Joint segmentation of optic cup and optic disc using deep convolutional generative adversarial network

Linfang Yu¹,*
¹No.4, Section 2, North Jianshe Road, Chengdu, P.R.China Zip/Postal Code: 610054
yulinfang@std.uestc.edu.cn

Abstract. Glaucoma, as one of the three major blinding ophthalmic diseases in the world, is usually accompanied by changes in the structure of the patient's optic disc, such as optic disc atrophy and depression. Clinical ophthalmologists tend to use the cup-disc ratio as an evaluation index to realize the screening and diagnosis of glaucoma. Therefore, the accurate measurement of optic cup (OC), optic disc (OD) and other parameters is of great clinical significance for early screening of glaucoma. Inspired by game theory, this paper combines deep convolutional neural networks (DCNN) with generative adversarial networks (GAN), and proposes a model for the joint segmentation of OC and OD. Specifically, the generator is composed of a deep convolutional encoder-decoder network to jointly segment the OC and OD, and the discriminator is composed of an eight layer full convolutional neural network. The discrimination results adjust the parameters in the structure of the generator and discriminator network through back propagation to achieve the effect of autonomous learning and optimization of the model. When the proposed network and the existing networks are evaluated on the public dataset Drishti-GS1, the research results demonstrate that the proposed network can achieve a significant improvement in the overall performance.

1. Introduction
The occurrence of glaucoma is mainly due to the damage of optic nerve caused by high intraocular pressure, and this damage is usually irreversible [1]. Its pathogenesis usually has no obvious characteristics, but it will be accompanied by the changes of OC, OD and optic nerve fiber layer. Therefore, the accurate measurement of various parameters such as OC and OD has important clinical significance for the early screening of glaucoma and inhibiting the development of glaucoma. In the current ophthalmic diagnosis, the vertical cup-disc ratio (VCDR) [2], as one of the most important evaluation indicator, is well accepted by clinicians and widely used in the diagnosis of glaucoma. Figure 1 shows the comparison of the OD structure between the normal fundus image and the fundus image of glaucoma patients. The region surrounded by the blue solid line is the OD, while the region surrounded by the green solid line is the OC. The VCDR can be obtained by calculating the ratio of vertical cup diameter to vertical disc diameter. Generally, the greater the ratio, the higher the risk of glaucoma. Therefore, how to accurately and efficiently segment the OC and OD from the fundus image has become the fundamental task.

Traditional machine learning relies on tedious manual feature extraction, which makes the performance of these methods vulnerable to low resolution images and lesion regions. In contrast, DCNN can automatically and effectively extract key features from fundus images without or less affected by lesions. In addition, the research and application of generative adversarial networks (GAN)
model inspired by game theory in the field of image and vision has also brought new development momentum. Based on the above, this paper considers the combination of GAN and DCNN, and proposes a method based on deep convolutional generative adversarial networks (DCGAN) to realize the joint segmentation of OC and OD. Specifically, in the generator model, a deep convolutional encoder-decoder network is used to jointly segment the OC and OD. The encoder consists of 13 convolutional network layers (the pooling layer is not included), and each encoder layer corresponds to a decoder layer. The feature maps generated by the decoder are finally sent to the soft-max classifier for pixel by pixel classification, that is, the category probability is generated independently for each pixel. The discriminator model is composed of an 8-layer convolutional neural network. The discrimination results adjust the parameters in the structure of the generator and discriminator network through back propagation to achieve the effect of autonomous learning and optimization of the model. The main contributions of this paper are as follows:

1) A joint segmentation method of OC and OD based on DCGAN is proposed. When using publicly available dataset for evaluation, the research results demonstrate that compared with most existing methods, the proposed method can achieve a significant improvement in the overall performance.

2) The proposed network combines the GAN with the deep convolution network, and trains the model in the way of confrontation. Specifically, the discriminator first learns the difference between the true and false labels, and then guides the generator to reduce the difference through back propagation, so as to improve the segmentation performance of the model.

3) Data augmentation technology is introduced to alleviate the problems of model over fitting caused by the small scale of medical image dataset, so as to improve the network robustness and segmentation performance.

The remaining chapters are arranged as follows. Related research works are introduced in Section 2. The proposed network is described in detail in Section 3. The evaluation results are presented in Section 4. Finally, the paper is summarized in Section 5.

2. Related work
At present, there are many researches on the segmentation of OC and OD in fundus images at home and abroad. In order to solve the problem of cross domain segmentation between different datasets, Zhang et al. [3] proposed a TAU model based on U-Net and attention mechanism for the segmentation of OD and OC. Wang et al. [4] proposed a pOSAL framework to segment the OC and OD in different datasets. By introducing adversarial learning into the output space, the framework effectively improves the segmentation performance of different test domains. Rashmi et al. [5] segmented OD based on the improved random walk algorithm. Assuming that the regions of the OC and OD are ellipses, Jiang et al. [6] introduced an end-to-end CNN called JointRCNN to jointly segment the OC and OD. Fu et al. [7] proposed an M-Net architecture, which can realize the joint segmentation of OD and OC in a multi label system. Chen et al. [8] introduced the domain adaptive method into the input and output space and proposed an unsupervised framework IOSUDA to alleviate the performance degradation caused by joint segmentation. Bhatkalkar et al. [9] proposed a novel CNN architecture to accurately segment the OD. Guo et al. [10] proposed an adaptive framework based on Faster R-CNN. The framework is composed of two networks: coarse network and fine network, which can be used to realize cross-domain OC and OD joint segmentation. Ega et al. [11] realized OD detection by using...
the innovative color fusion model generated by Markowitz's modern portfolio theory. In order to effectively extract context information, Yuan et al. [12] combined residual network with CNN and proposed a residual multiscale CNN to effectively segment OC and OD. Abdullah et al. [13] proposed a new method to find the boundary of the OD region through the initial fuzzy clustering algorithm, so as to realize the localization and segmentation of the OD in the fundus image.

3. Method

3.1. Model architecture
In this paper, the DCGAN-based model is designed to automatically segment the OC and OD. The architecture is shown in Figure 2. Set the fundus image dataset as $x_i \in X(1 \leq i \leq N)$ and the label dataset of the OC and OD as $y_i \in Y(1 \leq i \leq N)$. Inspired by the game theory in GAN, the model consists of generator and discriminator. The purpose of the generator is to jointly segment the OC and OD by learning the mapping relationship between the fundus image $x$ and its corresponding OC and OD (that is, $X \rightarrow Y$), so that the performance of the segmentation result $G(x)$ on the discriminator $D(G(x))$ is consistent with that of the label $y$ on the discriminator $D(y)$. Meanwhile, by learning the difference between $y$ and $G(x)$, the discriminator can correctly distinguish whether the OC and OD label input into the network comes from the label dataset or the generator, so as to guide the generator to reduce this difference, so that the model can segment the cup and disc more accurately.

Model training needs to continuously optimize the generator and discriminator, improve their segmentation ability and discrimination ability, and find the Nash equilibrium between them. When the output of the discriminator is equal to 1/2, that is, the source of the label input to the network cannot be correctly distinguished, the training is considered to be completed.

![Figure 2. The architecture of DCGAN.](image)

3.2. Generator network
The generator is an encoder-decoder network model, and its model architecture is shown in Figure 3. The encoder part uses the first 13 layers convolution structure of VGG16. Each convolution layer contains convolution, batch normalization and ReLU operations. The pooling layer adopts 2x2 window with a step of 2. Each pooling is equivalent to a down-sampling of the image by halving the resolution. In the process of maximum pooling, the position of the maximum value in each pooling window in the feature map will be recorded. In the decoder part, after the convolution-pooling operations, the feature maps of the input image are obtained, and the recorded maximum pooling index are used for nonlinear up-sampling. These operations will produce sparse feature maps, and then perform convolution operations to produce dense feature maps. It is worth noting that the last layer convolution of the decoder generates a 3-channel feature map corresponding to the three categories of
OC, OD and background in this paper, and then the feature map is sent to the soft-max classifier for pixel by pixel classification.

Figure 3. Generator network architecture.

3.3. Discriminator network
The discriminator is an 8-layer convolutional neural network model. The fundus image x and label image y (or G(x)) are spliced and input into the discriminator. The output of the discriminator is the probability that y is the segmented label image of fundus image x, that is, D(x, y), so as to optimize the parameters in the network structure. In the discrimination network, the pooling layer is removed, and each layer uses convolution with stride, which can not only quickly converge the model, but also effectively extract key features. The resolution of the feature map after each convolution is reduced to half of the previous feature map. In order to accelerate convergence and alleviate the influence of initialization weight on the discriminator, Leaky-ReLU is used as the activation function while batch normalization is performed in all convolution layers. After 6 down-sampling operations, the resolution of feature map is reduced to 1/64 of the original. Finally, the feature map is flattened and connected with the full connection layer to output a final discrimination result. After calculating the error between the discrimination result and the label, the parameters of the generator and discriminator are adjusted by back propagation according to the way of random gradient descent, so as to achieve the effect of model autonomous learning optimization.

4. Experiments and results

4.1. Implementation

4.1.1. Datasets. In this paper, the Drishti-GS1 public dataset is used for related experiments. The Drishti-GS1 dataset contains 101 color fundus images, 70 of which are collected from glaucoma patients, and the rest are normal fundus images. The dataset was divided into training set and test set in advance, of which 50 fundus images were used for training and the remaining 51 were used for testing. All fundus images were taken with OD as the center, and the image resolution was 2047×1760 pixels. In order to capture the labeling differences between observers, each fundus image provides manual labeling of the OC and OD from four glaucoma experts (with 3, 5, 9 and 20 years of clinical experience respectively). In this experiment, the manual labeling of four experts are fused, and the mean value is taken as the gold standard for the segmentation training of OC and OD.

Due to the small scale of the training data, in order to prevent over fitting of the model and improve the accuracy and robustness of the model, we have carried out the data augmentation. Specifically, the following operations are randomly taken for each training image: 1) horizontal or vertical flipping; 2) random rotation in the range of 0 to 360 degrees; 3) random reduction or amplification in the range of 0.8 to 1.6 times. Finally, for each fundus image, we randomly intercept the resolution size of 512 × 512 as input.
4.1.2. Evaluation metrics. We define the pixels constituting the OC and OD in the fundus image as the target pixels, and introduce the following measures: true positive (TP), false positive (FP), true negative (TN) and false negative (FN). Specifically, TP indicates that the target pixels are correctly classified into corresponding target; FP indicates that non-target pixels are misclassified into target; TN indicates that the non-target pixels are correctly classified as non-target; FN indicates that the target pixels are misclassified into non-target. In this paper, F1 score and accuracy (Acc) are used for model performance evaluation, where F1 score can be regarded as the harmonic average of model accuracy and recall. The mathematical expressions are as follows:

\[
F_1 \text{-score} = \frac{2TP}{2TP + FP + FN}
\]

\[
Acc = \frac{TN + TP}{TP + FN + FP + TN}
\]

4.1.3. Experimental configuration. In this experiment, the hardware facilities of an NVIDIA GTX1080 GPU graphics card and an Intel Core i7-7600T CPU are configured. The proposed model uses Adam as the optimizer and is trained by random gradient descent. The initial learning rate is 10^{-4}. After 400 iterations, the learning rate becomes 10^{-6}. To reduce the instability of random gradient during training, the mini batch=2.

4.2. Results
We used FCN, U-Net, SegNet and the proposed DCGAN model to carry out the OC and OD segmentation experiment on the Drishti-GS1 dataset. Figure 4 intuitively shows the visual segmentation results of different networks. From left to right: original color fundus image, groundtruth, segmentation results of the proposed network, and segmentation results of three other networks. The white part represents the OC and the gray part (including the white part) represents the OD. After comparison, it is obvious that the model proposed in this paper can segment the OC and OD more accurately, especially for the OC segmentation, which is obviously better than the other three networks.

![Figure 4. Comparison of visualization results of different networks on Drishti-GS1.](image)

**Table 1.** Performance comparison of different networks on Drishti-GS1.

| Method       | OC F1  | OC Acc | OD F1  | OD Acc |
|--------------|--------|--------|--------|--------|
| FCN          | 0.7601 | 0.9428 | 0.9518 | 0.9713 |
| SegNet       | 0.7738 | 0.9464 | 0.9565 | 0.9742 |
| U-net        | 0.7897 | 0.9519 | 0.9721 | 0.9829 |
| Proposed network | **0.8579** | **0.9709** | **0.9739** | **0.9841** |

Table 1 provides the performance comparison of four different network models on the Drishti-GS1 public dataset, in which the bold font represents the best result. It is not difficult to see from the table...
that compared with the other three network models, the DCGAN model proposed in this paper obtains the best performance. Especially in the OC segmentation, its performance is much better than the other three networks, and achieves an accurate rate of 97.09% and F1 score of 85.79%.

5. Conclusion
Considering the combination of DCNN and GAN, a model based on DCGAN is proposed to jointly segment the OC and OD. Specifically, the generator is composed of a deep convolutional encoder-decoder network to jointly segment the OC and OD, and the discriminator is composed of an eight layer full convolutional neural network. The discrimination results adjust the parameters in the structure of the generator and discriminator network through back propagation to achieve the effect of autonomous learning and optimization of the model. When the proposed network and the existing networks are evaluated on the public dataset Drishti-GS1, the research results demonstrate that the proposed network can achieve a significant improvement in the overall performance.

References
[1] Yih-Chung, Tham, Xiang, et al. Global Prevalence of Glaucoma and Projections of Glaucoma Burden through 2040: A Systematic Review and Meta-Analysis[J]. Ophthalmology, 2014.
[2] Jonas J B , Bergua A , Schmitzvalckenberg P , et al. Ranking of optic disc variables for detection of glaucomatous optic nerve damage.[J]. Investigative Ophthalmology & Visual Science, 2000, 41(7):1764.
[3] Zhang Y , Cai X , Zhang Y , et al. TAU: Transferable Attention U-Net for optic disc and cup segmentation[J]. Knowledge-Based Systems, 2021, 213(11):106668.
[4] Wang S , Yu L , Yang X , et al. Patch-based Output Space Adversarial Learning for Joint Optic Disc and Cup Segmentation[J]. IEEE Transactions on Medical Imaging, 2019:1-1.
[5] Rashmi P , Pulan N B , Ganapati P . Mean curvature and texture constrained composite weighted random walk algorithm for optic disc segmentation towards glaucoma screening[J]. Healthcare Technology Letters, 2018, 5(1):31-37.
[6] Jiang Y , Duan L , J Cheng, et al. JointRCNN: A Region-Based Convolutional Neural Network for Optic Disc and Cup Segmentation[J]. IEEE Transactions on Biomedical Engineering, 2020, 67(2):335-343.
[7] Fu H , Cheng J , Xu Y , et al. Joint Optic Disc and Cup Segmentation Based on Multi-label Deep Network and Polar Transformation[J]. IEEE Transactions on Medical Imaging, 2018, PP(99):1-1.
[8] Chen C , Wang G . IOSUDA: an unsupervised domain adaptation with input and output space alignment for joint optic disc and cup segmentation[J]. Applied Intelligence, 2020:1-19.
[9] Bhatkalkar B J , Reddy D R , Prabhu S , et al. Improving the performance of convolutional neural network for the segmentation of optic disc in fundus images using attention gates and conditional random fields[J]. IEEE Access, 2020, PP(99):1-1.
[10] Guo Y , Peng Y , Zhang B . CAFR-CNN: coarse-to-fine adaptive faster R-CNN for cross-domain joint optic disc and cup segmentation[J]. Applied Intelligence, 2021(1).
[11] Ega E , Tb B , Mg C , et al. A color fusion model based on Markowitz portfolio optimization for optic disc segmentation in retinal images - ScienceDirect[J]. Expert Systems with Applications, 2021.
[12] Yuan X , Zhou L , Yu S , et al. A Multi-scale Convolutional Neural Network with Context for Joint Segmentation of Optic Disc and Cup[J]. Artificial Intelligence in Medicine, 2021, 113(11):102035.
[13] Abdullah A S , Ra Hebi J , Zok Y E , et al. A new and effective method for human retina optic disc segmentation with fuzzy clustering method based on active contour model[J]. Medical & Biological Engineering & Computing, 2019, 58(8).