A represents the MBF module. To reduce the loss of detail information during convolution, the loss of spatial detail during convolution is reduced by sequentially increasing the number of channels in the convolution layer. As the spatial information in fundus images is relatively homogeneous, using a large number of convolutional channels is a small or even detrimental improvement in feature-extraction ability. At the same time, to keep the spatial-detail information-extraction module lightweight and efficient, the number of channels of the three standard convolutional layers is 32, 48, and 64 in turn; each convolutional layer is followed by a batch-normalization [20] layer and a ReLU activation function, each with a stride of 2 and a convolutional kernel size of 3. To obtain richer semantic information, the MBF module is used for the extraction of multiscale information and the boundary information of the OD and OC. The size of the feature map output from the spatial-detail information-extraction module is  $1/8$  of the size of the input image. The specific structure is shown as A module in Figure 2.

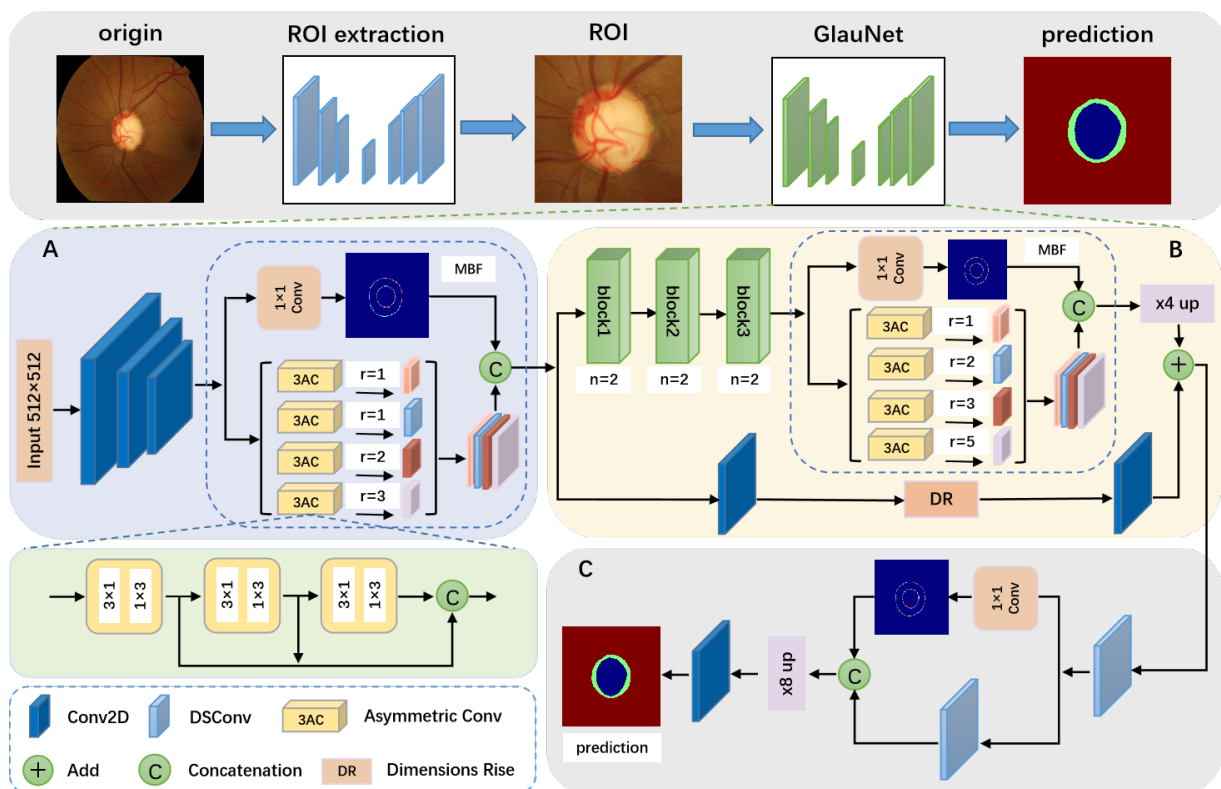
#### 2.1.2. Context-Information-Extraction Module

The contextual information extraction module consists of three blocks and an MBF module, where the blue dotted box in module B represents the MBF module. Each block consists of two bottleneck-inverted residual structures, and the function of the block is to efficiently extract the contextual information of the fundus image. To reduce model parameters and computation, this module uses depthwise-separable (DS) convolutions instead of standard convolutions. DS convolution consists of depthwise (DW) convolution and pointwise (PW) convolution. Compared with DS convolution, standard convolution is theoretically 8–9 times more costly in terms of the number of parameters and computation than DS convolution. The ratio of parameter amount and calculation amount of standard convolution and DS convolution is shown in Equation (1), the numerator is standard convolution, and the denominator is DS convolution.

$$\begin{cases} C_{cc} = \frac{F \times F \times M \times N \times H \times W}{F \times F \times M \times H \times W + M \times N \times H \times W} \\ C_{pc} = \frac{F \times F \times M \times N}{F \times F \times M + M \times N} \\ C_{cc} = C_{pc} = \frac{F^2 \times N}{F^2 + N} \end{cases} \quad (1)$$

where  $F$  is the size of the convolution kernel,  $M$  is the number of input channels,  $N$  is the number of output channels, and  $H$  and  $W$  are the height and width of the input image, respectively.  $C_{cc}$  is the proportion of the computational cost and  $C_{pc}$  is the proportion of the parametric cost.

Specifically, the efficient bottleneck-inverse residual structures from MobileNet-v2 [21] are used. The number of bottleneck-inverse residual structures in each block is 2, and the number of channels in the output feature map is 64, 96, and 128 in that order. In the first two blocks, the first bottleneck-inverse residual structures has a stride of 2, and the remaining bottleneck-inverse residual structures have a stride of 1. Then, an MBF module is connected to obtain a contextual feature map of size  $16 \times 16$ . To obtain more spatial detail information, the obtained contextual feature map is first upsampled four times, then the number of channels of the spatial detail feature map obtained from the spatial-detail information-extraction module is dimensionally increased to the same number of channels as the contextual feature map ( $65 \rightarrow 129$ ); then, the upsampled contextual feature map and the spatial detail feature map are fused with the features. The specific structure is shown in Module B in Figure 2.



**Figure 2.** The network structure diagram of GlauNet. Spatial-detail information-extraction module (A), context-information-extraction module (B), and decoding header module (C). The blue dashed boxes in modules A and B are MBF modules.

### 2.1.3. Decoder Header Module

The decoding head module consists of two DS convolutions, a boundary feature auxiliary branch, and a standard convolution. The feature map fused by the context-information-extraction module and the spatial-detail information-extraction module is sent to the decoding head module. In the decoding head module, a DS convolution is first

performed, and then the obtained feature maps are sent to a boundary feature auxiliary branch and a DS convolution branch, respectively. Then, the output feature maps of the two branches are concatenated. The obtained feature map is classified into pixels using standard convolution to obtain a segmented image with a size of  $64 \times 64$ , and the number of output channels is 129, 130, and 2. Finally, the segmentation map is upsampled eight times to obtain the segmentation map of the OD and OC with the same size as the input image. All convolution operations in this module have a stride of 1. The specific structure is shown in the C module in Figure 2.

## 2.2. Multiscale Boundary-Fusion Module

According to the features of fundus images, the MBF module is designed to better extract semantic information and boundary information. The module includes a boundary feature auxiliary (BFA) branch and a multiscale feature fusion (MFF) branch. The BFA branch is used to extract the boundary features of the input feature map, i.e., the boundaries of the OD and the OC. In this paper, the standard  $1 \times 1$  convolution is used for the boundary extraction of the OD and the OC, with a stride of 1. The BFA branching operation is computationally trivial but is very effective in extracting the boundaries of the OD and the OC, as experimentally demonstrated in the ablation experiments section. The MFF branch includes multiple feature-extraction branches with different scales; each branch performs feature extraction through AC of different expansion coefficients and then performs feature fusion on the feature maps of different branches. Finally, the obtained boundary information map and multiscale feature fusion map are concatenated to obtain richer semantic information.

The specific structure of the MBF in the blue dashed box in modules A and B in Figure 2 is shown in Figure 3, including four multiscale feature-extraction branches and one boundary feature auxiliary branch. Each multiscale feature-extraction branch consists of three asymmetric convolutions [17], and each AC consists of an  $n \times 1$  and  $1 \times n$  convolution. Here, we decompose the  $3 \times 3$  convolution into a  $3 \times 1$  convolution and a  $1 \times 3$  convolution. Compared with a  $3 \times 3$  convolution operation, using AC saves about 33% of the number of parameters for the same number of convolution kernels. The theoretical analysis is shown in Equation (2). The fusion of multiscale features is performed by setting different expansion coefficients. The size of the convolution kernels of the four multiscale feature-extraction branches is  $3 \times 3$ , and the size of the receptive field of the convolution kernels with different expansion coefficients is shown in Equation (3). For the MBF module in the spatial-detail information-extraction module, the expansion coefficients of the four branches are [1, 1, 2, 3] in sequence. For the MBF module in the context-information-extraction module, the expansion coefficients of the four branches are [1, 2, 3, 5] in sequence. At the same time, to reduce the feature loss in the convolution process, for each multiscale branch, the output feature maps of the three AC operations are concatenated as the output feature map of the branch.

$$P = \frac{n \times 1 + 1 \times n}{n \times n} \quad (2)$$

$P$  is the ratio of the parameters of the AC operation and the standard convolution operation, and  $n$  is the size of the convolution kernel. As the size of the convolution kernel increases, the proportion of the number of parameters saved by the AC operation continues to increase.

$$R = [c(n - 1) + 1]^2 \quad (3)$$

$c$  is the expansion coefficient, and  $n$  is the size of the convolution kernel, and  $R$  represents the size of the receptive field of the convolution kernel.



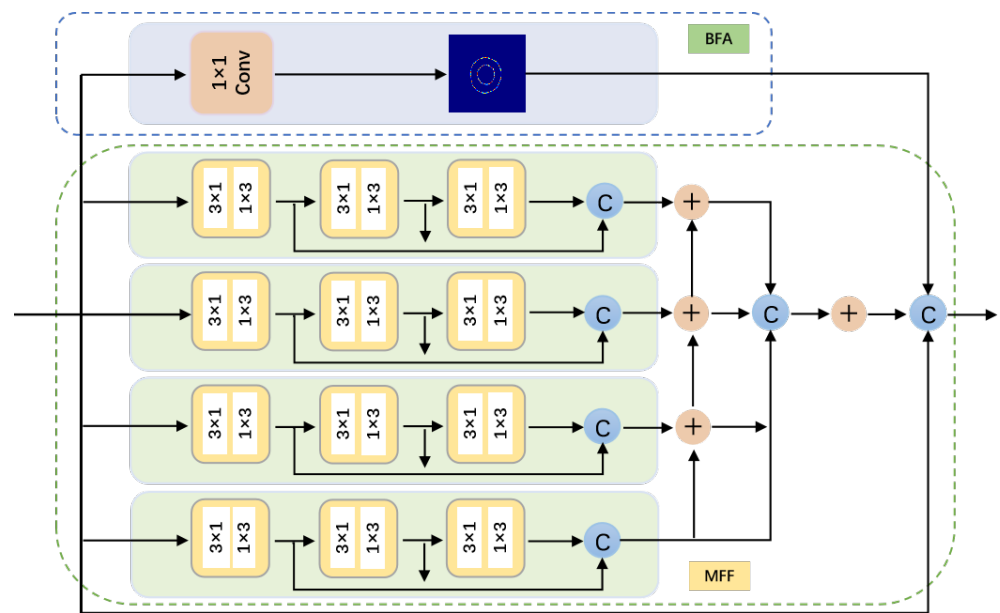


Figure 3. Structure of the multiscale boundary-fusion module.

2.3. Loss Function

This paper uses the binary cross-entropy function as the loss function of the segmentation algorithm. The loss function is defined as follows:

$$\sigma(z) = \frac{1}{1 + e^{-z}} \tag{4}$$

$$loss = -\frac{1}{N} \sum_i [y_i^m \log(\sigma(p_i^m)) + (1 - y_i^m) \log(1 - \sigma(p_i^m))] \tag{5}$$

$\sigma(z)$  is the Sigmoid function,  $N$  is the number of pixels,  $y_i^m$  is the annotation map, and  $p_i^m$  is the prediction map of GlauNet.

3. Experiments and Results

3.1. Datasets

Drishti-GS [22]: The Drishti-GS dataset consists of 101 fundus images with a resolution of  $2047 \times 1759$ . The dataset consists of 31 normal fundus images and 70 glaucoma fundus images, with 50 images in the training set and 51 in the testing set. Each image was manually annotated by four ophthalmologists with different clinical experiences.

RIM-ONE-r3 [23]: The RIM-ONE-r3 dataset consists of 159 fundus images with a resolution of  $2144 \times 1424$ , including 85 normal fundus and 74 glaucoma fundus images. Among the 159 fundus images, 99 fundus images were used as the training set and 60 fundus images were used as the testing set.

REFUGE [24]: The REFUGE dataset consists of 1200 fundus images, including 1080 normal fundus images and 120 glaucoma fundus images. The dataset consists of a training set, a validation set, and a testing set, each subset containing 400 fundus images. The training set has a resolution of  $2124 \times 2056$  and the validation and testing sets have a resolution of  $1634 \times 1634$ .

3.2. Implementation Details

The network model was implemented based on the PyTorch 1.10 deep learning framework with CUDA version 11.4, and all experiments were conducted on a single NVIDIA GTX 3070 GPU. The network model is trained using the Adam optimizer, and the Momentum is set to 0.9. The network model uses the poly learning strategy for model training. The decay strategy of the learning rate is  $lr = base-lr \times (1 - \frac{iter}{max-iter})^{power}$ , the initial learning

rate is set to 0.001, and the power is set to 0.9. Due to the small size of the Drishti-GS and RIM-ONE-r3 datasets, we achieved convergence of the loss function after training 1000 epochs on these two datasets with a batch size equal to 12. For the REFUGE dataset, this paper uses the REFUGE-train subset as the experimental dataset. We use 320 REFUGE-train subset fundus images as the training set and 80 REFUGE-train subset fundus images as the testing set, and the loss function converges after training 300 epochs with batch size equal to 12.

For the preprocessing of fundus images, the region of interest was cropped according to the literature [25] and the final input image size was  $512 \times 512$ . After an experimental comparison, it was found that the cropping of the region of interest not only reduces the computational effort of the network model but also has a greater improvement on the segmentation results. Due to the relative difficulty of acquiring fundus images, the current publicly available fundus dataset is relatively small and we have used many data-enhancement methods to increase the diversity of the data. The data-enhancement methods we use are random scaling, rotation, flipping, elastic transformation, contrast adjustment, adding noise, and random erasure. To optimize the output, we also perform morphological operations on the resulting segmented images to make the boundaries of the segmented images smoother and more natural using erosion and hole-filling operations, the principles of which are shown in (6) and (7).

$$A \ominus B = \{z|(B)_z \subseteq A\} \tag{6}$$

$A$  represents the target of post-processing,  $B$  is the structural element,  $\ominus$  represents that  $B$  performs the corrosion operation on  $A$ , and  $z$  is the size of the translation vector.

$$X_k = (X_{k-1} \oplus B) \cap A^c \quad k = 1, 2, 3, \dots \tag{7}$$

$X$  represents the set of all filled holes,  $B$  is the structural element,  $A^c$  is the complement of the fundus image,  $\oplus$  is the hole operation, and  $k$  is the number of iterations.

### 3.3. Evaluation Criteria

This paper uses DI (dice index), Jaccard (IoU), sensitivity (SEN), and  $CDR$  as the evaluation criteria for the GlauNet segmentation network. DI, Jaccard, SEN, and  $CDR$  are defined as follows:

$$DI = \frac{2 \times N_{TP}}{2 \times N_{TP} + N_{FP} + N_{FN}} \tag{8}$$

$$Jaccard = \frac{N_{TP}}{N_{TP} + N_{FP} + N_{FN}} \tag{9}$$

$$Sensitivity = \frac{N_{TP}}{N_{TP} + N_{FN}} \tag{10}$$

$$CDR = \frac{OC}{OD} \tag{11}$$

$$\delta = abs(CDR_p - CDR_g) \tag{12}$$

$N_{TP}$ ,  $N_{FP}$ , and  $N_{FN}$  represent the number of true positives, false positives, and false negatives, respectively,  $CDR_p$  and  $CDR_g$  represent the vertical cup-to-disc ratio of the OD and OC of the predicted segmentation map and the vertical cup-to-disc ratio of the OD and OC of the annotated map, respectively. This paper uses the average  $CDR$  error  $\delta$  to evaluate the difference between  $CDR_p$  and  $CDR_g$ , and lower  $\delta$  values represent better prediction results.

### 3.4. Experimental Results

This paper presents a comparative analysis of the experimental results in terms of both quantitative and qualitative aspects. For the Drishti-GS and RIM-ONE-r3 datasets, the results of the experimental comparison of the method in this paper with some classical methods and some current advanced methods are shown in Tables 1 and 2. Some methods are not open-source; we obtained experimental results from the original paper, and we conducted experiments under the same conditions. The results show that the proposed method achieves competitive or state-of-the-art experimental results on various evaluation metrics. Figures 4 and 5 present the qualitative experimental results comparison with U-Net [26], FCN-8s [27], and BGA-Net [3]. Green represents the OD boundary and blue represents the OC boundary.

**Table 1.** Quantitative results of different methods on the Drishti-GS dataset.

Methods	OD			OC			$\delta$
	DI	Jaccard	SEN	DI	Jaccard	SEN	
U-Net [26]	0.9487	0.9049	0.9198	0.8290	0.7254	0.8239	0.081
FCN-8s [27]	0.9559	0.9177	0.9367	0.8635	0.7730	0.8669	0.060
LR-ASPP [28]	0.9675	0.9378	0.9710	0.8645	0.7777	0.8864	0.064
Tabassum [29]	0.9597	0.9183	0.9754	0.924	0.8632	0.9567	-
Zilly [30]	0.973	0.914	-	0.871	0.85	-	-
AlBander [31]	0.949	0.9042	0.9268	0.8282	0.7113	0.7413	-
Xiao [2]	0.9638	0.9301	0.9488	0.8793	0.7846	0.8765	-
M-Net [25]	0.9678	0.9386	0.9711	0.8618	0.7730	0.8822	0.092
CE-Net [32]	0.9642	0.9323	0.9759	0.8818	0.8006	0.8819	0.076
MSMKU [33]	0.9780	0.9496	0.9792	0.8921	0.8232	0.9157	0.054
Proposed	0.9701	0.9422	0.9878	0.8959	0.8223	0.9343	0.043

**Table 2.** Quantitative results of different methods on the RIM-ONE-r3 dataset.

Methods	OD			OC			$\delta$
	DI	Jaccard	SEN	DI	Jaccard	SEN	
U-Net [26]	0.9273	0.8778	0.9055	0.7434	0.6350	0.6812	0.104
FCN-8s [27]	0.9543	0.9144	0.9318	0.8016	0.6898	0.7417	0.078
LR-ASPP [28]	0.9582	0.9219	0.9571	0.8330	0.7334	0.8414	0.077
Yu [34]	0.9610	0.9256	-	0.8445	0.7429	-	-
Tabassum [29]	0.9582	0.9101	0.9734	0.8622	0.7532	0.9517	-
AlBander [31]	0.9036	0.8289	0.8737	0.6903	0.5567	0.9052	-
Xiao [2]	0.9401	0.8870	0.9236	0.8397	0.7237	0.8133	-
CE-Net [32]	0.9527	0.9115	0.9502	0.8435	0.7424	0.8352	0.059
M-Net [25]	0.9526	0.9114	0.9481	0.8348	0.7300	0.8146	0.059
MSMKU [33]	0.9561	0.9172	0.9521	0.8564	0.7586	0.8515	0.051
Proposed	0.9650	0.9328	0.9809	0.8623	0.7666	0.8823	0.058

For the REFUGE-train dataset, this paper reproduces the methods in Table 3 under the same conditions. The results show that our method also achieves competitive or state-of-the-art results on various evaluation metrics. The qualitative experimental results are shown in Figure 6.



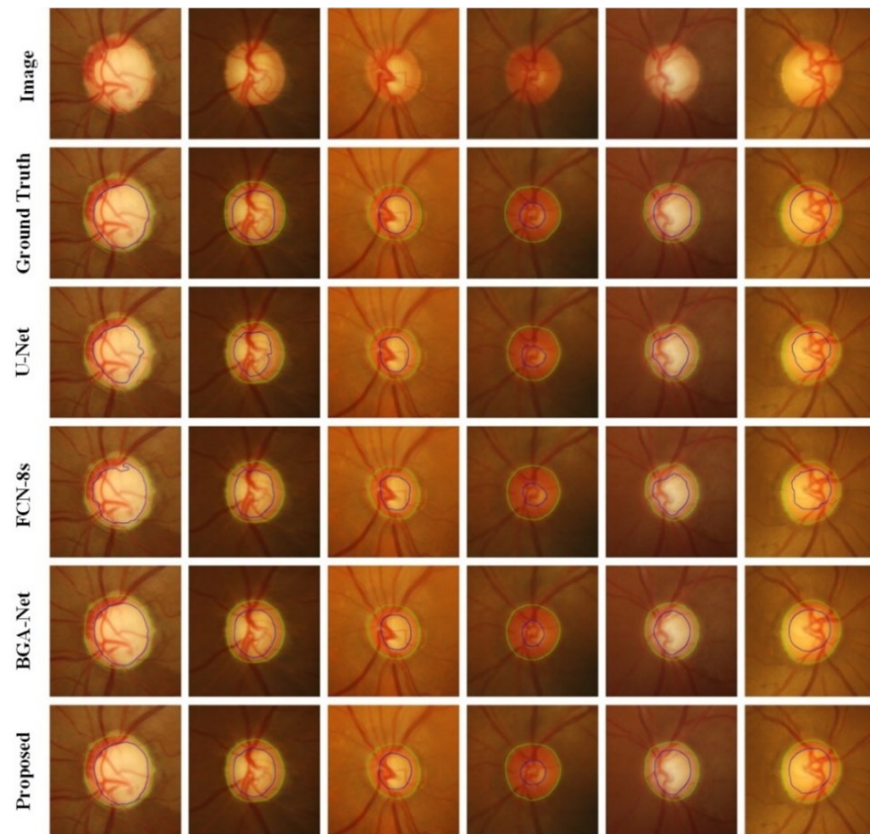


Figure 4. Visualization of optic disc and optic cup segmentation results in the Drishti-GS dataset.

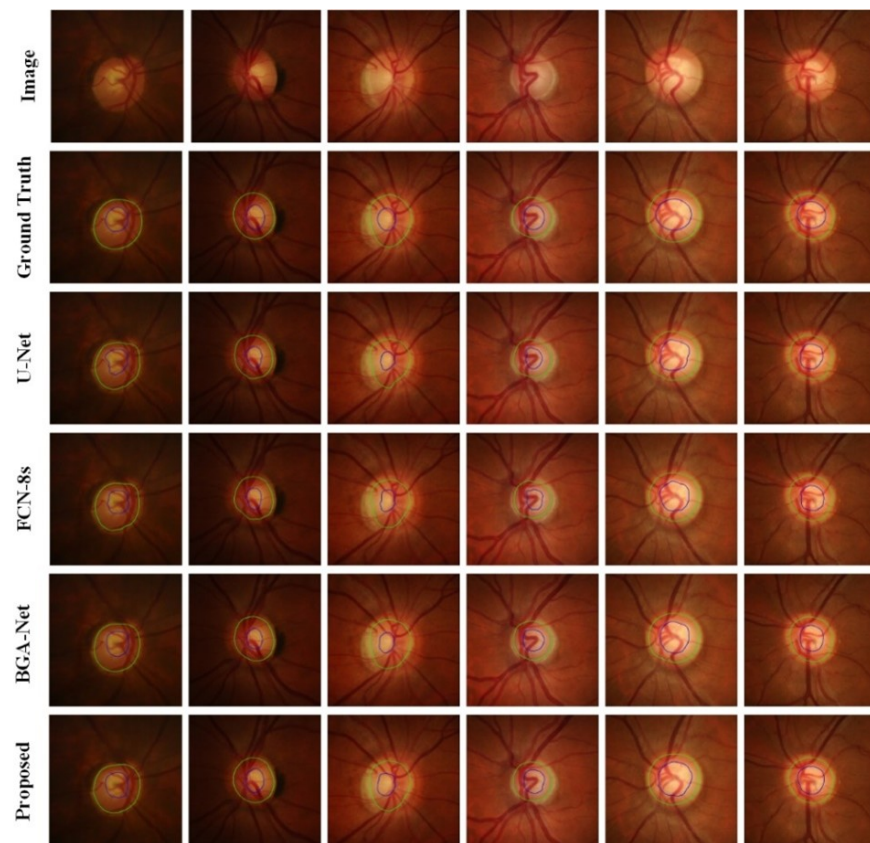
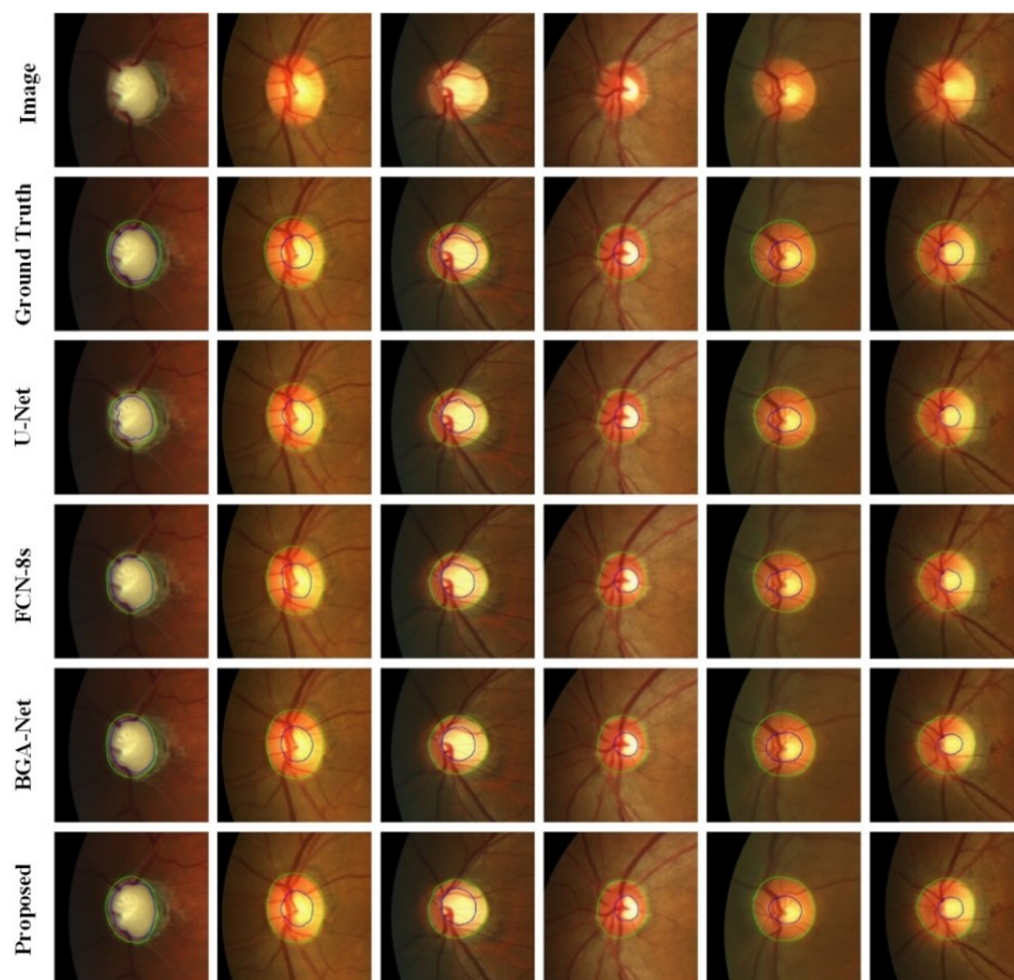


Figure 5. Visualization of optic disc and optic cup segmentation results in the RIM-ONE-r3 dataset.

**Table 3.** Quantitative results of different methods on the REFUGE-train dataset.

Methods	OD			OC			$\delta$
	DI	Jaccard	SEN	DI	Jaccard	SEN	
U-Net [26]	0.9457	0.8980	0.9137	0.8608	0.7621	0.8173	0.068
FCN-8s [27]	0.9530	0.9108	0.9325	0.8645	0.7674	0.8534	0.063
LR-ASPP [28]	0.9603	0.9252	0.9619	0.8779	0.7915	0.9029	0.059
BGA-Net [3]	0.9627	0.9286	0.9699	0.8779	0.7884	0.9001	0.056
Proposed	0.9594	0.9228	0.9741	0.8795	0.7927	0.9186	0.055



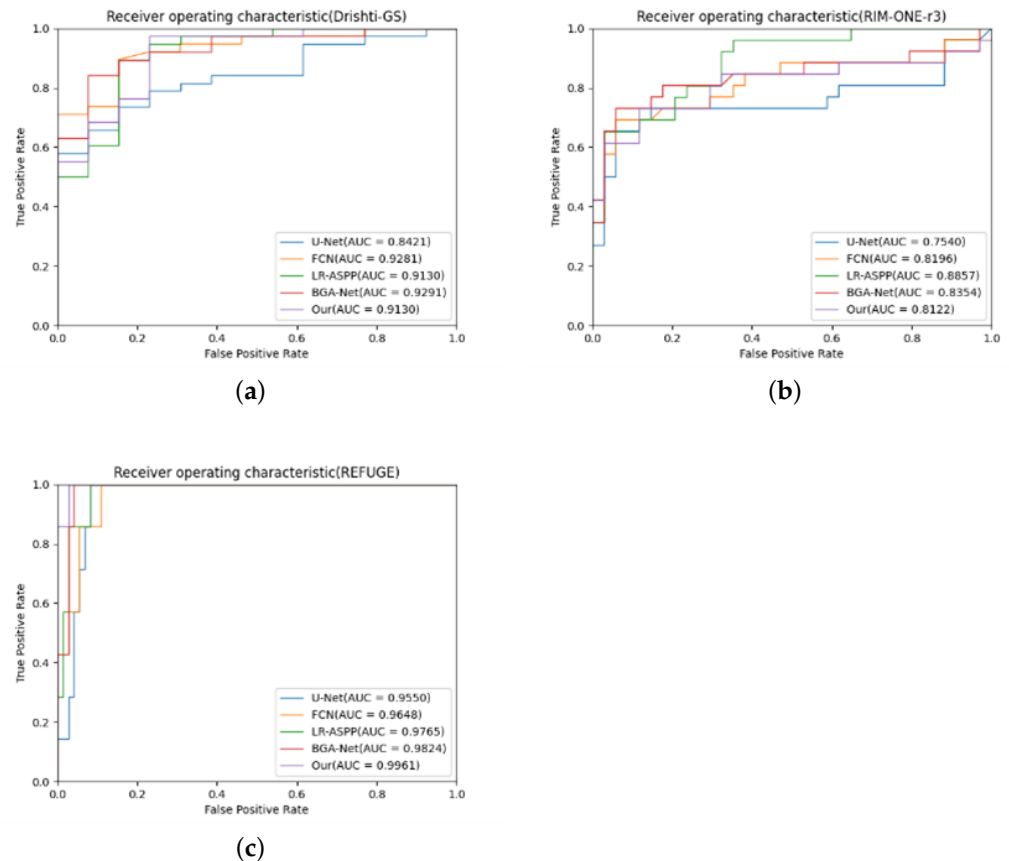
**Figure 6.** Visualization of optic disc and optic cup segmentation results in REFUGE-train dataset.

### 3.5. Model Performance

The receiver operating characteristic (ROC) curve and its corresponding area under the curve (AUC) [35] are used to assess the performance of the algorithm in detecting glaucoma. In general, the higher the AUC, the higher the diagnostic accuracy of the algorithm, indicating a better performance of the algorithm. This paper uses the ROC curve to evaluate the segmentation performance of GlauNet. Figure 7 shows the ROC curves and corresponding AUC values for the three datasets—Drishti-GS, RIM-ONE-r3, and REFUGE-train.

To better analyze the correlation between the mean *CDR* error  $\delta$  and AUC. A line graph of the relationship between  $\delta$  and AUC for the Drishti-GS, RIM-ONE-r3, and REFUGE-train datasets is given in Figure 8. The results show that AUC is not always strictly negatively correlated with  $\delta$ , and can only reflect the correlation between average *CDR* error  $\delta$  and

AUC to a certain extent, but it still has a certain guiding significance. In current glaucoma research, the mean *CDR* error  $\delta$  is still used as an important indicator in the diagnosis of glaucoma.



**Figure 7.** ROC curves for the dataset. (a) is the ROC curve for the Drishti-GS dataset, (b) is the ROC curve for the RIM-ONE-r3 dataset, and (c) is the ROC curve for the REFUGE-train dataset.

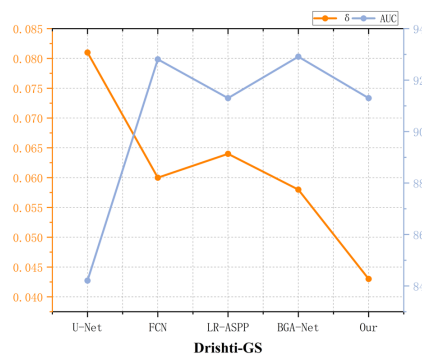
This paper compares the performance of model parameters, required memory, computational complexity, and inference time with some advanced methods in this field. In order to ensure the comparability of the results, the methods in Table 4 were all carried out under the same experimental conditions. The input image is an  $800 \times 800$  RGB image, and the experimental equipment is an NVIDIA GTX 3070 GPU. The results show that the method outperforms other methods on most evaluation metrics while obtaining competitive OD and OC segmentation results.

This paper analyzes the network model size, FLOPs, inference time, and Jaccard on the REFUGE-train dataset. The area of the circle represents the size of the network model, FLOPs, and inference time, respectively.

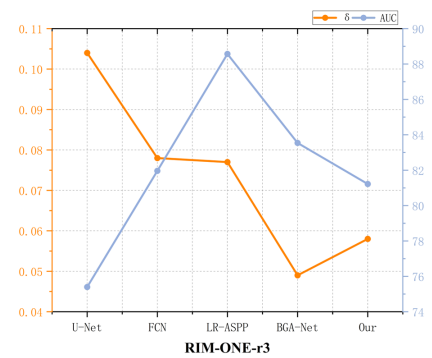
First, the relationship between the size of the network model and the Jaccard accuracy is analyzed, as shown in Figure 9. Our method achieves competitive results with the smallest network model size. The relationship between FLOPs and Jaccard accuracy is then analyzed, as shown in Figure 10. Our method achieves competitive results with minimal FLOPs. Figure 11 analyzes the relationship between inference time and Jaccard accuracy. The inference time of our method is second only to LR-ASPP [28], which is only about half of the other methods while achieving competitive segmentation results.

**Table 4.** Model performance of different methods.

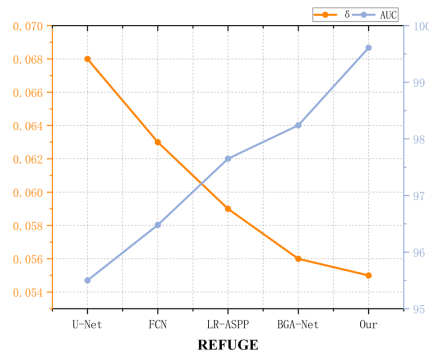
Methods	Parameters	Memory	MAdd	FLOPs	Inference Time
YAO [36]	109.3 M	-	-	78.0 G	-
DDSC-Net [5]	46.3 M	23,012.8 M	924 G	458.6 G	-
FCN-8s [27]	13.1 M	1182 M	217 G	95.8 G	25 ms
UNetFormer [37]	11.7 M	652 M	62.6 G	32.8 G	16 ms
DCUnet [38]	10.2 M	4692 M	381.7 G	173.6 G	33 ms
BGA-Net [3]	5.8 M	1576 M	129 G	64.6 G	22 ms
ISFA [6]	7.0 M	8022 M	200.6 G	98 G	-
U-Net [26]	4.3 M	1857 M	196.4 G	98.3 G	27 ms
LR-ASPP [28]	3.2 M	773 M	9.8G	4.9 G	12 ms
Fast-SCNN [13]	1.1 M	377 M	4.2 G	2.1 G	11 ms
Proposed	0.8 M	327 M	4.9 G	2.5 G	13 ms



(a)



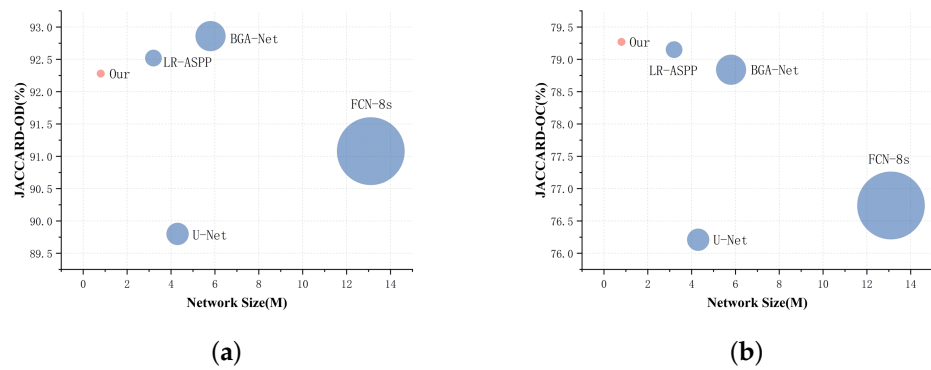
(b)



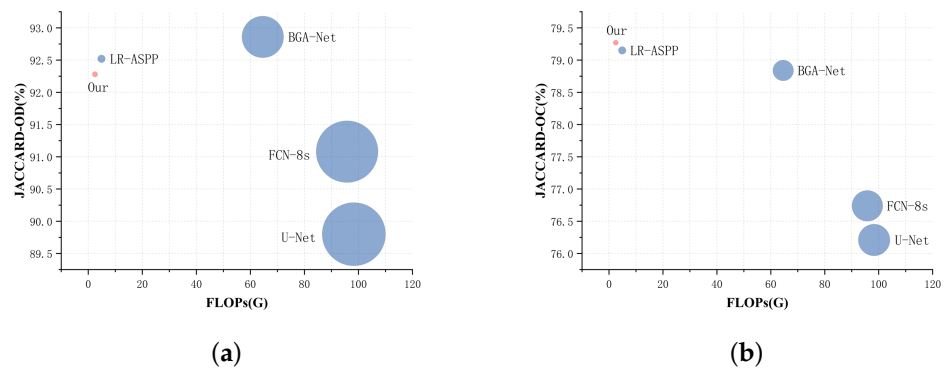
(c)

**Figure 8.** Correlation of average CDR error  $\delta$  and AUC: (a) is the correlation between the average CDR error  $\delta$  and AUC for the Drishti-GS dataset; (b) is the correlation between the average CDR error  $\delta$  and AUC for the RIM-ONE-r3 dataset; (c) is the correlation between the average CDR error  $\delta$  and AUC for the REFUGE-train dataset.

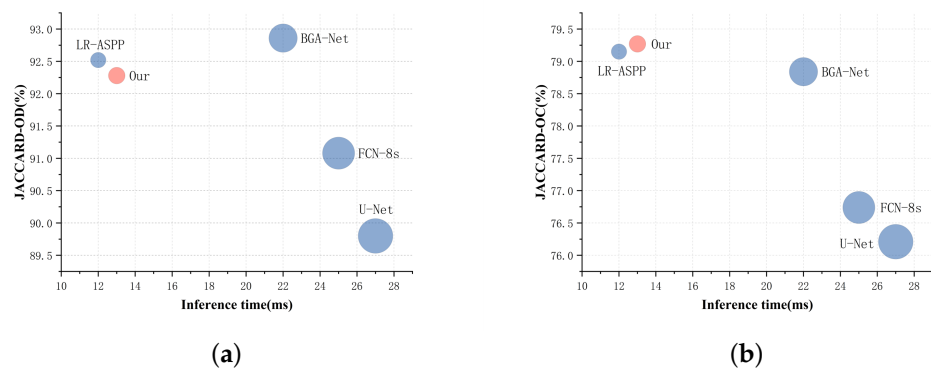




**Figure 9.** Network model size and Jaccard accuracy: (a) is the Jaccard accuracy and network model size of OD; (b) is the Jaccard accuracy and network model size of OC.



**Figure 10.** FLOPs and Jaccard accuracy: (a) is the Jaccard accuracy and FLOPs of OD; (b) is the Jaccard accuracy and FLOPs of OC.



**Figure 11.** Inference time and Jaccard accuracy: (a) is the Jaccard accuracy and inference time of OD; (b) is the Jaccard accuracy and inference time of OC.

#### 4. Ablation Experiments

To verify the effectiveness of the multiscale feature fusion (MFF) module and the boundary feature auxiliary (BFA) module, we used the model with the MFF and BFA modules removed as the baseline to perform ablation experiments on the RIM-ONE-r3 dataset and gave quantitative results and qualitative results in Table 5 and Figure 12. The results demonstrate the effectiveness of the MFF and BFA modules.

The results of ablation experiments show that the MFF module and BFA module can significantly improve the segmentation effect of the optic cup. Figure 13 shows the

entropy map of the OC in baseline, baseline+MFF, and baseline+MFF+BFA (GlauNet). Figure 13 shows that the MFF module can effectively reduce the entropy of the cup boundary prediction map, but there is still some boundary noise. Based on the MFF module, the BFA module further suppresses the boundary noise and highlights the edge structure information of the optic cup. The entropy graph in Figure 13 demonstrates the effectiveness of the MFF module and the BFA module from another perspective.

Table 5. Quantitative results of ablation experiments on RIM-ONE-r3 dataset.

Methods	OD			OC			$\delta$
	DI	Jaccard	SEN	DI	Jaccard	SEN	
Baseline	0.9640	0.9312	0.9645	0.8325	0.7315	0.7931	0.071
Baseline+MFF	0.9609	0.9254	0.9842	0.8515	0.7524	0.8922	0.060
Baseline+MFF+BFA	0.9650	0.9328	0.9809	0.8621	0.7666	0.8823	0.058

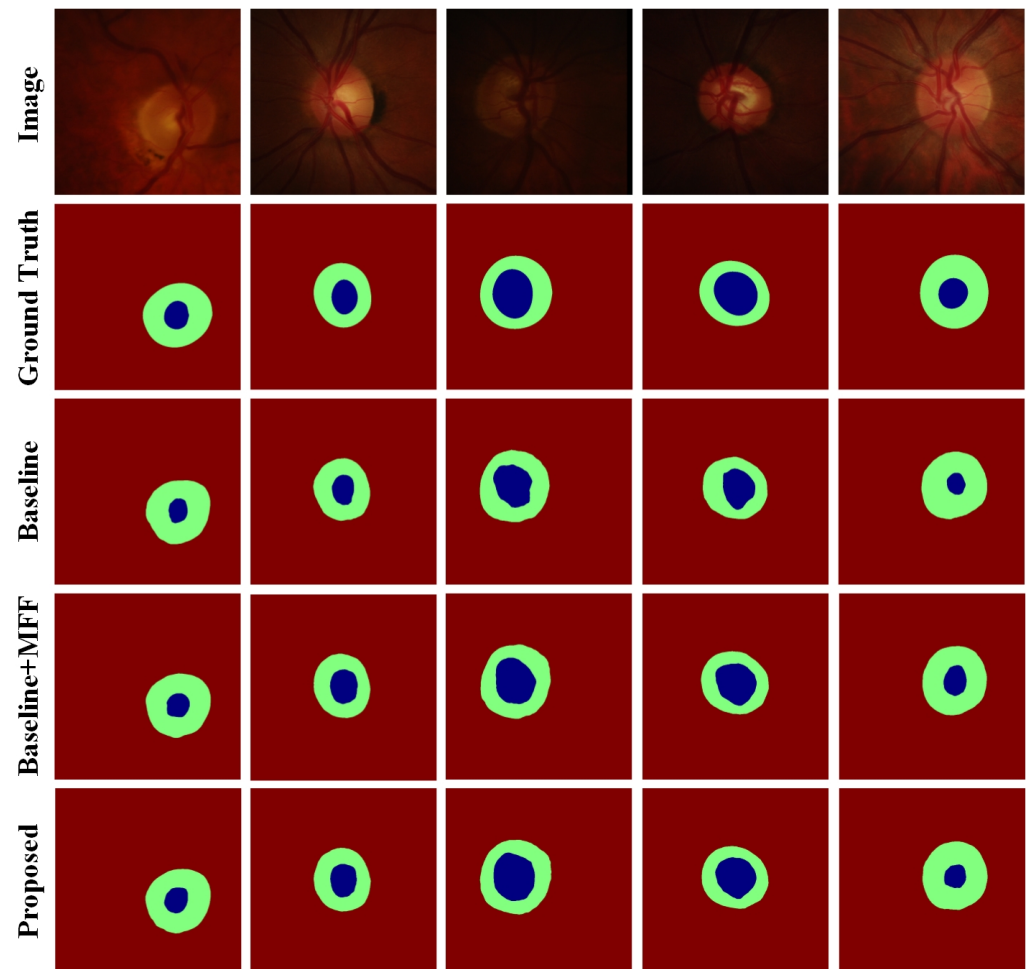


Figure 12. Ablation experiment on RIM-ONE-r3 dataset. Green represents the optic disc and blue represents the optic cup.



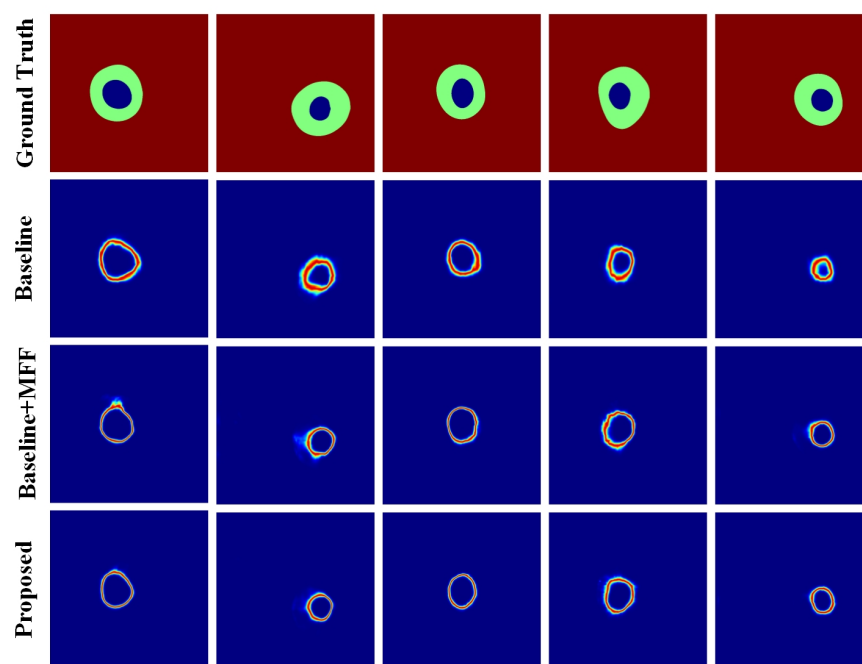


Figure 13. Entropy map of optic cup for ablation experiment of RIM-ONE-r3 dataset.

## 5. Discussion

To meet the performance requirements of mobile devices and edge devices, this paper proposes a simple and efficient fundus-image-segmentation algorithm GlauNet. We experimentally verify the proposed algorithm on Drishti-GS, RIM-ONE-r3, and REFUGE-train public datasets. The performance of the proposed algorithm is evaluated from two aspects: quantitative analysis and qualitative analysis. In this paper, four indices—DI, Jaccard, sensitivity, and *CDR*—are used as the evaluation criteria for the segmentation results. Tables 1–3 show the results of quantitative experiments on the Drishti-GS, RIM-ONE-r3, and REFUGE-train datasets, which show that our algorithm achieves competitive or state-of-the-art segmentation results for each evaluation criterion compared with current state-of-the-art algorithms. Figure 4–6 are partial segmentation renderings randomly selected on the Drishti-GS, RIM-ONE-r3, REFUGE-train datasets, including the region-of-interest images and the corresponding manual annotation maps, U-Net, FCN-8s, the segmentation map of the BGA-Net algorithm, and the algorithm in this paper. The segmentation effect of our proposed algorithm is significantly better than that of U-Net and FCN-8s, with a 51% and 48% improvement in inference time compared with U-Net and FCN-8s, respectively. Our proposed algorithm achieves segmentation results comparable to the hand-labeled graph and BGA-Net algorithms, while the inference time is improved by 40% compared with BGA-Net.

In this paper, the ROC curve was used to evaluate the performance of the proposed algorithm in diagnosing glaucoma, and its ROC curve is shown in Figure 7. Our algorithm achieved a diagnostic accuracy of 91.30, 81.22, and 99.61 on the Drishti-GS, RIM-ONE-r3, and REFUGE-train datasets, respectively. To validate the effectiveness of the MFF and BFA modules in this paper, we conducted ablation experiments on the RIM-ONE-r3 dataset. The quantitative results in Table 5 and the qualitative results in Figure 12 show that the MFF and BFA modules bring a significant improvement in segmentation performance. To verify the computational performance of the proposed algorithm, we compare and analyze the performance of the model with some current state-of-the-art methods. The results in Table 4 and Figures 9–11 show that our algorithm is much smaller than some of the current state-of-the-art algorithms in terms of the number of model parameters, computational complexity, and memory, while achieving competitive or state-of-the-art segmentation results, demonstrating the lightweight nature and efficiency of the proposed algorithm.

However, as the algorithm does not use pretrained weights, our training efficiency is low, requiring 1000 epochs for the Drishti-GS and RIM-ONE-r3 datasets and 300 epochs for the REFUGE-train dataset for the algorithm's loss function to reach convergence.

## 6. Conclusions

In this paper, we propose a lightweight medical-image-segmentation algorithm GlauNet for joint OD and OC segmentation, which consists of a spatial-detail information-extraction module, a contextual-information-extraction module, and a decoding head module. To obtain richer boundary information, an MBF module is proposed, which is beneficial to the segmentation results through experimental comparison. The algorithm has high real-time performance, low algorithm complexity, and a small memory footprint compared with current state-of-the-art algorithms. We conducted extensive experimental comparisons on the Drishti-GS, RIM-ONE-r3, and REFUGE-train datasets and showed that the GlauNet algorithm achieved competitive or better segmentation results than current state-of-the-art methods on the three fundus datasets with model parameters of 0.8M. In the future, we will continue to work on lightweight segmentation algorithms and apply the proposed methods to more medical image segmentation tasks.

**Author Contributions:** Conceptualization, Z.L. (Zhijie Liu) and J.L.; methodology, Z.L. (Zhijie Liu), B.L., and Z.L. (Zhan Li); writing—original draft preparation, Z.L. (Zhijie Liu) and Y.C.; writing—review and editing, J.L. and Z.L. (Zhan Li); visualization, Z.L. (Zhijie Liu); supervision, J.L.; data curation, X.X.; funding acquisition, J.L.; All authors have read and agreed to the published version of the manuscript.

**Funding:** The work was supported by National Natural Science Foundation of China (61962023)

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Publicly available datasets were analyzed in this study. The code can be accessed at <https://github.com/liu1037342030/GlauNet> (accessed on 1 November 2022). These data can be found at: <https://ai.baidu.com/broad/download> (accessed 1 November 2022).

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Akram, M.U.; Tariq, A.; Khalid, S.; Javed, M.Y.; Abbas, S.; Yasin, U.U. Glaucoma detection using novel optic disc localization, hybrid feature set and classification techniques. *Australas. Phys. Eng. Sci. Med.* **2015**, *38*, 643–655. [[CrossRef](#)] [[PubMed](#)]
2. Zhao, X.; Wang, S.; Zhao, J.; Wei, H.; Xiao, M.; Ta, N. Application of an attention u-net incorporating transfer learning for optic disc and cup segmentation. *Signal, Image Video Process.* **2021**, *15*, 913–921. [[CrossRef](#)]
3. Luo, L.; Xue, D.; Pan, F.; Feng, X. Joint optic disc and optic cup segmentation based on boundary prior and adversarial learning. *Int. J. Comput. Assist. Radiol. Surg.* **2021**, *16*, 905–914. [[CrossRef](#)]
4. Wang, S.; Yu, L.; Li, K.; Yang, X.; Fu, C.W.; Heng, P.A. Boundary and entropy-driven adversarial learning for fundus image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Shenzhen, China, 13–17 October 2019; pp. 102–110.
5. Liu, B.; Pan, D.; Song, H. Joint optic disc and cup segmentation based on densely connected depthwise separable convolution deep network. *BMC Med. Imaging* **2021**, *21*, 1–12. [[CrossRef](#)] [[PubMed](#)]
6. Lei, H.; Liu, W.; Xie, H.; Zhao, B.; Yue, G.; Lei, B. Unsupervised domain adaptation based image synthesis and feature alignment for joint optic disc and cup segmentation. *IEEE J. Biomed. Health Inform.* **2021**, *26*, 90–102. [[CrossRef](#)] [[PubMed](#)]
7. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
8. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
9. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Semantic image segmentation with deep convolutional nets and fully connected crfs. *arXiv* **2014**, arXiv:1412.7062.
10. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *40*, 834–848. [[CrossRef](#)]
11. Chen, L.C.; Papandreou, G.; Schroff, F.; Adam, H. Rethinking atrous convolution for semantic image segmentation. *arXiv* **2017**, arXiv:1706.05587.

12. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European conference on computer vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 801–818.
13. Poudel, R.P.; Liwicki, S.; Cipolla, R. Fast-scnn: Fast semantic segmentation network. *arXiv* **2019**, arXiv:1902.04502.
14. Wu, H.; Zhang, J.; Huang, K.; Liang, K.; Yu, Y. Fastfcn: Rethinking dilated convolution in the backbone for semantic segmentation. *arXiv* **2019**, arXiv:1903.11816.
15. Paszke, A.; Chaurasia, A.; Kim, S.; Culurciello, E. Enet: A deep neural network architecture for real-time semantic segmentation. *arXiv* **2016**, arXiv:1606.02147.
16. Yu, C.; Wang, J.; Peng, C.; Gao, C.; Yu, G.; Sang, N. Bisenet: Bilateral segmentation network for real-time semantic segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 325–341.
17. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826.
18. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid scene parsing network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2881–2890.
19. Jiang, T.; Jin, Y.; Liang, T.; Wang, X.; Li, Y. Boundary Corrected Multi-scale Fusion Network for Real-time Semantic Segmentation. *arXiv* **2022**, arXiv:2203.00436.
20. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the International Conference on Machine Learning, PMLR, Lille, France, 7–9 July 2015; pp. 448–456.
21. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4510–4520.
22. Sivaswamy, J.; Krishnadas, S.; Joshi, G.D.; Jain, M.; Tabish, A.U.S. Drishti-gs: Retinal image dataset for optic nerve head (onh) segmentation. In Proceedings of the 2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI), Beijing, China, 29 April–2 May 2014; pp. 53–56.
23. Fumero, F.; Alayón, S.; Sanchez, J.L.; Sigut, J.; Gonzalez-Hernandez, M. RIM-ONE: An open retinal image database for optic nerve evaluation. In Proceedings of the 2011 24th International Symposium on Computer-Based Medical Systems (CBMS), Bristol, UK, 27–30 June 2011; pp. 1–6.
24. Orlando, J.I.; Fu, H.; Breda, J.B.; van Keer, K.; Bathula, D.R.; Diaz-Pinto, A.; Fang, R.; Heng, P.A.; Kim, J.; Lee, J.; et al. Refuge challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs. *Med. Image Anal.* **2020**, *59*, 101570. [[CrossRef](#)] [[PubMed](#)]
25. Fu, H.; Cheng, J.; Xu, Y.; Wong, D.W.K.; Liu, J.; Cao, X. Joint optic disc and cup segmentation based on multi-label deep network and polar transformation. *IEEE Trans. Med. Imaging* **2018**, *37*, 1597–1605. [[CrossRef](#)] [[PubMed](#)]
26. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; pp. 234–241.
27. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
28. Howard, A.; Sandler, M.; Chu, G.; Chen, L.C.; Chen, B.; Tan, M.; Wang, W.; Zhu, Y.; Pang, R.; Vasudevan, V.; et al. Searching for mobilenetv3. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 27 October–2 November 2019; pp. 1314–1324.
29. Tabassum, M.; Khan, T.M.; Arsalan, M.; Naqvi, S.S.; Ahmed, M.; Madni, H.A.; Mirza, J. CDED-Net: Joint segmentation of optic disc and optic cup for glaucoma screening. *IEEE Access* **2020**, *8*, 102733–102747. [[CrossRef](#)]
30. Zilly, J.; Buhmann, J.M.; Mahapatra, D. Glaucoma detection using entropy sampling and ensemble learning for automatic optic cup and disc segmentation. *Comput. Med. Imaging Graph.* **2017**, *55*, 28–41. [[CrossRef](#)]
31. Al-Bander, B.; Williams, B.M.; Al-Nuaimy, W.; Al-Taei, M.A.; Pratt, H.; Zheng, Y. Dense fully convolutional segmentation of the optic disc and cup in colour fundus for glaucoma diagnosis. *Symmetry* **2018**, *10*, 87. [[CrossRef](#)]
32. Gu, Z.; Cheng, J.; Fu, H.; Zhou, K.; Hao, H.; Zhao, Y.; Zhang, T.; Gao, S.; Liu, J. Ce-net: Context encoder network for 2d medical image segmentation. *IEEE Trans. Med. Imaging* **2019**, *38*, 2281–2292. [[CrossRef](#)]
33. Xu, Y.L.; Lu, S.; Li, H.X.; Li, R.R. Mixed maximum loss design for optic disc and optic cup segmentation with deep learning from imbalanced samples. *Sensors* **2019**, *19*, 4401. [[CrossRef](#)]
34. Yu, S.; Xiao, D.; Frost, S.; Kanagasigam, Y. Robust optic disc and cup segmentation with deep learning for glaucoma detection. *Comput. Med. Imaging Graph.* **2019**, *74*, 61–71. [[CrossRef](#)] [[PubMed](#)]
35. Fawcett, T. An introduction to ROC analysis. *Pattern Recognit. Lett.* **2006**, *27*, 861–874. [[CrossRef](#)]
36. Zheng, Y.; Zhang, X.; Xu, X.; Tian, Z.; Du, S. Deep level set method for optic disc and cup segmentation on fundus images. *Biomed. Opt. Express* **2021**, *12*, 6969–6983. [[CrossRef](#)]
37. Wang, L.; Li, R.; Zhang, C.; Fang, S.; Duan, C.; Meng, X.; Atkinson, P.M. UNetFormer: A UNet-like transformer for efficient semantic segmentation of remote sensing urban scene imagery. *ISPRS J. Photogramm. Remote. Sens.* **2022**, *190*, 196–214. [[CrossRef](#)]
38. Lou, A.; Guan, S.; Loew, M. DC-UNet: rethinking the U-Net architecture with dual channel efficient CNN for medical image segmentation. In *Medical Imaging 2021: Image Processing*; SPIE: Bellingham, WA, USA, 2021; pp. 758–768.