IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON DEEP LEARNING ALGORITHMS FOR INTERNET OF MEDICAL THINGS

Received June 29, 2020, accepted July 10, 2020, date of publication July 15, 2020, date of current version July 29, 2020.

*Digital Object Identifier 10.1109/ACCESS.2020.3009442*

# A Novel Adaptive Weighted Loss Design in Adversarial Learning for Retinal Nerve Fiber Layer Defect Segmentation

**SHUAI LU[1], MAN HU[2,3], RUIRUI LI[4], (Member, IEEE), AND YONGLI XU[1]**

[1]Department of Mathematics, Beijing University of Chemical Technology, Beijing 100029, China
[2]Department of Ophthalmology, Beijing Children's Hospital, Capital Medical University, Beijing 100045, China
[3]Beijing Tongren Eye Center, Beijing Tongren Hospital, Capital Medical University, Beijing 100045, China
[4]College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China

Corresponding author: Yongli Xu (xuyongli2312@sina.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 11571031 and Grant U1830107, and in part by the National Key Research and Development Project under Grant 2019YFB2006602.

**ABSTRACT** Glaucoma is a chronic eye disease that can cause permanent visual loss and is difficult to detect early. Retinal nerve fiber layer defect (RNFLD) is clinical evidence for the diagnosis of glaucoma. Classical deep learning based methods can be used to segment RNFLD from fundus images. However, the segmentation results of these methods do not have the specific geometry of RNFLD, and the segmentation errors of fundus images with special styles are large. In this paper, we present a novel conditional adversarial shuffle U-shaped network (CASU-Net) to segment RNFLD, which consists of a generator and a discriminator. For the generator, a mixed loss is designed, which consists of an adaptive weighted segmentation loss and an adversarial loss. This adaptive weighted segmentation loss can balance the segmentation accuracy of the target and background region, and assign more attention to the hard samples, thus ensuring the consistent improvement of the segmentation accuracy of all fundus images. The adversarial loss not only helps to improve the pixel-wise segmentation accuracy but also makes the geometry of the RNFLD segmentation closer to the ground truth. In addition, in the generator, a shuffle module was designed to fully mine the information of all channels to improve the feature extraction capability of the model. The proposed CASU-Net is verified on a RNFLD dataset from Beijing Tongren Hospital. The experiments show that the CASU-Net achieves state-of-the-art results on this dataset.

**INDEX TERMS** Glaucoma, retinal nerve fiber layer defect segmentation, deep learning.

## I. INTRODUCTION

Glaucoma is the second causing blindness disease in the world, and it will cause irreversible visual loss to patients. By 2040, the number of glaucoma patients will reach 110 Million [1]. Patients usually have no obvious symptoms at the beginning of glaucoma until the visual loss appears [2]. If initial glaucoma patients can be found in the glaucoma screening, prompt treatments can be adopted to decrease the vision loss effectively. Therefore, early diagnosis and treatment of glaucoma are important to protect the vision of the patients. Generally, the diagnosis of glaucoma requires rich clinical experience. However, a small number of glaucoma professional physicians cannot meet the needs of large-scale glaucoma screening. Therefore, there is an urgent need for automatic and accurate diagnosis methods for glaucoma.

Optical coherence tomography (OCT) and color fundus images are two methods of glaucoma screening [3]–[6]. Because OCT is expensive, and color fundus images are highly efficient and economical, color fundus images are more suitable for large-scale initial screening work [6]. If the retinal nerve fiber layer defect (RNFLD) is detected in the color fundus image, it can be used as an indicator for glaucoma diagnosis. In the color fundus image, the optic disc is a bright yellow oval region, and the RNFLD is a wedge-shaped dark region close to the optic disc [7], as shown in Figure. 1.

A number of works have been proposed to segment the RNFLD. These methods mainly include image segmentation methods based on traditional techniques and segmentation

The associate editor coordinating the review of this manuscript and approving it for publication was Wei Wei.

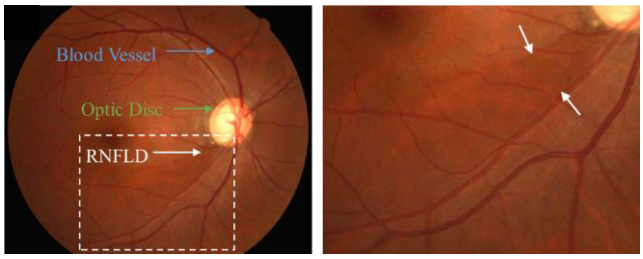


**FIGURE 1.** An example of a color fundus image with retinal nerve fiber layer defect (RNFLD). The green arrow points to the optic disc region, the white arrow points to the RNFLD, and the blue arrow points to the blood vessels in the fundus image.

methods based on deep learning. However, traditional techniques are mainly based on hand-crafted features, which lack effective representations and are susceptible to low contrast quality. Although existing deep learning methods of RNFLD segmentation can automatically extract features, they also have three main problems:

1) The RNFLD segmentation obtained by the classic CNNs cannot be trusted by doctors, because the segmentation results neither have a specific morphology nor conform to the intuitive perception of the doctor. In contrast, RNFLD marked by the doctor generally has a wedge-shaped geometry and smooth boundaries. Therefore, a method that meets the morphological characteristics of true RNFLD needs to be proposed.
2) The prediction accuracy of a few fundus images with special styles is not high enough. In the case in which there are multiple pieces of RNFLD or the contrast between the inner and outer areas of the RNFLD outline is not obvious, the RNFLD extracted by the above deep learning methods have large errors compared to the RNFLD marked by glaucoma experts.
3) The information in red and blue channels of color fundus images is not fully exploited, since the RNFLD information is mainly distributed in the green channel of the color fundus images, and the green channel is easy to completely dominate the training of the CNN. A method that can fully mine feature information needs to be proposed.

To solve the above three problems in segmenting RNFLD in fundus images, we proposed a novel method based on adversarial learning. The main contributions of this work include:

1) We designed a novel conditional adversarial shuffle U-shaped network (CASU-Net), which consists of a generator and a discriminator. The discriminator is used to supervise the segmentation results of the generator. This design not only improves the pixel-wise segmentation accuracy, but also makes the geometry of the target area obtained by the generator closer to the ground truth, which helps to strengthen the doctor's trust in the segmentation result.
2) In the CASU-Net, an adaptive weighted (AW) segmentation loss was designed for the generator. This AW loss can adaptively adjust the weight of each fundus image, so that the fundus image with large segmentation error gets more attention in the training. The AW loss tends to improve the segmentation accuracy of all fundus images simultaneously during the entire training process, instead of improving the majority in the early stage and then improving the remaining few in the later stage, so as to avoid overfitting for fundus images with special styles.
3) In the CASU-Net, a channel shuffle module was designed for the generator. Through feature rearrangement and random inactivation of connections between feature channels, the shuffle module can not only focus on the information of the green channel, but also fully explores the information of the red and blue channels, thus enhancing the feature extraction capability of the model.
4) We evaluate the effectiveness and generalization capability of the proposed CASU-Net and existing methods on a RNFLD dataset from Beijing Tongren Hospital. The proposed CASU-Net achieves state-of-the-art segmentation performance.

The rest of this paper is organized as follows. We first review techniques related to the RNFLD segmentation in the second part. The framework of CASU-Net is described in the third part, and the experimental setup and results are presented in the fourth part. We further obtain the conclusion in the fifth part.

## II. RELATED WORK

Some traditional methods have been proposed for the detection of RNFLD in the fundus image. Muramatsu *et al.* applied the Gabor filters to the enhancement of RNFLDs after the removal of the major blood vessels. By using LDA and ANN classifier, true RNFLDs were identified from the darker bandlike regions [8]. In [9], Oh *et al.* applied Hough transformation to detect the candidates after illumination correction and polar transformation. Knowledge-based rules were used to reduce false detection candidates for RNFLD. Lamani *et al.* proposed a method based on texture and fractal description for glaucomatous retina detection and used a support vector machine classifier for classification [10]. In [11], Panda *et al.* classified RNFLD boundary pixels using random forest with cost-effective red-free fundus images. However, there is a large gap between the RNFLDs predicted by these methods and those marked by ophthalmologists.

In recent years, deep learning has developed rapidly in the field of computer vision. It has made great progress in image classification [12]–[18], object detection [19], [20] and image segmentation [22]–[27]. Compared with traditional methods, deep neural networks can automatically extract features from the input data and achieve higher accuracy. There have also been breakthroughs in the detection of RNFLDs. Panda *et al.* [28] proposed a deep learning method to detect RNFLD boundaries. In this method, the visibility of the RNFLD region was further enhanced by contrast-limited adaptive histogram equalization after the removal of

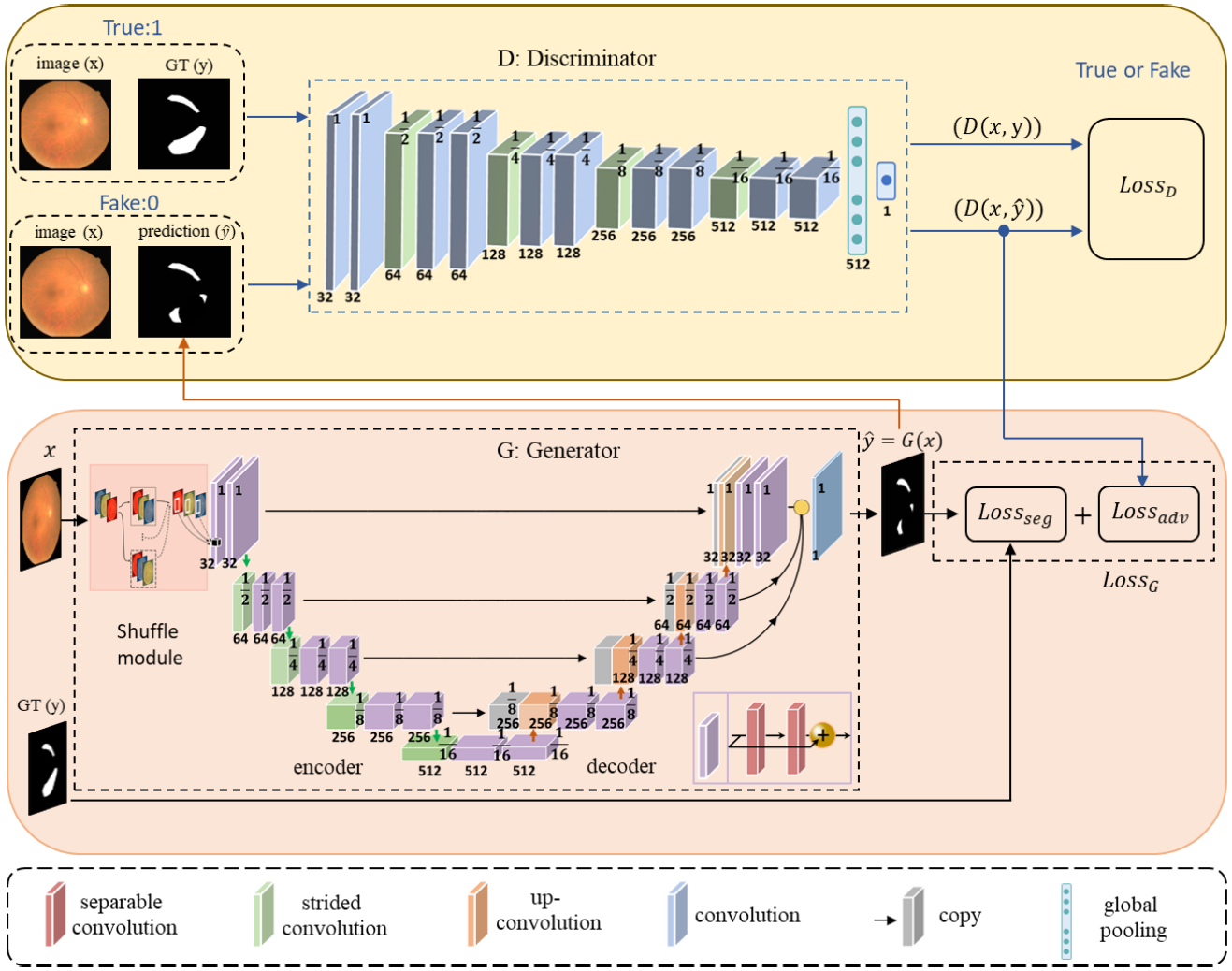


**FIGURE 2.** Structure of the proposed CASU-Net. *G* represents generator, *D* represents discriminator, *x* represents a retinal fundus image, *y* represents ground truth, and $D(x, y)$ and $D(x, G(x))$ represent the probability of true samples and false samples predicted as ground truth by the discriminator, respectively. $Loss_D$ represents the discriminator loss, $Loss_{adv}$ represents the adversarial loss of the generator, and $Loss_{seg}$ represents the segmentation loss of the generator.

blood vessels. The RNFLD boundary pixels were selected for training and testing. A patch-based deep convolutional neural network (DCNN) was initially used to detect RNFLD boundaries. The detected RNFLD boundary pixels were fitted into lines by the random sample consensus algorithm. In [29], Watanabe *et al.* proposed a DCNN with deconvolutional layers to detect RNFLD. DCNN training was carried out using different input image sets, such as original images of abnormal cases, original images of both normal and abnormal cases, and transformed half images.

## III. METHODS

We propose a novel CASU-Net framework for the segmentation of RNFLD in fundus images. The CASU-Net consists of a generator and a discriminator, as shown in Figure. 2. The generator is a U-shaped convolutional neural network whose input is the fundus image ($x$) with three channels, and the output is the probability map ($y$) of the RNFLD segmentation. The overall loss function of the generator $Loss_G$ includes an AW segmentation loss $Loss_{seg}$ and an adversarial loss $Loss_{adv}$.The parameters of the generator network are optimized by minimizing $Loss_G$. The discriminator is a convolutional classification network. The input of the discriminator network consists of the fundus image ($x$) and the RNFLD segmentation probability map ($y$ or $G(x)$). Herein, the RNFLD segmentation probability map is divided into ground truth ($y$) annotated by glaucoma experts and the probability map ($\hat{y} = G(x)$) generated by the generator network. The output of the discriminator is the probability of the network classifying the input as ground truth. The loss function of the discriminator $Loss_D$ can be calculated based on the output, and parameters of the discriminator network are optimized by minimizing $Loss_D$.

During training, the optimization and update of parameters of generator and discriminator are implemented alternately. Optimizing the parameters of the generator network can

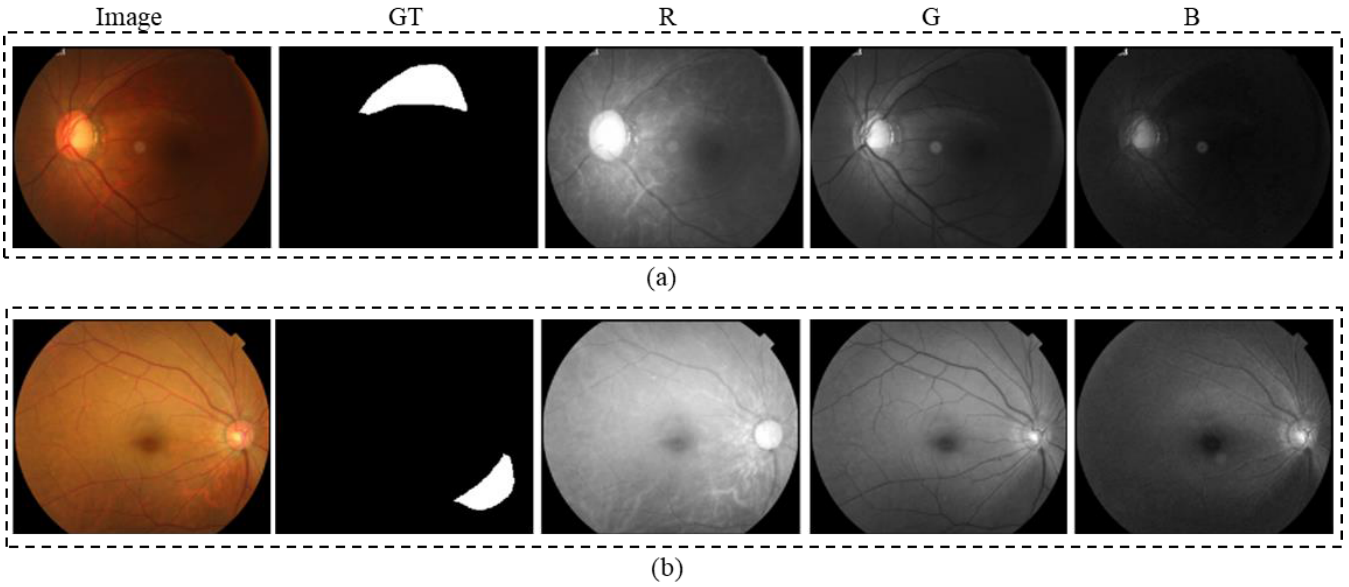


**FIGURE 3.** Examples of the pixel intensity distribution in the three channels of RGB. Image represents the color fundus image, GT represents the RNFLD area marked by the doctor, and R, G, and B represent the pixel intensity distribution of the fundus image on the three channels of red, green, and blue, respectively.

obtain segmentation with higher accuracy, and it makes the discriminator difficult to distinguish the source of the RNFLD segmentation probability map. Optimizing the parameters of the discriminator network can improve the discriminator's ability to distinguish between the ground truth and the RNFLD segmentation generated by the generator network. In the process of alternate optimization and update of generator parameters and discriminator parameters, the performance of the discriminator network and the generator network are both enhanced. Finally, RNFLD segmentation results with high segmentation accuracy and highly consistent with the geometry of ground truth is obtained.

### A. DISCRIMINATIVE NETWORK

The discriminator network of the proposed CASU-Net is a classification network, which consists of ten convolutional layers, four pooling layers, and a global pooling layer. The convolutional layers are composed of $3 \times 3$ convolution kernels. Each convolution layer is followed by a rectified linear unit (ReLU) activation function. By using the padding operation, the convolutional operation does not change the size of the feature map. After the feature map goes through the pooling layer, the height and width of the feature map are reduced to half of the original size. In addition, we use the global pooling layer instead of the conventional fully-connected layer. The network structure is shown in Figure. 2. The input of the discriminator network is a four-channel structure composed of the fundus image ($x$) and the RNFLD segmentation probability map ($y$ or $G(x)$). Herein, the sample labeled by the glaucoma expert is represented as ($x$, $y$), and the sample generated by the generator is represented as ($x$, $G$ ($x$)). The sizes of $x$, $y$, and $G(x)$ are ($H$, $W$, 3), ($H$, $W$, 1), and ($H$, $W$, 1), respectively. Here, $H$ and $W$ represent the height and width of the feature, respectively. The output of discriminator is a real value mapped by Sigmoid function to [0, 1], which means that the discriminator network judges the input as the probability of ground truth. The loss function of the discriminator is:

$$Loss_D = -E_{x,y\sim p_{data(x,y)}}\left[logD\left(x,y\right)\right] - E_{x\sim p_{data(x)}}\left[log\left(1-D\left(x,G\left(x\right)\right)\right)\right] \quad (1)$$

During discriminator training, the parameters of generator $G$ remain unchanged, and the parameters of discriminator $D$ are optimized and updated by minimizing the loss function $Loss_D$ of discriminator. The optimization objective of the discriminator is to distinguish between the RNFLD probability map generated by the generator and the ground truth of RNFLD.

### B. GENERATIVE NETWORK

The generator of the proposed CASU-Net is an end-to-end shuffle U-shaped network (SU-Net), which consists of three components, as shown in Figure. 2. The first part is the shuffle module, which is mainly used to enhance the generalization ability of the network. The second part is an encoder network and a decoder network, which are used to generate multi-level representations and the final prediction. The third part is a mixed loss, which is used to optimize the generator.

#### 1) SHUFFLE MODULE

Figure. 3. shows two fundus images and their pixel intensity distribution in the three channels of RGB. As can be seen from Figure. 3, the pixel intensity distribution of the three channels of RGB in an image has strong consistency and certain difference. For most fundus images, the RNFLD is more obvious on the green channel, as shown in Figure. 3 (a); for a few fundus images, the RNFLD is more obvious

on the blue or red channel, as shown in Figure. 3 (b). Due to the higher correlation between the green channel and the RNFLD, the parameter update of the traditional CNN will be dominated by the information of the green channel and ignore the information of the blue and red channels. In order to avoid the network from over-reliance on green channels, we designed a shuffle module to mine the overall correlation information and detail difference information of the three channels at the same time.

In the shuffle module, the order of three channels in RGB is randomly changed to achieve reordering. Let $X = [x_1, x_2, x_3]$ be the original feature map of a fundus image. $X$ is transformed into $\tilde{X}$ by a feature rearrangement with a probability of 0.5. Here, $\tilde{X}$ is defined as $\tilde{X} = \left[\tilde{x}_1, \tilde{x}_2, \tilde{x}_3\right] = \left(\sum_i e_{i1}x_i, \sum_i e_{i2}x_i, \sum_i e_{i3}x_i\right)$, where $e_{ij} \in \{0, 1\}$ is a random binary weight and satisfies the following condition: $\sum_i e_{ij} = 1, \sum_j e_{ij} = 1 (i = 1, 2, 3; j = 1, 2, 3)$. At the same time, we are inspired by dropout [30], connections between feature channels of the shuffle module and feature channels of the encoder network are randomly deactivated according to a certain probability. Let $Y = [y_1, y_2, \ldots, y_{n-1}, y_n]$ represent a feature vector of the first layer in the encoder network and n represent the number of features of this layer. Here, the formula of $y_j$ is expressed as follows: $y_j = f\left(\sum_i (w_{ij}\tilde{x}_i r_i + b_j)\right)$, where $w_{ij}$ is the weight of the convolution kernel, $r_i$ satisfies the Bernoulli distribution, $b_j$ is the bias of the j-th feature, and $f$ is the ReLU activation function. When there holds $\sum_i r_i \geq 2 (i = 1, 2, 3)$, $r_i$ can be used for network training, otherwise $r_i$ is generated randomly again. The shuffle module only works during the training process and not during the prediction process.

### 2) FEATURE ENCODER AND DECODER MODULE

In our work, we modify the U-shape convolutional network (U-Net) in [31] as the main part of the proposed SU-Net. The modified U-Net serves as the base model in our paper. Baseline contains an encoder network and a decoder network similar to U-Net. The last features of the decoder are processed by $1 \times 1$ convolution and sigmoid function operations to obtain the prediction of the RNFLD. Compared with U-Net, we make the following improvements in the feature encoder network and decoder network. Our baseline replaces conventional convolutions and pooling layers with deep separable convolutions [32] and strided convolutions, respectively. Each separable convolution is followed by a ReLU activation function and a batch normalization. Skip connections are introduced between two adjacent separable convolutions for residual correction. Second, multiple layers of information are combined for the final feature prediction.

### 3) ADVERSARIAL LOSS AND ADAPTIVE WEIGHTED LOSS

In order to balance the background and target area, and optimize the network from both the pixel level and the picture level, we propose the mixed loss function of the generator ($Loss_G$), which consists of an adversarial loss and a novel adaptive weighted segmentation loss:

$$Loss_G = Loss_{adv} + \lambda L_{seg}(G) \tag{2}$$

In the above expression, $\lambda$ is a parameter used to balance $Loss_{adv}$ and $L_{seg}$, and the adversarial loss is defined based on predictions by the discriminator:

$$Loss_{adv} = E_{x \sim p_{data(x)}}\left[\log\left(1 - D\left(x, G\left(x\right)\right)\right)\right] \tag{3}$$

And, the adaptive weighted segmentation loss $L_{seg}$ is defined based on the difference in pixel level between the probability map generated by the generator and ground truth. In the expression of adversarial loss $Loss_{adv}$, $1 - D\left(x, G\left(x\right)\right)$ represents the probability that discriminator judges $G\left(x\right)$ as a fake sample. The network parameters of the generator are adjusted by minimizing $Loss_{adv}$. The segmentation loss $L_{seg}\left(G\right)$ is composed of a RNFLD segmentation loss and a background segmentation loss. The background and the RNFLD is defined as class 1 and 2, respectively. First of all, the true positive, false negative and false positive of the two classes are defined for a fundus image: $TP_k^i = \sum_{j=1}^N p_j^i(k) g_j^i(k)$, $FN_k^i = \sum_{j=1}^N (1 - p_j^i(k)) g_j^i(k)$, and $FP_k^i = \sum_{j=1}^N p_j^i(k)(1 - g_j^i(k))$. Here, $k$ represents the label of the object ($k = 1, 2$), i represents the sequence number of the fundus image, j is the sequence number of the pixel, and N represents the total number of pixels in a fundus image. And, $p_j^i(k) \in [0, 1]$ represents the probability that the j-th pixel of the i-th fundus image predicted by the generator network belongs to the objective region of the k-th class feature; $g_j^i(k) \in \{0, 1\}$ represents the true label of the j-th pixel of the i-th fundus image of the k-th class feature. Based on the above definitions, the Dice value of the k-th class feature of the i-th fundus image is defined as:

$$\mathrm{DE}_k^i = \frac{2TP_k^i}{2TP_k^i + \alpha_k FN_k^i + \beta_k FP_k^i + \epsilon_1}. \tag{4}$$

Considering that there is no RNFLD in some fundus images, it is likely to appear that $TP_2^i$, $FN_2^i$ and $FP_2^i$ are all 0. In order to avoid the situation where the denominator is 0 in the expression of $\mathrm{DE}_k^i$, we add a small value $\epsilon_1$ to its denominator. Herein, $\alpha_k$ and $\beta_k$ are penalties for false negatives and false positives for class k. The proposed $\mathrm{DE}_k^i$ is an improved Dice coefficient index, which is used to describe the similarity between the prediction result of the k-th feature of the i-th fundus images and the corresponding ground truth. Furthermore, we can define the adaptive weighted segmentation loss of the generator:

$$L_{seg} = -\frac{1}{m}\sum_{i=1}^{m}\sum_{k=1}^{2} w_k^i \left(1 + \theta\left(1 - \mathrm{DE}_k^i\right)^{\gamma}\right) \times ln\left(\mathrm{DE}_k^i + \epsilon_2\right) \tag{5}$$

where $w_k^i$ represents the weight of the k-th class feature of the i-th sample. When there is no true RNFLD in a fundus image, set $w_1^i = 1, w_2^i = 0$, otherwise set $w_1^i = w_2^i = \frac{1}{2}$. In the

expression of $L_{seg}(G)$, $(1+\theta(1-\mathrm{DE}_k^i)^\gamma)$ is a novel adaptive weighted method for sample weights. Here, $\theta$ and $\gamma$ are parameters for balancing weights of hard and easy samples. In the training process of traditional deep neural networks, simple samples dominate the updating of network parameters, while hard samples cannot receive enough attention. The adaptive weighted loss can automatically increase the weight of hard samples while reducing the weight of easy samples, thereby uniformly improving the prediction accuracy of all samples.

#### 4) COMPARISON WITH RELATED WORKS

##### *a: COMPARISON WITH CONDITIONAL ADVERSARIAL NETWORK*

The proposed method has the general architecture of a conditional adversarial network. The loss functions of the discriminator and generator can be unified as the following objective function:

$$G^* = arg\ min_G max_D \left[E_{x,y\sim p_{data(x,y)}}\left[\log D(x,y)\right] + E_{x\sim p_{data(x)}}\left[log\left(1-D\left(x,G\left(x\right)\right)\right)\right]+\lambda L_{seg}\left(G\right)\right] \quad (6)$$

In this objective function, the objective of discriminator D is to maximize the objective function to accurately distinguish the RNFLD probability map generated by the generator and ground truth, and the objective of generator G is to minimize the objective function to generate RNFLD probability map that is indistinguishable by the discriminator and reduce the pixel-level deviation between the RNFLD probability map and ground truth. The training of discriminator and generator is performed alternately. The optimization process of the model is as follows: first, the parameters in the discriminator network and generator network are assigned using a random initialization method. Then, the network parameters of the discriminator and the generator alternately perform n rounds of optimization. In each round of optimization, the network parameters of the discriminator are optimized $k_1$ times, and then the network parameters of the generator are optimized $k_2$ times. The detailed optimization process is shown in Algorithm 1.

Compared with the classic conditional adversarial network [33], the proposed CASU-Net has the following advantages: 1) The objective function contains not only the objective function of conditional GAN, but also the segmentation loss of the RNFLD. It pays attention to the difference between the prediction result of RNFLD and ground truth in the overall geometry and pixel level. 2) A strategy is designed to adaptively adjust the sample weight based on the segmentation accuracy of samples. As a result, the fundus image difficult to segment is focused more by the model. The final model is effectively improved in predicting both the hard and easy samples. 3) The segmentation loss $L_{seg}(G)$ takes the segmentation performances of the RNFLD region and the background region as two optimization objectives to solve the absence of RNFLD and the imbalance data of pixel-wise segmentation.

##### *b: COMPARISON WITH FOCAL LOSS*

The focal loss is a loss function designed to prevent easy samples from dominating CNN training [34]. It can be applied in the field of target detection and image segmentation. For image segmentation, the formula for focal loss can be described as follows:

$$L_{Focal} = -\frac{1}{mN}\sum_{i=1}^{m}\sum_{j}^{N}\left[(1-p_j^i)^\gamma g_j^i \log\left(p_j^i\right) + \left(p_j^i\right)^\gamma (1-g_j^i)\log(1-p_j^i)\right] \quad (7)$$

where $\gamma$ represents the focusing parameter, i represents the sequence number of the fundus image, j represents the sequence number of the pixel, N represents the total number of pixels in a fundus image, m represents the number of fundus images used in one iteration, $p_j^i(k) \in [0, 1]$ represents the probability that the j-th pixel of the i-th fundus image predicted by the CNN to the RNFLD region, and $g_j^i \in \{0, 1\}$ represents the RNFLD label of the j-th pixel of the i-th fundus image.

The basic design idea of focal loss and the proposed AW loss are the same: both give more attention to difficult objects in the process of training the network. However, there are obvious differences between the two loss functions. First of all, in the calculation of the focal loss, the weight is given to pixels, and more attention is paid to pixels with large prediction deviations during the training process; but in the calculation of the AW loss, the weight is given to images, and more attention to is paid to images with large prediction deviations. In addition, in order to overcome the problem of unbalanced pixel distribution in the target area and the background area, we designed a weighting strategy for the segmentation accuracy of the two types of areas in AW loss.

## IV. EXPERIMENTS

### A. DATASETS AND EVALUATION METHOD

In the experiment of this paper, we evaluated the proposed and compared algorithms in a dataset from Beijing Tongren hospital, which included 474 fundus images with a resolution of 1924 × 1924. A glaucoma expert judged whether there was RNFLD in the fundus image, and manually marks the boundary line of RNFLD for the fundus image with RNFLD. We subtract the fundus image from the labeled fundus image according to the pixel position to obtain the RNFLD boundary. The boundary line divides a fundus image into multiple areas, the largest one is marked as the background, and the remaining areas are marked as RNFLD. There were 223 fundus images with RNFLD and 251 fundus images without RNFLD.

In order to comprehensively evaluate performances of the proposed and compared methods, we used the following evaluation metrics to compare different methods, including F-score, sensitivity, specificity, the Receiver Operating Characteristic (ROC) curves and the area under ROC curve (AUC). Here, F-score, sensitivity, and specificity are defined as $F_{\text{score}} = \frac{2TP}{2TP+FN+FP}$, sensitivity $= \frac{TP}{TP+FN}$,

**Algorithm 1** Training of CASU-Net. $k_1$, $k_2$ Are the Number of Optimization Steps in Discriminator and Generator, Respectively, $\lambda$ Is Used to Balance the Two Parts of Loss, $n$ Is the Number of Training Iterations

Randomly initialize critic parameter $\theta_d$ and generator's parameter $\theta_g$
for n in training iterations do
for $k_1$ steps do

- Sample minibatch of m retinal images samples $\{(x_1, y_1), \ldots, (x_m, y_m))$ from retinal images and corresponding ground truth dataset
- Sample minibatch of m retinal images samples $\{x_1, \ldots, x_m\}$ from retinal images dataset
- Optimize the convolution kernel parameter in the discriminator by ascending its gradient:

$$\nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^{m} \left[ logD\left(x_i, y_i\right) + \log\left(1 - D\left(x_i, G\left(x_i\right)\right)\right)\right]$$

end for
for $k_2$ steps do

- Sample minibatch of m retinal images samples $\{(x_1, y_1), \ldots, (x_m, y_m))$ from retinal images and corresponding ground truth dataset
- Optimize the convolution kernel parameter in the generator by descending its gradient:

$$\nabla_{\theta_g} \frac{1}{m} \sum_{i=1}^{m} \left[ \log\left(1 - D\left(x_i, G\left(x_i\right)\right)\right) + \lambda L_{seg}\left(\mathrm{G}\left(x_i\right), \mathrm{y_i}\right)\right]$$

end for
end for

specificity $= \frac{TN}{TN+FP}$, where *TP*, *FN*, and *FP* are the true positives, false negatives, and false positives. In addition, considering that the area of the objective segmentation is much smaller than that of the background area, we also used Mean Intersection over Union (MIoU) and Mean Average Precision (MAP) [35] to evaluate different methods. Herein, MAP and MIoU are defined as: MAP $= \frac{1}{K}\sum_{i=1}^{K} \frac{p_{ii}}{\sum_{j=1}^{K} p_{ij}}$, and MIoU $= \frac{1}{K}\sum_{i=1}^{K} \frac{p_{ii}}{\sum_{j=1}^{K} p_{ij} + \sum_{j=1}^{K} p_{ji} - p_{ii}}$. Here, $p_{ij}$ represents the number of pixels that belong to the i-th class and are predicted into the j-th class, and $p_{ii}$ represents the number of pixels that belong to the i-th class and are predicted into the i-th class.

We used the two-fold cross-validation method to evaluate the performance of the proposed method and the compared methods. The fundus image dataset was randomly divided into two folds. One fold contained 237 fundus images, which included 111 fundus images with RNFLD. The other fold contained 237 fundus images, which included 112 fundus images with RNFLD. The average value of the two-fold cross-validation was used as the evaluation metric. Five replicate experiments were performed on the two-fold cross-validation. Finally, the average value of five repeated experiments was used as the metric evaluation value.

### B. EXPERIMENTAL SETUP

In the experiments, both the original fundus image and the RNFLD segmentation image marked by the doctor were compressed to a resolution of 256 × 256 pixels. Standard techniques were used to perform data enhancement on the training set, such as random rotation, flipping,

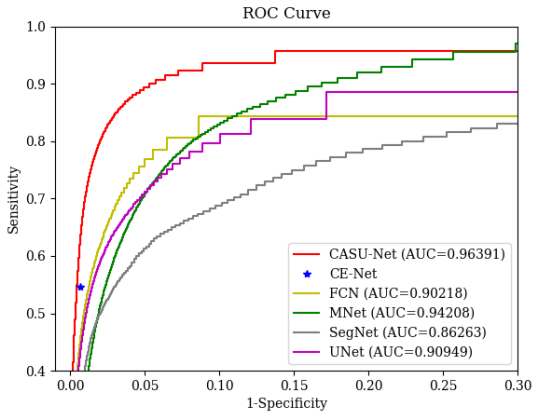


**FIGURE 4.** The ROC curves of different methods on the RNFLD dataset.

and translation. The proposed CASU-Net model was implemented based on the Keras framework with Tensorflow backend. During the network training process, Adam optimization method was used to optimize the model parameters. The initial learning rate was set to $10^{-4}$. For the parameters of the shuffle module, $p$ was set to be 0.8. The parameters in the loss function were set as follows: $\alpha_1 = 4$, $\alpha_2 = 1, \beta_1 = 1, \beta_2 = 1, \epsilon_1 = 10^{-7}, \epsilon_2 = 10^{-7}, \lambda = 10, \theta = 20$, and $\gamma = 2$. When the parameters are set as $\alpha_1 = \alpha_2 = \beta_1 = \beta_2 = 1$, the generalized DICE is the standard Dice: Dice coefficient $= \frac{2TP}{2TP+FN+FP}$. In order to emphasize the false positives of the defect area, we set $\alpha_1$ to be slightly higher than the other three parameters: $\alpha_1 = 4$ and $\alpha_2 = \beta_1 = \beta_2 = 1$. The parameters $\epsilon_1$ and $\epsilon_2$ are

**TABLE 1.** Performance comparisons of the different methods.

| Methods | F-score | MAP | MIoU | sensitivity | specificity |
|---|---|---|---|---|---|
| FCN [29] | 0.52198 | 0.86156 | 0.83547 | 0.43562 | 0.99390 |
| SegNet[21] | 0.45395 | 0.85610 | 0.83151 | 0.40924 | 0.98998 |
| U-Net [22] | 0.52279 | 0.87259 | 0.84298 | 0.48418 | 0.99023 |
| M-Net[6] | 0.43410 | 0.88261 | 0.83288 | 0.60291 | 0.96901 |
| CE-Net[37] | 0.59401 | 0.89007 | 0.86193 | 0.51541 | **0.99434** |
| CASU-Net | **0.67391** | **0.92293** | **0.87649** | **0.67559** | 0.99173 |

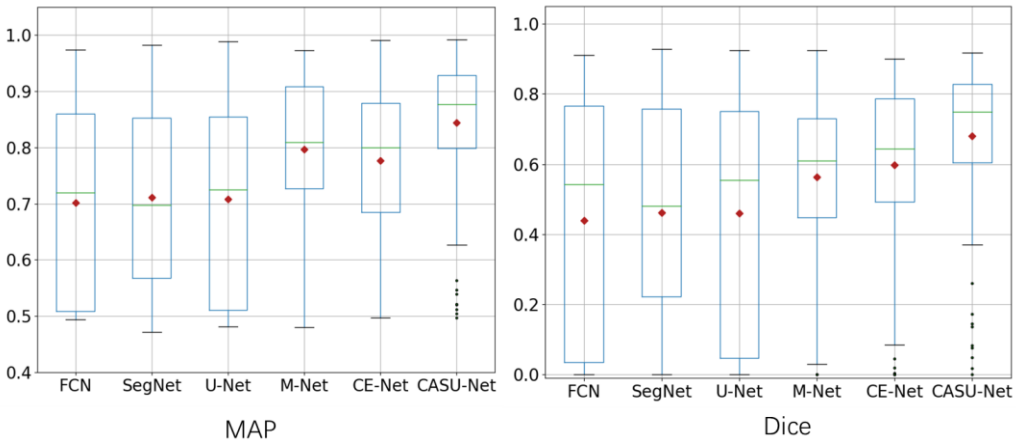


**FIGURE 5.** Boxplot of MAP and Dice on fundus images with RNFLD predicted by the proposed and compared methods.

two infinite quantities set to avoid meaningless expressions (the denominator is zero and ln0). We refer to the default infinitesimal value recommended in the Keras deep learning framework and set these two parameters to $10^{-7}$. The parameters $\theta$ and $\gamma$ are parameters in the proposed AW Loss, which are used to reflect the importance of the sample with large prediction error. When $\theta = 20$ is satisfied, for the sample with DICE $= 0$, the value of the loss is roughly as 10 times of the original value. Therefore, in this study, we set $\theta = 20$. We refer to the $\gamma$ parameter setting of focal loss in [34]. The parameter $\lambda$ is the weight used to balance the adversarial loss and the AW loss in the total loss function. The parameter $\lambda$ is set to 10 according to the weighted ratio of the multiple losses in the reference network [27]. In the comparative experiments in this paper, the hyperparameters of three methods (FCN, SegNet, and U-Net) are consistent with the literature [21], [22], and [29]. In this paper, 0.5 was used as the threshold to convert the probability map predicted by the network into a binary image.

### C. EXPERIMENTAL RESULTS

In the experiments, we compared the proposed CASU-Net with the following methods: FCN [29], SegNet [21], U-Net [22], M-Net [6], and CE-Net [37]. Experimental results are presented in Table. 1. As shown in Table.1, on four metrics of F-score, MAP, MIoU, and sensitivity, the proposed CASU-Net achieved the highest value. For specificity, the proposed CASU-Net is slightly lower than FCN and CE-Net. Additionally, compared with the three classical image segmentation methods (FCN, SegNet, and U-Net), the proposed CASU-Net obtains obvious prediction advantages. Receiver Operating Characteristic (ROC) curves predicted by different methods in the first experiment of the first-fold dataset is shown in Figure. 4.

As can be seen from Table.1, the sensitivity differences of various methods are more obvious than the specificity. In other words, compared with the background region, different methods have a more significant difference in the segmentation accuracy of the objective region (RNFLD). In order to show more details of the prediction accuracy of different methods for fundus images with RNFLD, we made boxplots of MAP and Dice predicted by different methods in the first experiment of the first-fold dataset. As can be seen in Figure.5, the 0.25, 0.5, and 0.75 quantiles of MAP for CASU-Net reach 0.7968, 0.8771, and 0.9303. The 0.25, 0.5, and 0.75 quantiles of Dice for CASU-Net reach 0.6009, 0.7492, and 0.8314, which are significantly higher than other methods.

Further, we showed visualization examples of methods proposed in this paper and other methods to predict the RNFLD in the first experiment of the first-fold dataset, as shown in Figure.6. For fundus images with RNFLD

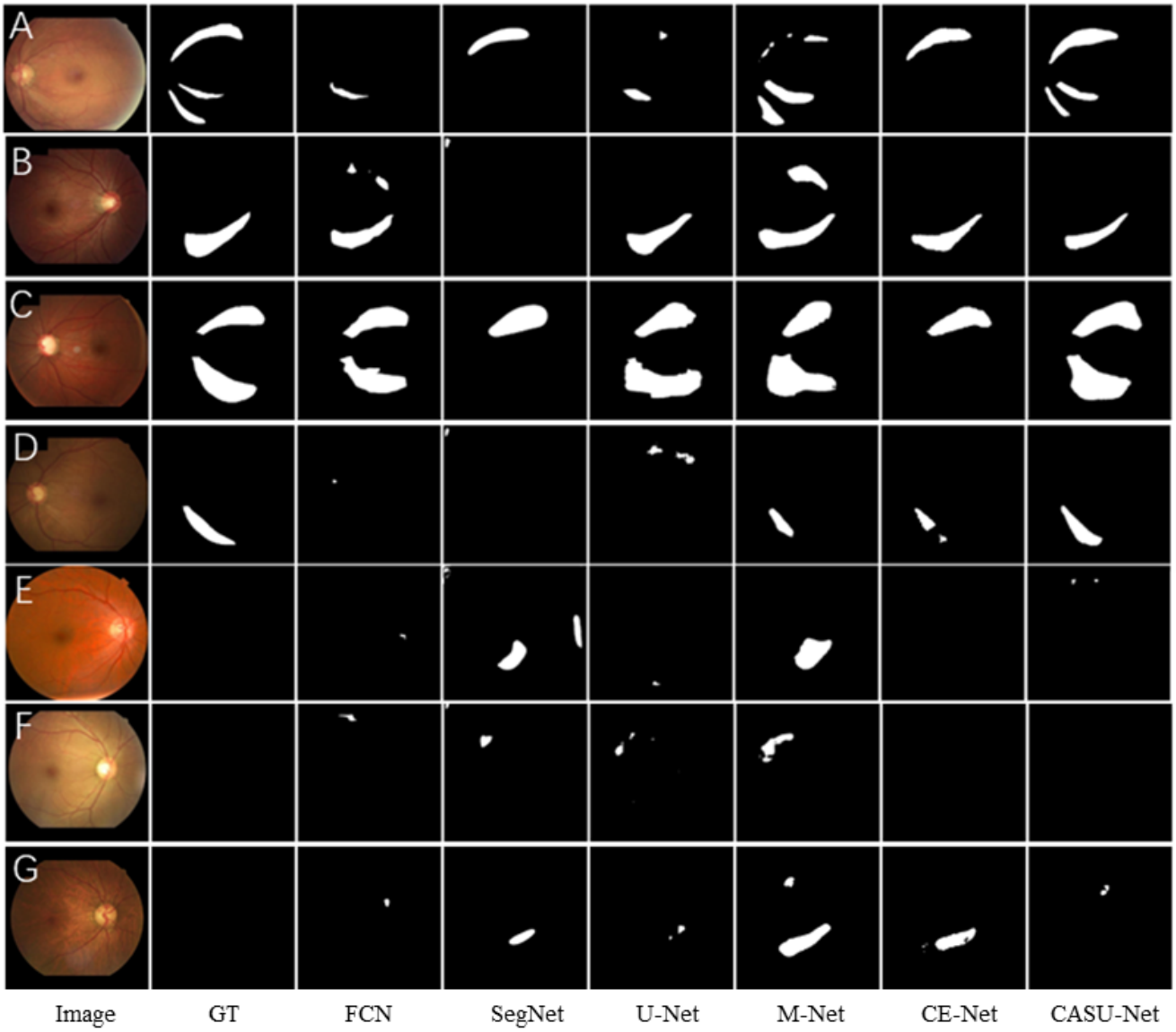


**FIGURE 6.** Visualization examples of different segmentation methods. The white area represents the prediction of the RNFLD, and the black represents the background. The pictures from left to right present fundus image, ground truth (GT), FCN, SegNet, U-Net, M-Net, CE-Net, and CASU-Net in this paper.

(fundus images labeled A-D), the segmentation of CASU-Net for the objective area is the most complete compared with those of other methods. The FCN, SegNet, and U-Net have obvious problems in incomplete segmentation for the objective area. The M-Net and CE-Net are improved compared with the above three methods, but still have an obvious gap compared with the ground truth. Additionally, the RNFLD region predicted by the proposed CASU-Net is closer to ground truth in morphology: on the one hand, the outline of the RNFLD region segmented by CASU-Net is smoother; on the other hand, the shape also conforms to the wedge structure. Especially for a few fundus images with special styles, the comparison methods have poor prediction performance. In the case in which there are multiple pieces of RNFLD or the contrast between the inner and outer areas of the RNFLD outline is not obvious, CASU-Net has a more obvious improvement, as shown in fundus images labeled A and D. For the fundus images without RNFLD (fundus images labeled E-G), the proposed CASU-Net has slightly wrong predictions, but predicts almost the entire fundus image as the background area.

In order to verify the effectiveness of the proposed shuffle module, AW loss, and GAN framework, we conducted further comparative experiments. The experimental results are presented in Table.2. The baseline (BL) is a modified U-shaped network with separable convolutions. Four BLs use four images of red channel (R), green channel (G), blue channel (B), and color fundus image (RGB) as their input for training. The experimental results show that when the green channel is used as the input, the BL can obtain better performance than when the red channel or the blue channel is used as the input. When the color fundus image is used

**TABLE 2. Performance comparisons of different components.**

| Methods | Input | F-score | MAP | MIoU | sensitivity | specificity |
|---|---|---|---|---|---|---|
| BL+CE loss | R | 0.46486 | 0.84466 | 0.82251 | 0.36379 | 0.99473 |
| BL+CE loss | G | 0.53502 | 0.88402 | 0.85261 | 0.52721 | 0.98798 |
| BL+CE loss | B | 0.50209 | 0.85434 | 0.83223 | 0.39339 | **0.99501** |
| BL+CE loss | RGB | 0.58673 | 0.88686 | 0.85468 | 0.56678 | 0.99001 |
| SU-Net(BL+shuffle)+CE loss | RGB | 0.61353 | 0.90560 | 0.86831 | 0.63078 | 0.98921 |
| SU-Net + AW loss | RGB | 0.63750 | 0.91600 | 0.87294 | 0.65206 | 0.98952 |
| SU-Net + GAN+AW loss | RGB | **0.67391** | **0.92293** | **0.87649** | **0.67559** | 0.99173 |

**TABLE 3. Performance comparisons of loss functions on BL model.**

| Methods | F-score | MAP | MIoU | sensitivity | specificity |
|---|---|---|---|---|---|
| BL+CE loss | 0.58673 | 0.88686 | 0.85468 | 0.56678 | 0.99001 |
| BL+focal loss | 0.58269 | 0.89267 | 0.85682 | 0.54273 | **0.99112** |
| BL+AW loss | **0.63185** | **0.90578** | **0.87001** | **0.61498** | 0.99089 |

as the input of BL, compared with when the green channel is used as the input, the prediction results are improved in all the five metrics, but the improvement is not obvious. This is consistent with the results reported in [29]. When the proposed shuffle module is combined with the BL, the SU-Net has achieved significant improvements on four metrics: F-score, MAP, MIoU, and sensitivity. The effectiveness of the proposed GAN framework and AW loss is further verified. After SU-Net is combined with the AW loss, the performance of SU-Net has been improved. When the GAN with SU-Net is combined with the AW loss, the proposed CASU-Net (SU-Net + GAN + AW) has achieved the best prediction performance on four metrics: F-score, MAP, MIoU, and sensitivity. Additionally, the improvement on F-score is obvious.

We further compare the impact of different loss functions on the BL model, and the results are shown in Table.3. From Table.3, we can see that the BL model with the proposed AW loss achieves the best prediction results on four metrics: F-score, MAP, MIoU, and sensitivity, and the advantages are obvious. On the specificity metric, the BL model with the focal loss achieves the best prediction result, but the advantage is slight.

## V. CONCLUSION

We designed a novel conditional adversarial shuffle U-shaped network to segment RNFLD from fundus images. The proposed CASU-Net consists of a generator and a discriminator. The SU-Net was first proposed as a generator, which employed a U-shape network with separable convolutions as the main structure. To achieve more efficient feature extraction, a shuffle module is constructed in SU-Net to make full use of the feature information in RGB three channels. For obtaining prediction results that are closer to the doctor's annotation morphologically, a discriminator was employed to supervise the spatial structure and geometry of the RNFLD predicted by the SU-Net. Furthermore, an adaptive weighted segmentation loss was designed to deal with the imbalance data of pixel-wise segmentation. Besides, the loss adaptively adjusts the weight of each fundus image in training, so that fundus images with large segmentation errors get more attention in training and generalization performance of the model can be enhanced. In the experiment, the proposed method obtained the state-of-the-art result, which can promote the automatic positioning of the RNFLD for glaucoma screening and alleviate the urgent need for professional glaucoma physicians.

## ACKNOWLEDGMENT

The authors would like to thank anonymous reviewers for their constructive suggestions. *(Shuai Lu and Man Hu contributed equally to this work.)*

## 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] J. I. Orlando *et al.*, "REFUGE challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs," *Med. Image Anal.*, vol. 59, Jan. 2020, Art. no. 101570.

[3] K. Lee, M. Niemeijer, M. K. Garvin, Y. H. Kwon, M. Sonka, and M. D. Abramoff, "Segmentation of the optic disc in 3-D OCT scans of the optic nerve head," *IEEE Trans. Med. Imag.*, vol. 29, no. 1, pp. 159–168, Jan. 2010.

[4] M. Wu, T. Leng, L. de Sisternes, D. L. Rubin, and Q. Chen, "Automated segmentation of optic disc in SD-OCT images and cup-to-disc ratios quantification by patch searching-based neural canal opening detection," *Opt. Express*, vol. 23, no. 24, pp. 31216–31229, 2015.

[5] H. Fu, Y. Xu, S. Lin, X. Zhang, D. W. K. Wong, J. Liu, A. F. Frangi, M. Baskaran, and T. Aung, "Segmentation and quantification for angle-closure glaucoma assessment in anterior segment OCT," *IEEE Trans. Med. Imag.*, vol. 36, no. 9, pp. 1930–1938, Sep. 2017.

[6] 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 Trans. Med. Imag.*, vol. 37, no. 7, pp. 1597–1605, Jul. 2018.
[7] M. S. Haleem, L. Han, J. van Hemert, and B. Li, "Automatic extraction of retinal features from colour retinal images for glaucoma diagnosis: A review," *Comput. Med. Imag. Graph.*, vol. 37, nos. 7–8, pp. 581–596, Oct. 2013.
[8] C. Muramatsu, Y. Hayashi, A. Sawada, Y. Hatanaka, T. Hara, T. Yamamoto, and H. Fujita, "Detection of retinal nerve fiber layer defects on retinal fundus images for early diagnosis of glaucoma," *J. Biomed. Opt.*, vol. 15, no. 1, 2010, Art. no. 016021, doi: 10.1117/1.3322388.
[9] J. E. Oh, H. K. Yang, K. G. Kim, and J. M. Hwang, "Automatic computer-aided diagnosis of retinal nerve fiber layer defects using fundus photographs in optic neuropathy," *Investigative Ophthalmol. Vis. Sci.*, vol. 56, no. 5, pp. 2872–2879, 2015, doi: 10.1167/iovs.14-15096.
[10] D. Lamani, T. C. Manjunath, M. Mahesh, and Y. S. Nijagunarya, "Early detection of glaucoma through retinal nerve fiber layer analysis using fractal dimension and texture feature," *Int. J. Res. Eng. Technol.*, vol. 3, no. 10, pp. 158–163, 2014.
[11] R. Panda, N. B. Puhan, A. Rao, D. Padhy, and G. Panda, "Automated retinal nerve fiber layer defect detection using fundus imaging in glaucoma," *Comput. Med. Imag. Graph.*, vol. 66, pp. 56–65, Jun. 2018, doi: 10.1016/j.compmedimag.2018.02.006.
[12] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
[13] A. Diaz-Pinto, A. Colomer, V. Naranjo, S. Morales, Y. Xu, and A. F. Frangi, "Retinal image synthesis and semi-supervised learning for glaucoma assessment," *IEEE Trans. Med. Imag.*, vol. 38, no. 9, pp. 2211–2218, Sep. 2019.
[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[15] X. Zhang, X. Zhou, M. Lin, and J. Sun, "ShuffleNet: An extremely efficient convolutional neural network for mobile devices," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6848–6856, doi: 10.1109/CVPR.2018.00716.
[16] 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 Trans. Med. Imag.*, vol. 37, no. 11, pp. 2493–2501, Nov. 2018, doi: 10.1109/TMI.2018.2837012.
[17] L. Li, M. Xu, H. Liu, Y. Li, X. Wang, L. Jiang, Z. Wang, X. Fan, and N. Wang, "A large-scale database and a CNN model for attention-based glaucoma detection," *IEEE Trans. Med. Imag.*, vol. 39, no. 2, pp. 413–424, Feb. 2020, doi: 10.1109/tmi.2019.2927226.
[18] A. Hassan and A. Mahmood, "Convolutional recurrent deep learning model for sentence classification," *IEEE Access*, vol. 6, pp. 13949–13957, 2018.
[19] Q. Lu, C. Liu, Z. Jiang, A. Men, and B. Yang, "G-CNN: Object detection via grid convolutional neural network," *IEEE Access*, vol. 5, pp. 24023–24031, 2017.
[20] Y. Shin, H. A. Qadir, L. Aabakken, J. Bergsland, and I. Balasingham, "Automatic colon polyp detection using region based deep CNN and post learning approaches," *IEEE Access*, vol. 6, pp. 40950–40962, 2018.
[21] L. Wu, Y. Liu, Y. Shi, B. Sheng, P. Li, L. Bi, and J. Kim, "Optic disc and cup segmentation based on enhanced SegNet," in *Proc. 32nd Int. Conf. Comput. Animation Social Agents (CASA)*, 2019, pp. 33–36.
[22] A. Sevastopolsky, "Optic disc and cup segmentation methods for glaucoma detection with modification of U-Net convolutional neural network," *Pattern Recognit. Image Anal.*, vol. 27, no. 3, pp. 618–624, Jul. 2017.
[23] S. A. Taghanaki, K. Abhishek, J. P. Cohen, J. Cohen-Adad, and G. Hamarneh, "Deep semantic segmentation of natural and medical images: A review," *Artif. Intell. Rev.*, to be published, doi: 10.1007/s10462-020-09854-1.
[24] Y.-L. Xu, S. Lu, H.-X. Li, and R.-R. Li, "Mixed maximum loss design for optic disc and optic cup segmentation with deep learning from imbalanced samples," *Sensors*, vol. 19, no. 20, p. 4401, Oct. 2019.
[25] J. Zilly, J. M. Buhmann, and D. Mahapatra, "Glaucoma detection using entropy sampling and ensemble learning for automatic optic cup and disc segmentation," *Comput. Med. Imag. Graph.*, vol. 55, pp. 28–41, Jan. 2017.
[26] C. Du and S. Gao, "Image segmentation-based multi-focus image fusion through multi-scale convolutional neural network," *IEEE Access*, vol. 5, pp. 15750–15761, 2017.
[27] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1125–1134.
[28] R. Panda, N. B. Puhan, A. Rao, B. Mandal, D. Padhy, and G. Panda, "Deep convolutional neural network-based patch classification for retinal nerve fiber layer defect detection in early glaucoma," *J. Med. Imag.*, vol. 5, no. 4, p. 44003, 2018.
[29] R. Watanabe, C. Muramatsu, K. Ishida, A. Sawada, Y. Hatanaka, T. Yamamoto, H. Fujita, "Automated detection of nerve fiber layer defects on retinal fundus images using fully convolutional network for early diagnosis of glaucoma," *Proc. SPIE*, vol. 10134, Mar. 2017, Art. no. 1013438, doi: 10.1117/12.2254574.
[30] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.
[31] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2015, pp. 234–241.
[32] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1251–1258.
[33] M. Mirza and S. Osindero, "Conditional generative adversarial nets," 2014, *arXiv:1411.1784*. [Online]. Available: http://arxiv.org/abs/1411.1784
[34] T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 2999–3007, doi: 10.1109/ICCV.2017.324.
[35] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, and J. Garcia-Rodriguez, "A review on deep learning techniques applied to semantic segmentation," 2017, *arXiv:1704.06857*. [Online]. Available: http://arxiv.org/abs/1704.06857
[36] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.
[37] 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 Trans. Med. Imag.*, vol. 38, no. 10, pp. 2281–2292, Oct. 2019, doi: 10.1109/tmi.2019.2903562.

**SHUAI LU** received the B.Sc. degree from the Beijing University of Chemical Technology, Beijing, China, in 2017, where he is currently pursuing the master's degree. His current research interests include machine learning, deep learning, and ensemble learning.

**MAN HU** received the M.D. degree from Capital Medical University, Beijing, China, in 2019. She did her research work at Beijing Tongren Hospital, Capital Medical University. She has been with the Department of Ophthalmology, Beijing Children's Hospital, Capital Medical University, since 2009, where she is currently an Associate Senior Doctor. Her current research interests include artificial-intelligence-aided glaucoma diagnosis and genetics of congenital glaucoma.

**RUIRUI LI** (Member, IEEE) received the Ph.D. degree in computer science and technology from Tsinghua University, Beijing, China, in 2014. She joined the College of Information Science and Technology, Beijing University of Chemical Technology, as a Postdoctoral Researcher, where she was appointed as a Lecturer, in 2017. She has published more than 20 peer-reviewed articles. She is the Inventor or Co-Inventor of two patents. Her research interests include image processing in remote sensing, machine learning and pattern analysis, and computer vision. She is a Peer Reviewer of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS and *Remote Sensing*.

**YONGLI XU** received the B.E. and Ph.D. degrees from the Beijing University of Aeronautics and Astronautics, Beijing, China, in 2005 and 2010, respectively. He has been with the Department of Mathematics, Beijing University of Chemical Technology, Beijing, since 2010, where he is currently an Associate Professor. His current research interests include machine learning, system identification, and biomedical image analysis. He is a Reviewer for a number of international journals and conferences, such as the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, the IEEE TRANSACTIONS ON AUTOMATIC CONTROL, *Neurocomputing*, and the American Control Conference.