**Research on Biomedical Engineering manuscript No.**
(will be inserted by the editor)

# Inter-Dataset Performance Analysis of Generative Adversarial Networks for Optic Disc Segmentation Using Digital Fundus Images

**Ambika Sharma · Monika Agrawal · Sumantra Dutta Roy · Vivek Gupta**

Received: date / Accepted: date

**Abstract**
***Purpose*** Optic disc segmentation helps in non-invasive diagnosis of incurable and blindness-causing diseases of the eye such as glaucoma, diabetic retinopathy and optic disc drusen. Optic disc segmentation also plays a vital role in locating anatomical structures like retinal blood vessels, macula, and optic cup. Contemporary systems are based on image processing/computer vision and/or machine learning. These generally fail to handle the challenging pathologies under poor lightning conditions and the presence of clinical exudates and other distractions. Authors typically evaluate the performance of their algorithms on a few datasets, thus limiting their generalization and extrapolation ability on new datasets.
***Methods*** We propose a U-Net-based adversarial strategy to segment the optic disc, based on the localization of the probable region using a U-Net-based regressor. We perform extensive experimentation with challenging datasets, and inter-dataset training, to validate the generalization capability of the proposed method.
***Results*** The proposed method has been validated on seven publicly available retinal datasets and a challenging private community camp dataset obtained from the All India Institute of Medical Sciences (AIIMS), New Delhi. The generated results show an average Dice coefficient, disc overlap, and sensitivity value of 97.31%, 93.78% and 97.72% respectively on eight datasets. Moreover, the work reaches the benchmark performance for the Drishti-GS and Refugee datasets with a Dice coefficient of 97.31% and 95.12% respectively.

Ambika Sharma
Indian Institute of Technology Delhi
E-mail: ambika.sharma@dbst.iit.ac.in

Monika Agrawal
Indian Institute of Technology Delhi
E-mail: maggarwal@care.iitd.ac.in

Sumantra Dutta Roy
Indian Institute of Technology Delhi
E-mail: sumantra@ee.iitd.ac.in

Vivek Gupta
All India Institute of Medical Sciences, New Delhi
E-mail: vgupta@aiims.ac.in

***Conclusion*** The proposed method works efficiently on poor resolution images with various retinal artifacts such as blurring, poor illumination, and other pathologies such as peripapillary atrophy and disc hemorrhages.

**Keywords** Retinal Images · Image Processing · Deep Convolutional Neural Network · Generative Adversarial Network · Image Segmentation

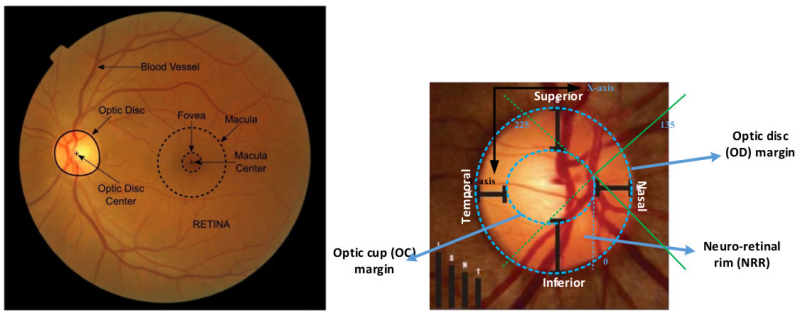
## 1 Introduction

Optic disc segmentation is one of the primary steps to screen for diseases such as glaucoma, diabetic retinopathy, and age-related macular degeneration. Early screening of such diseases is required to avoid any permanent loss of vision. Current diagnostic procedures employ fundus (or retinal) imaging or optical coherence tomography (OCT) to capture images and a specialized doctor to analyze them (Trucco et al. 2019). Simple retinal imaging (as opposed to OCT) can provide cost-effective solutions for disease detection, particularly in developing nations without easy access to good healthcare (Acharjya et al. 2019). Retinal images are a two-dimensional picture of the eye, with three anatomical landmarks i.e., optic disc with the optic cup in it, macula (with the fovea inside it), and retinal blood vessel structure and the neuro-retinal rim (Fig. 1).

Robust optic disc detection holds the key to non-invasive low-cost detection of potential blindness-causing diseases such as glaucoma. In a healthy individual, the neuro-retinal rim widely follows the ISNT rule: the inferior being the thickest section, followed by the superior, nasal, and finally, the temporal region. The violation of the ISNT rule strictly implies the presence of glaucoma disease in the retina. Moreover, a large CDR (ratio of the vertical cup to disc diameter) value indicates the high risk of glaucoma (Dada and Coote 2010). However, optic disc segmentation is a difficult problem. Fig. 2(a) and (b) depict a healthy optic nerve head with a good intensity gradient at the optic disc margin. Fig. 2(c) and (d) show common situation with ill-defined boundaries due to poor illumination and peripapillary atrophy (uneven texture and color outside the optic disc margin).

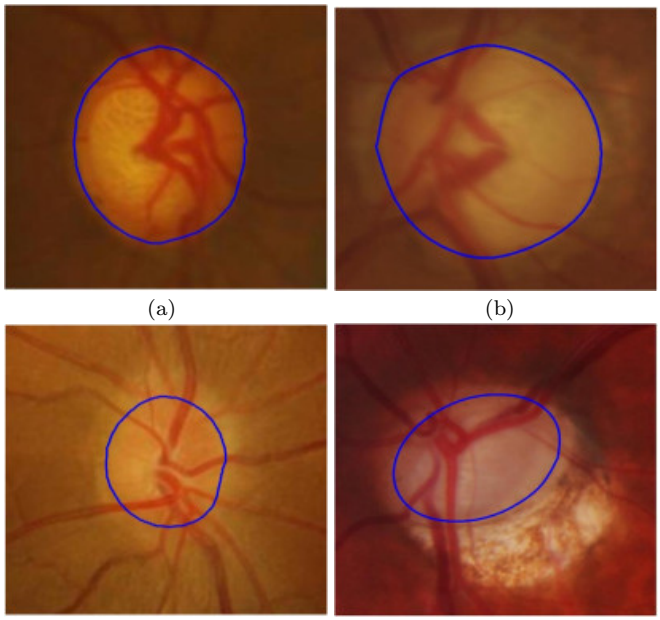
### 1.1 Image Processing/Computer Vision-based Approaches

One can group the available literature into two classes of methods: those based on image processing or computer vision, and those primarily based on machine learning. In a healthy retinal image, the optic disc is circular in shape and has a high edge gradient. Using this knowledge, (Aquino et al. 2010) suggest an edge detector and thresholding-based method to find the probable optic disc boundary candidates. A polar transform-based local thresholding (Zahoor and Fraz 2017) with a series of image normalization pre-processing steps is another approach. A set of works (Gao et al. 2019a), (Zou et al. 2019) propose an adaptive level set contour extraction approach using saliency and thresholding techniques. (Gao et al. 2019b) also incorporate prior shape information into a modified local image fitting (LIF) model for abnormal retinal images. (Fraga et al. 2012) use a fuzzy Hough transform to locate the bright circular regions and to find the disc edge




(a) (b)

**Fig. 1** Sample retinal images with marked landmarks. (a) The optic disc (with the optic cup inside), macula and blood vessels: three major anatomical structures. (b) An enlarged annotated version with the optic disc and optic cup



(a) (b)

(c) (d)

**Fig. 2** (a), (b): Healthy optic discs with a sharp gradient at the disc-background margin. (c), (d): Challenging cases with ill-defined disc boundaries due to non-uniform lighting and the peripapillary atrophy condition

points. (Abdullah et al. 2016) also use the Hough transform and a grow-cut-based algorithm in addition to some morphological operations to enhance the region of interest. Another approach (Joshi et al. 2011) use local image information present around each point of interest in multi-dimensional feature space (from color and texture information) and an active contour model. (Cheng et al. 2013) use super-pixel classification to segment the disc region. The method uses histograms and

center-surround statistics to classify each pixel. Using the basic assumption of circularity and high edge gradient (Lu 2011) perform an image variation study along multiple evenly-oriented radial lines, with the knowledge of optic disc center. (Gao et al. 2019a) propose the use of multi-view information from shape and appearance using a modified active contour model.

### 1.2 **Machine Learning-based Methods**

Deep neural networks are popular for various localization, segmentation, and classification problems. (Fu et al. 2018) use a transformed polar image for training a deep M-Net architecture. The network feeds a series of multi-scale pyramid images at each level in order to concatenate it with the down-sampled feature matrix of a U-Net model. A deep learning-based approach (Qin et al. 2019) uses the Inception structure in the GoogleNet models. (Jiang et al. 2020) use Region-Based Convolutional Neural Network (R-CNN, hereafter) architecture to locate a bounding box around the optic disc and fit an ellipse to segment the true disc margin. (Singh et al. 2018), (Jiang et al. 2019), (Liu et al. 2019) use generative adversarial networks to segment the disc region.

In more recent work, (Pachade et al. 2021) use an adversarial strategy with a rather complex architecture. This has a nested EfficientNet encoder. The decoder includes pre-activated residual blocks and attention gates before and after concatenation operations. Our work achieves a much better performance (for instance, a 0.99% higher Dice coefficient on the representative Drishti-GS dataset), in spite of a much simpler architecture. The approach in (Veena et al. 2022) leverages a combination of deep learning (a U-Net model) and image processing techniques (equalization, morphology, shape detection, and watershed segmentation) to locate the disc. However, this approach is limited in its generalization due to the use of the Drishti-GS dataset alone. Performance on other datasets is not guaranteed, with differing characteristics due to inter-dataset variability. The authors in (Pande Darma Suardika et al. 2022) use a R-CNN-based model for segmentation. However, the evaluation is restricted to a solitary dataset (IDRiD), a limitation similar to the previous reference, above.

A cascaded two-stage framework in (Wang et al. 2022) has a U-Net in the first-stage to locate the optic disc. The second stage integrates multi-task framework (estimating the mask, contours and distance maps) and adversarial learning. A similar approach in (Luo et al. 2021) uses boundary and map prediction for adversarial learning. The work in (Wang et al. 2022) involves large resource requirements due to training a U-Net model twice, once for optic disc detection and then for multiple tasks performed separately. Furthermore, the evaluation of the work is limited to two public datasets alone. This may result in reduced generalization when tested on other diverse datasets. The authors in (Fu et al. 2021) propose a U-Net model-based method which utilizes a model-driven probability bubble approach. This takes into account the positional relationship between vessels and the optic disc. The single-dataset testing and the lack of clarity in the training for localization, are limitations of the work.

Existing methods fail to handle challenging atrophy discs under poor lightning conditions and lack generalization among large public optic disc datasets (Lowell et al. 2004) (Lowell et al., 2004), (Aquino et al. 2010). In this paper, we propose

an adversarial training to extract the optic disc pixels from the given RGB retinal image. The deep adversarial network is based on a U-Net architecture as a generator to produce optic disc maps with the condition of input RGB retinal image. In addition to this, a cascade of convolutional layers network is used as a discriminator to classify an input image as real or machine generated. A combination of segmentation and adversarial loss is employed to train the network with Adam optimizer. The major contributions of the proposed work are as follows:

1. We propose a novel optic disc localization methodology using a fused grayscale image generated from distance and intensity maps when trained on a U-Net model. The optic disc center coordinates are located using a Laplacian of Gaussian filter on a fused map. We use a Tukey's biweight robust loss function to handle the pathologies similar in appearance to the optic disc. This is more robust to outliers than a mean square error.
2. We propose a fully automatic end-to-end generative adversarial network (GAN) is proposed to segment the optic disc, with adversarial and segmentation losses. The generator generates the disc map, and the discriminator classifies it from the ground truth map. We validate the model on a wide range of datasets to address its generalization and extrapolation capability.
3. The paper performs an exhaustive study on eight different datasets (seven public and one challenging AIIMS community camp (private) dataset. The AIIMS dataset is a community camp dataset collected by the All India Institute of Medical Sciences, New Delhi). This has various retinal artifacts such as peri-papillary atrophy and disc hemorrhage. These also include various imaging/camera artifacts such as non-uniform illumination, blurring and lens dust.
4. The proposed algorithm handles small- to large-sized disc images, irrespective of their shape, size, and color, belonging to different datasets (across a variety of cameras used to capture them.) The AIIMS community camp (private) dataset has images captured under a wide variety of difficult imaging and lighting conditions with cheap fundus cameras, and has a wide variety of pathologies.

The rest of the paper is organized as follows. Sec. 2 gives details of the proposed method, in terms of the U-Net-based regressor for localization of the optic disc (Sec. 2.1), the U-Net-based Generative adversarial network for optic disc segmentation (Sec. 2.2), and the Post-processing (Sec. 2.3). Sec. 3 presents results of extensive experiments with a large number of datasets. Sec. 4 concludes the paper.

## 2 Methods

Fig. 3 shows the three main parts of the proposed system. The optic disc segmentation procedure has three main stages.

1. Pre-process the retinal images which include optic disc detection (or localization) in order to extract the region of interest (a square-shaped cropped area containing the optic disc), and augmentation of training samples to avoid model over-fitting.
2. The second major stage of the proposed work is adversarial training that consists of the discriminator and generator networks for disc segmentation, with segmentation and adversarial losses.

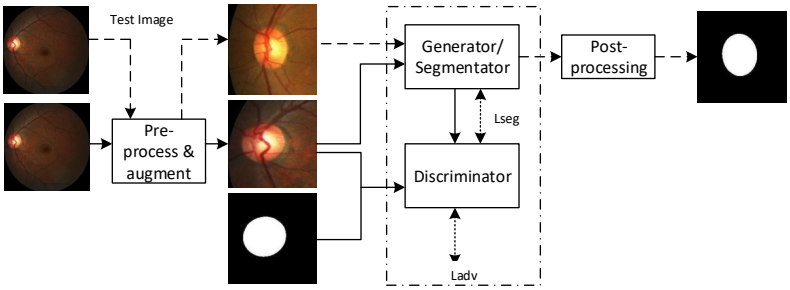



**Fig. 3** The three stages in the proposed method: pre-processing and augmentation (Sec. 2.1), Adversarial system (Sec. 2.2), and post-processing (Sec. 2.3). Solid lines represent the training path and dashed ones, the testing path.

3. Post-processing to smooth and highlight the probable optic disc pixels present at the margin.

Fig. 3 shows the complete block diagram. The following sections gives details in the following sections.

### 2.1 **Pre-processing: Optic Disc Location with a U-Net-based Regressor**

To reduce the computational complexity of the optic disc segmentation task, it is important to locate the disc region in the full retinal image. We propose a fused Distance-Intensity (DI) map to model the distance and intensity variations with respect to the center of the optic disc, to obliterate bright exudate pathologies. The two-dimension fused map is generated with pixel-wise multiplication of normalized distance $D(x,y)^N_{GT}$ and intensity $I(x,y)^N$ map images.

$$DI_{GT}(x,y) = D(x,y)^N_{GT} \times I(x,y)^N \tag{1}$$

$$D(x,y)^N_{GT} = \left(1 - \frac{D(x,y)}{\max_{(x,y)} D(x,y)}\right)^{\gamma} \tag{2}$$

Here, the distance map $D(x,y)_{GT}$ can be calculated by finding the Euclidean distance of each pixel $(x,y)$ from optic disc center $(x_{od}, y_{od})$.

$$I(x,y)^N = \left(\frac{I(x,y)}{\max_{(x,y)} I(x,y)}\right)^{\beta} \tag{3}$$

$\gamma$ and $\beta$ are distance and intensity map decay parameters, respectively. An Euclidean Distance Error (EDE)-based empirical method estimates $\gamma = 7$ and $\beta = 7$ for the challenging AIIMS dataset (with a large number of pathologies, poor illumination and motion blur), as it gives the minimum EDE with an acceptable spread.

Further the fused distance-intensity map is considered as a ground-truth for training a U-Net model using a set of retinal images to predict the grayscale DI

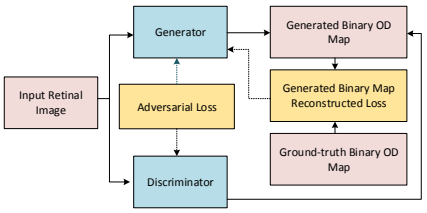



**Fig. 4** A block diagram of the proposed GAN architecture (Sec. 2.2).

map image. The loss function used for back-propagation is Tukey's biweight loss, which is more robust to outliers than mean square error.

The predicted image is used to obtain the optic disc center coordinates using a Laplacian-of-Gaussian (LOG) operator over the entire image. The filter acts as a blob detector (Singh et al. 2018) to detect the brightest optic disc center coordinate from the grayscale predicted image. This work has been verified on a wide range of community-based retinal images consisting of pathologies or imaging artifacts such as non-uniform illumination, noise, motion artifacts, and blurring.

Further, the optic disc localized coordinates are used to crop a square-shaped region of interest and their corresponding optic disc binary masks from the RGB image and ground-truth mask respectively, as shown in Fig. 3. To feed the deep model with images of fixed dimensional input sizes, the size of cropped images from all datasets is kept at $512 \times 512$ pixels, as it empirically gives the best accuracy without affecting the computation complexity. These cropped images are augmented with re-scaling and standard color normalization operations. In order to avoid over-fitting the augmentation involves vertical flipping, zooming, and normalizing to zero mean operations.

### 2.2 **A U-Net-based Generative Adversarial Network for Optic Disc Segmentation**

Fig. 4 outlines the GAN-based proposed method. Considering the output of pre-processing stage (i.e., $512 \times 512$ dimension RGB image) as the training samples, a generative adversarial network (GAN) has been employed to generate the grayscale disc maps. Fig. 5 gives a layer-wise description of the U-Net-based generator, and the discriminator. In this work, the optic disc extraction is posed as a conditional image generation problem where the binary disc map is conditioned by the region of interest-specific RGB sub-images. The transformation (or mapping) has to be learned in such a way that the optic disc and background (containing peri-papillary atrophy or healthy retina) regions can be distinguished in the feature space without any mis-assignment.

The presented model consists of two main networks: an optic disc binary map generation network ($Gen_R$) and a generated binary map Discriminator network ($Dis_B$). The detailed structure of these networks is shown in Fig. 5, where generator network is motivated by U-Net and discriminator follows a series of convolution layer architecture. Further, the reconstruction and adversarial losses have been calculated for back-propagation using above networks.

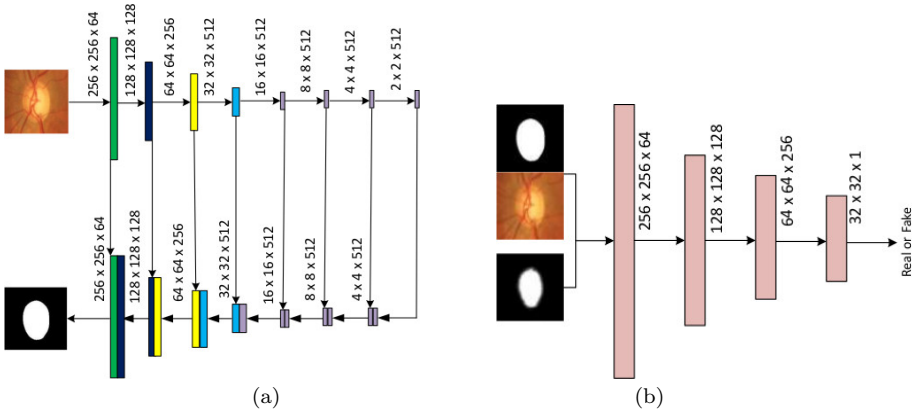



**Fig. 5** The basic GAN structure (Sec. 2.2), with the (a) U-Net-based generator, and (b) the discriminator.

### *2.2.1* ***GAN Loss function***

The proposed model incorporates two losses for disc segmentation i.e., adversarial ($L_{adv}$) and segmentation ($L_{seg}$) loss. Assume $i$ be the input RGB cropped retinal image and $d$ be the binary ground-truth disc map corresponding to it. Moreover, $Pr(i)$ is the predicted gray-scale disc map for image $i$ generated by the segmentation (or generator, D) network and $S(i,d)$ represent the discriminator network (S) output. Both generator (D) and discriminator (S) try to minimize the adversarial loss. The goal of the generator (D) is to generate disc maps($Pr(i)$) that looks as real as the ground-truth disc maps (d). On the other hand, the discriminator S network trains to correctly classify the maps as fake (generated) or real(ground-truth) (Goodfellow et al. 2014). To meet these goals, the generator and discriminator networks need to learn the most discriminating features that can certainly classify between a fake and real disc map. The objective function (or Loss, $L_{adv}$) can be defined as follows:

$$\begin{aligned} L_{adv}(Pr, GT) = {} & E_{i,d\sim P_{data}(i,d)}[log(S(i,d))]+ \\ & E_{x\sim P_{data}(i)}[log(1 - S(i, D(i)))] \end{aligned} \tag{4}$$

The aim is to maximize the discriminator output $S(i,d)$ when true disc map $d$ is fed as an input, while minimizing the output $S(i,D(i))$ when generated map $Pr(i)$ acts as an input. In other terms, the segmentation network is penalized in case the generated disc map is classified as fake by the discriminator, thus the generator fights against the adversarial network. On the contrary, the discriminator is penalized when it classifies the generated disc map as real.

In addition to this, segmentation ($L_{seg}$) loss ensures that the generator network is penalized in case, the generated disc map($Pr(i)$) deviates from the corresponding ground-truth disc image($d$). For disc segmentation, the goal is to focus on the region-of-interest i.e., optic disc region rather than the background. We proposed

a Dice coefficient-based function for segmentation loss calculations.

$$\begin{aligned} L_{seg}(Pr) &= L_{Dice}(d, Pr) \\ &= E_{i,d} \sim P_{data}(i) \left[ \frac{1}{e^{Dice(d,Pr(i))}} \right] \end{aligned} \tag{5}$$

where the Dice coefficient *Dice* can be calculated as:

$$Dice(d, Pr(i)) = 2 \frac{\sum_x^n d_x Pr(i)_x}{\sum_{x=1}^n d_x^2 + \sum_{x=1}^n Pr(i)_x^2} \tag{6}$$

Here $n$ total number of pixels in the image, and $x$ denotes each pixel location, $d_x$ is the ground-truth disc map value at each $x$ and $Pr(i)$ is the predicted disc map respectively. The final objective function is

$$Pr^* = \arg \min_{Pr} \max_{S} \alpha L_{adv}(Pr, S) + L_{seg}(Pr) \tag{7}$$

$\alpha$ represents the adversarial to segmentation network ratio. It specifies the trade-off between reconstruction loss (mismatch between ground-truth and predicted) and for the output to look as real as the input.

### 2.3 **Post-processing (Optic Disc Margin Smoothing)**

In the post-processing stage of disc segmentation, the output of the segmentation network is passed through a series of morphological operations i.e., opening, filtering small objects, and smoothing. The opening operation with a disc shape structuring element is employed to remove the unwanted objects lying around the optic disc boundary. Moreover, to enhance the margin pixels between disc-background and suppress noise, a filtering step is applied. An extra ellipse fitting step is applied with the centroid of segmented region as center of ellipse. The maximum and minimum Euclidean distance between object boundary and centroid acts as the major and minor axis radius respectively. It gives a defined shape for optic disc boundary and results in an increase in overall performance. An Otsu thresholding (Otsu 1979) gives the final binary disc.

## 3 Results

### 3.1 **Datasets**

For our experiments, we use seven publicly available datasets namely Drishti-GS (Sivaswamy et al. 2014), DRIONS, DRIVE, Refugee, RIM (Maninis and Pont-Tuset 2010), RIGA-Magrabi, and RIGA-BinRushed (Almazroa et al. 2018). We also use the challenging AIIMS community camp (private) dataset. This dataset has retinal images collected and ground-truthed by the Ophthalmology Department, All India Institute of Medical Sciences (AIIMS), New Delhi. Table 1 gives a detailed description of these datasets. Experienced ophthalmologists from AIIMS have labeled the optic disc ground truth segmentation. For experimentation purposes dataset has been divided into a 65% training, 5% validation and 30% testing ratio, except Drishti-GS and Refugee which are split into 50:50 ratio, as other algorithms have validated the results on this *de facto* standard split.

**Table 1** Details of Retinal Image Datasets used in the Work

| S.No | Dataset | Train, Test images | Image Dimension | Camera specification |
|---|---|---|---|---|
| 1 | DRIONS | 77, 33 | 600 × 400 | Color analogical fundus camera, digitized with HP-PhotoSmart-S20 high-resolution scanner |
| 2 | DRIVE | 28, 12 | 768 × 584 | Canon CR5 non-mydriatic 3CCD camera with a 45° field of view |
| 3 | Drishti-GS | 50, 51 | 2896 × 1944 | Dilated eye, 30° field of view |
| 4 | Refugee | 400, 400 | 1634 × 1634 | Canon CR-2 device |
| | | | 2124 × 2056 | Zeiss Visucam 500 fundus camera |
| 5 | RIM-ONE v3 | 127, 37 | 2144 × 1424 | Zeiss FF450+camera at 50° field of view |
| | RIM-ONE v2 | 364, 91 | Not available | Nidek AFC-210 with a body of a Canon EOS 5D Mark II, 21.1 megapixels |
| 6 | RIGA-Magrabi | 66, 28 | 2743 × 1936 | Non-mydriatic 3CCD camera (Top-Con TRC NW6) at 45° field of view |
| 7 | RIGA-BinRushed | 136, 58 | 2376 × 3168 | Non-mydriatic 3CCD camera (Top-Con TRC NW6) at 45° field of view |
| 8 | AIIMS (private) | 255, 108 | 1536 × 1584 | Miscellaneous low-resolution cameras with cheap and robust hand-held ophthalmoscopes |

### 3.2 **Evaluation Measures for Optic Disc Segmentation**

ROC parameters are a standard method to encapsulate statistical performance measures. True positives (TP) indicate the total number of disc pixels that are correctly predicted by the algorithm. True negatives (TN) are the number of pixels correctly predicted as background pixels. False positives (FP) are the number of disc region pixels which are wrongly predicted as background. False negative (FN) are background pixels incorrectly predicted as belonging to the optic disc. The three commonly used performance measures for a segmentation problem are the Dice coefficient (F1 score) (Eq. 8), the Jaccard coefficient (Intersection-over-Union (IOU), or the optic disc percentage overlap) (Eq. 9), and the Sensitivity (Recall) (Eq. 10).

$$\textit{Dice Coefficient (F1 score)} = 2 \times \frac{(TP)}{2 \times TP + FN + FP} \tag{8}$$

$$\textit{Intersection over Union (Jaccard coefficient, Disc Overlap)} = \frac{(TP)}{TP + FN + FP} \tag{9}$$

$$\textit{Sensitivity (Recall)} = \frac{(TP)}{TP + FN} \tag{10}$$

Sensitivity is often preferred in medical segmentation problems over another related parameter, Specificity (the relative ratio of TN to all negatives). Sensitivity is often considered more significant since it gives a direct estimate of false negatives (which are crucial in medical diagnosis). To this end, Tables 2, 3, 4, 5 all mention the Dice coefficient, the optic disc overlap percentage, and the sensitivity values.

Tables 6 and 7 present confusion matrices corresponding to the Dice coefficient and the optic disc overlap.

### 3.3 **System Implementation Details**

To build the training dataset, each input RGB retinal image, and its corresponding binary map is cropped and resized with the center of the square-shaped region as coordinates of the optic disc. A $512 \times 512$ pixel size region-of-interest is selected, as the average optic disc size is $1.80 - 1.95mm$ vertically, so the specific dimension incorporates the complete optic disc region including the background area around it. Any down-scaling more than this does not help with improvement in performance parameters (Dice coefficient) and also degrades the visualization quality of results. On the other hand, an up-scale size of $512 \times 512$ results in unnecessarily increased computations without offering many advantages in terms of performance. The above-mentioned steps are performed for each dataset except RIM-ONE which already contains cropped images with 85% area occupied by optic disc (as shown in the second column of Fig 2). Thus, no cropping operation is required and one only needs to resize the RIM-ONE dataset images to $512 \times 512$. Further, image augmentation such as rotation by $30^{\circ}$, flipping, and cropping is performed to prevent the model from over-fitting.

The implementation has been done on a single GPU. An Adam optimizer with a learning rate of 0.0001 is used and step decay is performed after 30 epochs with 0.1 decay. Moreover, a batch size between 8 to 16 is employed depending upon the dataset used for training. We run the adversarial and segmentation network alternatively for 50 epochs, while at the testing time only segmentation is performed. In the final post-processing procedure, image closing morphological operation is done with disc shape structuring element of radius 10. It not only removes small unwanted objects but also smoothes out the edges of the predicted disc map. Moreover, ellipse fitting has been shown to increase the Dice coefficient value by a small margin of 0.4% and 0.6% respectively for Drishti-GS and Refugee datasets.

### 3.4 **Experimental Determination of GAN parameters $\alpha$ and Discriminator Losses**

Eqn. 7 (Sec. 2.2.1) has parameter $\alpha$, which captures the relative trade-off between segmentation and adversarial losses. Table 2 shows that a value of $\alpha = 10$ strikes a good balance between the three common measures of segmentation performance i.e., the Dice coefficient, the disc overlap and the sensitivity. To fine-tune all the learning parameters we initiated the procedure by experimenting with the representative Drishti-GS dataset consisting of 50 training and 51 testing images. Fig. 6 shows the three parameter values corresponding to these losses. A pixel-based loss function gives the best Dice coefficient and optic disc overlap, with an acceptable sensitivity.

**Table 2** GAN parameter $\alpha = 10$ (Eqn. 7: the trade-off between adversarial and segmentation losses, Sec. 2.2.1) strikes a good balance between performance parameters sensitivity, optic disc overlap and Dice coefficient values for the representative Drishti-GS dataset. (Details in Sec. 3.4)

| S.No | $\alpha$ | Sensitivity (%age) | Disc Overlap (%age) | Dice coefficient (%age) |
|---|---|---|---|---|
| 1 | 0 | 97.2685 | 92.24 | 95.94 |
| 2 | 5 | 97.0329 | 92.50 | 96.07 |
| 3 | 10 | 97.72 | **93.78** | **97.31** |
| 4 | 15 | 98.62 | 90.3261 | 94.84 |
| 5 | 20 | 96.47 | 92.44 | 96.07 |

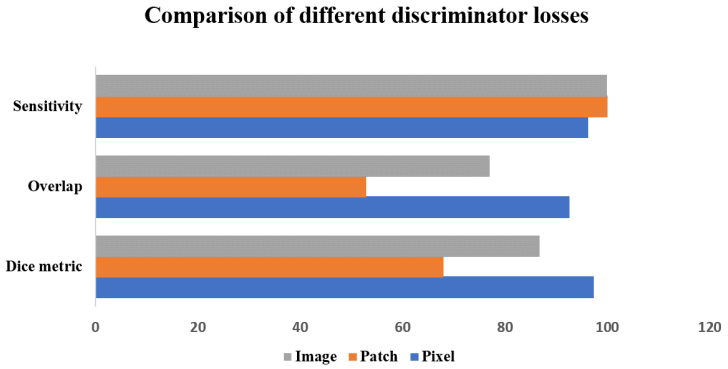



**Fig. 6** A pixel-based discriminator loss function (Sec. 2.2.1) gives the best Dice coefficient and optic disc overlap, with an acceptable sensitivity (Sec. 3.4).

### 3.5 **Inter-Dataset Performance of the Proposed Method**

Table 3 shows all three performance parameters used for evaluating the efficacy of the proposed work. The average Dice coefficient, disc overlap, and sensitivity obtained are 92.93%, 91.01%, 94.83% respectively. Fig. 7 shows some representative test images. Here red- and blue-colored markings show the ground truth and predicted optic disc boundaries, respectively. Fig. 7(a) shows 5 representative examples of atrophy-affected discs where the textural properties of the region just outside the disc boundary vary from the background. For the 5 representative examples of Fig. 7(b), each disc boundary is diffused into the background, thus making the segmentation process tough. In spite of this, the proposed algorithm is able to segment the disc precisely. Next, the third row (Fig. 7(c)) shows oval-shaped discs from the DRIONS dataset with different color illuminations. Fig. 7(d) shows some miscellaneous challenging examples, where the disc boundary is not even clearly visible to an observer and non-uniform illumination persists in the image. The proposed method is able to segment the optic disc margin even in such difficult cases with clinical artifacts. as improper illumination, segmenting optic

**Table 3** The proposed adversarial optic disc segmentation method gives consistently good results across eight different datasets: sensitivity, optic disc overlap and the Dice coefficient

| S.No | Dataset | Sensitivity (%age) | Disc Overlap (%age) | Dice coefficient (%age) |
|---|---|---|---|---|
| 1 | Drishti-GS (Liu et al. 2019) | 97.72 | 93.78 | 97.31 |
| 2 | RIM (Sharma et al. 2019) | 94.23 | 85.72 | 91.88 |
| 3 | Refugee (Jiang et al. 2019) | 93.45 | 89.41 | 95.12 |
| 4 | DRIVE (Wang et al. 2019) | 86.63 | 98.94 | 86.63 |
| 5 | AIIMS [private] | 94.27 | 85.93 | 91.91 |
| 6 | DRIONS | 97.22 | 85.95 | 92.34 |
| 7 | RIGA-BinRushed | 93.86 | 89.695 | 94.51 |
| 8 | RIGA-Magrabi | 96.25 | 90.83 | 95.17 |
| | All datasets | 97.72 | 93.78 | 97.31 |

disc becomes a challenging task.

The generalizability of the proposed generative adversarial network has been described with eight different datasets. To better investigate the dataset variability problem along with the performance of the proposed approach, for each dataset the prediction capability is analyzed in independent local and external generalization scenarios. In the first part of the experimentation, a split of 70:30 ratio for each dataset is done for training and test the corresponding sets. The learning model is trained over each local dataset (train and testing data belongs to the same dataset) with tuned parameters. Further, in the second step of experimentation, the trained model corresponding to each dataset is validated across other remaining datasets. In the inter-dataset (train and testing data belongs to the different datasets) setup the split ratio is kept the same for each dataset. The motivation for such experimentation is to compare both setups (local and inter-dataset) and thus to perform better for any data irrespective of which the dataset is extracted.

### 3.6 **Comparative Performance with The State-of-the-art across Datasets**

In addition to inter-dataset performance (Sec. 3.5), we compare our approach with the state-of-the-art across various reported performance measures across popular datasets. Table 4 shows a comparison for the popular for the Drishti-GS dataset. The first method by (Liu et al. 2019) uses the full retinal image of dimension $512 \times 512$ as the region of interest for disc extraction. It is based on the GAN model with a patch-based discriminator for classifying a $64 \times 64$ sized patch into a disc or the background region. Next, an image processing technique of constructing atlases of optic disc region to find the best match to the tested optic disc image is suggested in (Sharma et al. 2019) with obtained average sensitivity and overlap of 97.38% and 93.44% respectively. Moreover, some deep learning approaches (Jiang

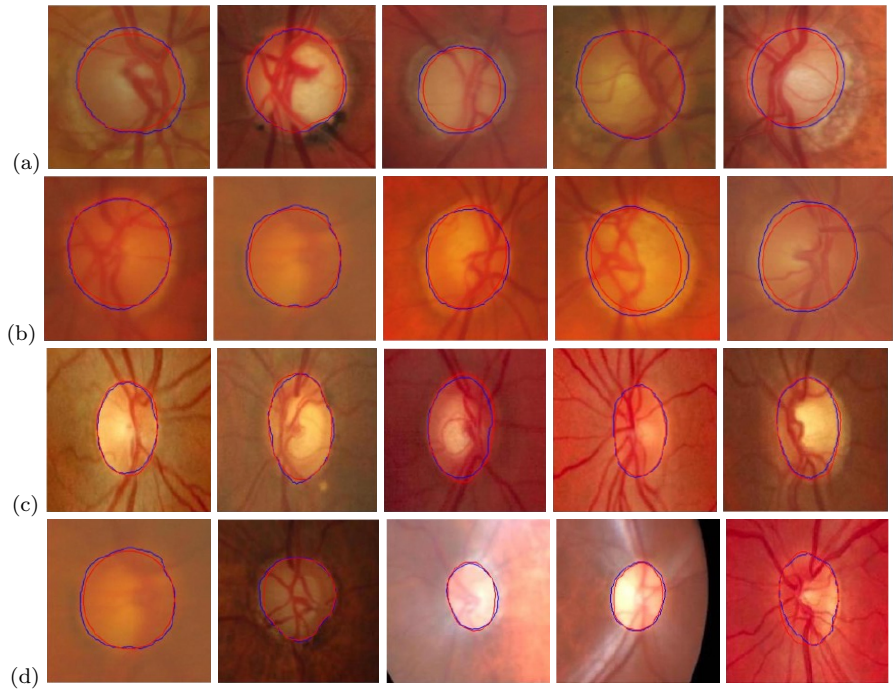



**Fig. 7** Some representative results across challenging inputs: (a) peri-papillary atrophy, (b) diffuse disc boundary, (c) oval disc with varying texture, and (d) pathologies and other artifacts. (The predicted optic disc boundary is in blue.)

et al. 2019), (Wang et al. 2019) also rely on patch-based generative adversarial networks which result in Dice coefficient values of 97.0% and 96.5% respectively. Finally, a multi-resolution approach using a deep convolutional network results in a Dice coefficient of 97.13%. In comparison, our proposed network achieves the benchmark performance when compared with state-of-art methods and obtains an average sensitivity, overlap, and Dice coefficient of 97.72%, 93.78%, and 97.31% respectively on all datasets. In another experimentation where the training samples from all datasets are used together for training and tested on the remaining images results in much lower performance with average parameters of 89.86%, 77.44%, and 85.63% respectively.

Table 5 shows a similar comparison for the Refugee dataset, across all three performance parameters. It is clear from the results that the proposed generative network outperforms all the state-of-art algorithms with obtained average Dice coefficient and overlap of 95.12% and 89.39% respectively, and an acceptably large value of the sensitivity.

### 3.7 **Generalization Capability of the Proposed Method**

A Dice coefficient-based confusion matrix is a good measure of performance in segmentation tasks (Bertels et al. 2019). In order to visualize the generalization

**Table 4** A comparison of performance parameters (sensitivity, overlap and the Dice coefficient) with state-of-the-art methods, for the Drishti-GS dataset

| S.No | Detection method | Sensitivity (%age) | Disc Overlap (%age) | Dice coefficient (%age) |
|---|---|---|---|---|
| 1 | Liu et al. (Liu et al. 2019) | | 93.8 | 96.7 |
| 2 | Atlas method (Sharma et al. 2019) | 97.38 | 93.44 | 96.0 |
| 3 | Jiang et al. (Jiang et al. 2019) | - | - | 97.0 |
| 4 | Wang et al. (Wang et al. 2019) | - | - | 96.5 |
| 5 | Multi-resolution (Mohan et al. 2019) | - | - | 97.13 |
| 6 | M-Net (Fu et al. 2018) | - | 85.88 | 96.58 |
| 7 | Proposed Approach | **97.72** | **93.78** | **97.31** |

**Table 5** A comparison of performance parameters (sensitivity, overlap and the Dice coefficient) with state-of-the-art methods, for the Refugee dataset

| S.No | Detection method | Sensitivity (%age) | Disc Overlap (%age) | Dice coefficient (%age) |
|---|---|---|---|---|
| 1 | U-Net (Ronneberger et al. 2015) | | 83.12 | 93.08 |
| 2 | M-Net (Fu et al. 2018) | - | 84.02 | 93.59 |
| 3 | Multi-Task et al. (Chen et al. 2016) | - | 84.36 | 94.01 |
| 4 | Wang et al. (Wang et al. 2019) | - | - | 94.60 |
| 5 | ET-Net (Zhang et al. 2019) | - | 86.7 | 95.2 |
| 6 | Proposed Approach | 93.4461 | 89.40 | **95.12** |

capability of the proposed network, we first show a Dice coefficient-based confusion matrix proposed for eight different datasets (Table 6). Each row and column represent the training and testing datasets respectively. The Dice coefficient value resulting from the train-test combination is shown in each cell. In general, one expects the same dataset used in training and testing to give the best Dice coefficients, on an average. The AIIMS, DRIVE, DRIONS and RIGA-Magrabi datasets are exceptions. The AIIMS private dataset seems to perform better when trained on RIGA-BinRushed rather than on its own dataset. The reason for such performance is observed that the AIIMS-private dataset is limited in size and contains a large variety of images. Further, it shows high similarity with the RIGA-BinRushed dataset in terms of disc size, location, and color. In the case of the other three, the

**Table 6** Dice coefficient-based confusion matrix of 8 different datasets (Sec. 3.7).

| | | Testing Datasets | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | AIIMS | RIGA-BR | DRIONS | Drishti-GS | DRIVE | Refugee | RIGA-Mgr | RIM-ONE |
| Training Datasets | AIIMS | 0.880 | 0.870 | 0.710 | 0.740 | 0.790 | 0.887 | 0.811 | 0.596 |
| | RIGA-BR | **0.920** | **0.960** | 0.890 | 0.920 | 0.780 | 0.930 | 0.930 | 0.660 |
| | DRIONS | 0.910 | 0.920 | 0.940 | 0.930 | 0.840 | 0.900 | 0.860 | 0.720 |
| | Drishti-GS | 0.870 | 0.930 | **0.945** | **0.973** | 0.843 | 0.806 | **0.950** | 0.850 |
| | DRIVE | 0.840 | 0.880 | 0.680 | 0.730 | 0.800 | 0.880 | 0.820 | 0.370 |
| | Refugee | 0.870 | 0.940 | 0.920 | 0.920 | **0.870** | **0.950** | 0.920 | 0.690 |
| | RIGA-Mgr | 0.896 | 0.896 | 0.930 | 0.940 | 0.800 | 0.770 | 0.940 | 0.859 |
| | RIM-ONE | 0.820 | 0.850 | 0.920 | 0.950 | 0.780 | 0.760 | 0.834 | **0.920** |

**Table 7** Optic disc overlap-based confusion Matrix of 8 different datasets. (Sec. 3.7)

| | | Testing Datasets | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | AIIMS | RIGA-BR | DRIONS | Drishti-GS | DRIVE | Refugee | RIGA-Mgr | RIM-ONE |
| Training Datasets | AIIMS | 78.18 | 77.94 | 54.67 | 59.72 | 65.65 | 80.28 | 69.45 | 42.83 |
| | RIGA-BR | **85.93** | **91.47** | 79.939 | 84.76 | 68.91 | 86.37 | 87.88 | 50.98 |
| | DRIONS | 84.43 | 85.92 | 88.47 | 86.67 | 73.90 | 82.32 | 77.40 | 58.15 |
| | Drishti-GS | 77.34 | 87.10 | **89.70** | **93.78** | 74.01 | 69.45 | **90.83** | 75.39 |
| | DRIVE | 72.82 | 79.71 | 52.43 | 58.41 | 72.49 | 78.78 | 70.85 | 23.99 |
| | Refugee | 79.66 | 89.64 | 84.94 | 86.30 | **89.40** | **88.42** | 85.47 | 54.41 |
| | RIGA-Mgr | 81.76 | 82.08 | 87.61 | 89.44 | 67.74 | 63.94 | 89.76 | 76.32 |
| | RIM-ONE | 70.62 | 75.95 | 85.95 | 89.99 | 66.89 | 63.69 | 73.23 | **85.71** |

difference from the performance on the original dataset is not very significant. Table 7 presents a similar exercise for the optic disc overlap-based confusion matrix for all eight datasets. As with the Dice coefficient, one has a similar explanation for why the AIIMS dataset gives the best performance when trained on the RIGA-BR dataset (and DRIVE, on the Refugee dataset). The discrepancy for the DRIONS and Riga-Magrabi is not very significant.

An interesting exercise in generalization across datasets orders them in decreasing order of size (Refugee, RIM-ONE, AIIMS, RIGA-BR, Drishti-GS, DRIONS,

RIGA-Magrabi and DRIVE). Let us consider the Dice coefficients in Table 6 as an example. The Refugee dataset gives the highest test score (0.95) for its own training set. We put Refugee in a new collection of reduced number of datasets. RIM-ONE has a similar behaviour: we add this to the reduced collection. The next largest dataset AIIMS gives its best performance when trained on RIGA-BR. We add RIGA-BR to our reduced collection. RIGA-BR gives its best performance when trained on its own training dataset. This is already in the reduced collection. Drishti-GS shows a similar behaviour on its own training dataset: this is added to the reduced collection. The next in line are the DRIONS and RIGA-Magrabi datasets, which give best results with training on Drishti-GS, which is already in the reduced collection. The last is DRIVE, which gives its best performance with the Refugee dataset, which is already in the reduced collection. Thus, reduced collection of 4 training datasets with a sufficient number of training samples, helps the GAN-based system generalize well over a wider range of datasets (8 in this example).

## 4 Conclusions

Optic disc segmentation is a vital step for diagnosing diseases such as peri-papillary atrophy and hemorrhages present near the disc. Blindness-causing conditions such as glaucoma benefit from non-invasive techniques such as computation of the cup-to-disc ratio. We propose a unified U-Net-based optic disc region locator, and a U-Net-based generative adversarial network for the optic disc segmentation. We have evaluated the proposed method on as many as seven public dataset, and one challenging (private) AIIMS community camp dataset. The novel experimentation performed by the proposed algorithm is in terms of inter-dataset train and testing. It has been observed that with the combination of four training datasets, the proposed method is able to obtain maximum average Dice coefficient, disc overlap, and sensitivity of 97.31%, 93.78%, and 97.72% respectively on eight datasets taken together. In addition to this, for the Drishti-GS and Refugee datasets (on which other methods typically report results), the proposed work achieves a state-of-art performance with Dice coefficient values of 97.31% and 95.12%, respectively.

## Declarations

**Conflict of interest** The authors declare no competing interests.

## References

Abdullah M, Fraz MM, Barman SA (2016) Localization and Segmentation of Optic Disc in Retinal Images using Circular Hough Transform and Grow-cut Algorithm. PeerJ 4

Acharjya K, Sahoo G, Sharma SK (2019) An Extensive Review on Various Fundus Databases use for development of computer-aided diabetic retinopathy screening tool. In: Wang J, et al. (eds) Soft Computing and Signal Processing, Springer Nature, pp 407 – 418

Almazroa A, Alodhayb S, Osman E, Ramadan E, Hummadi M, Dlaim M, Alkatee M, Raahemifar K, Lakshminarayanan V (2018) Retinal Fundus Images for Glaucoma Analysis: the RIGA Dataset. In: Proc. SPIE Medical Imaging 2018: Imaging Informatics for Healthcare, Research, and Applications, pp 55 – 62

Aquino A, Gegúndez-Arias ME, Marín D (2010) Detecting the Optic Disc Boundary in Digital Fundus Images using Morphological, Edge Detection, and Feature Extraction Techniques. IEEE Transactions on Medical Imaging 29(11):1860 – 1869

Bertels J, Eelbode T, Berman M, Vandermeulen D, Maes F, Bisschops R, Blaschko MB (2019) Optimizing the Dice Score and Jaccard Index for Medical Image Segmentation: Theory and Practice. In: Proc. Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp 92 – 100

Chen H, Qi X, Yu L, Heng PA (2016) DCAN: Deep Contour-Aware Networks for Accurate Gland Segmentation. In: Proc. IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)

Cheng J, Liu J, Xu Y, Yin F, Wong DWK, Tan N, Tao D, Cheng C, Aung T, Wong TY (2013) Superpixel Classification based Optic Disc and Optic Cup Segmentation for Glaucoma Screening. IEEE Transactions on Medical Imaging 32(6):1019 – 1032

Dada T, Coote M (2010) Clinical Evaluation of Optic Nerve Head. International Society of Glaucoma Surgery.

Fraga A, Barreira N, Ortega M, Penedo MG, Carreira MJ (2012) Precise Segmentation of the Optic Disc in Retinal Fundus Images. In: Moreno-Díaz R, Pichler F, Quesada-Arencibia A (eds) In Proc. Computer Aided Systems Theory (EUROCAST), pp 584 – 591

Fu H, Cheng J, Xu Y, Wong DWK, Liu J, Cao X (2018) Joint Optic Disc and Cup Segmentation based on Multi-Label Deep Network and Polar Transformation. IEEE Transactions on Medical Imaging 37(7):1597 – 1605

Fu Y, Chen J, Li J, Pan D, Yue X, Zhu Y (2021) Optic Disc Segmentation by U-Net and Probability Bubble in Abnormal Fundus Images. Pattern Recognition 117:107971

Gao Y, Yu X, Wu C, Zhou W, Wang X, Chu H (2019a) Accurate and Efficient Segmentation of Optic Disc and Optic Cup in Retinal Images Integrating Multi-View Information. IEEE Access 7:148183 – 148197

Gao Y, Yu X, Wu C, Zhou W, Lei X, Zhuang Y (2019b) Automatic Optic Disc Segmentation based on Modified Local Image Fitting Model with Shape Prior Information. Journal of Healthcare Engineering 2019:1 – 10

Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2014) Generative Adversarial Networks. In: Proc. Advances in Neural Information Processing Systems (NeurIPS)

Jiang Y, Tan N, Peng T (2019) Optic disc and cup segmentation based on deep convolutional generative adversarial networks. IEEE Access 7:64483–64493, DOI 10.1109/ACCESS.2019.2917508

Jiang Y, Duan L, Cheng J, Gu Z, Xia H, Fu H, Li C, Liu J (2020) Jointrcnn: A Region-based Convolutional Neural Network for Optic Disc and Cup Segmentation. IEEE Transactions on Biomedical Engineering 67(2):335 – 343

Joshi GD, Sivaswamy J, Krishnadas SR (2011) Optic Disk and Cup Segmentation from Monocular Color Retinal Images for Glaucoma Assessment. IEEE Transactions on Medical Imaging 30(6):1192 – 1205

Liu Y, Fu D, Huang Z, Tong H (2019) Optic Disc Segmentation in Fundus Images using Adversarial Training. IET Image Processing 13(2):375 – 381

Lowell J, Hunter A, Steel D, Basu A, Ryder R, Fletcher E, Kennedy L (2004) Optic Nerve Head Segmentation. IEEE Transactions on Medical Imaging 23(2):256 – 264

Lu S (2011) Accurate and Efficient Optic Disc Detection and Segmentation by a Circular Transformation. IEEE Transactions on Medical Imaging 30(12):2126 – 2133

Luo L, Xue D, Pan F (2021) Optic Joint Optic Disc and Optic Cup Segmentation based on Boundary Prior and Adversarial Learning. International Journal of Computer Assisted Radio Surgery 16:905 – 914

Maninis KK, Pont-Tuset J (2010) Retinal Databases. http://www.vision.ee.ethz.ch/ cvlsegmentation/driu/downloads.html

Mohan D, Kumar J, Seelamantula CS (2019) Optic Disc Segmentation using Cascaded Multiresolution Convolutional Neural Networks. In: Proc. IEEE International Conference on Image Processing (ICIP), pp 834 – 838

Otsu N (1979) A Threshold Selection Method from Gray-Level Histograms. IEEE Transactions on Systems, Man and Cybernetics 9(1):62 – 66

Pachade S, Porwal P, Kokare M, Giancardo L, Mériaudeau F (2021) NENet: Nested EfficientNet and Adversarial Learning for Joint Optic Disc and Cup Segmentation. Medical Image Analysis 74(4):102253

Pande Darma Suardika IG, Dendi Maysanjaya IM, Antara Kesiman MW (2022) Optic Disc Segmentation based on Mask R-CNN in Retinal Fundus Images. In: Proc. International Conference on Biomedical Engineering (IBIOMED), pp 71 – 74

Qin P, Wang L, Lv H (2019) Optic Disc and Cup Segmentation based on Deep Learning. In: Proc. IEEE Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), pp 1835 – 1840

Ronneberger O, Fischer P, Brox T (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Proc. Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp 234 – 241

Sharma A, Aggarwal M, Dutta Roy S, Gupta V, Vashist P, Sidhu T (2019) Optic Disc Segmentation in Fundus Images using Anatomical Atlases with Non-rigid Registration. In: Arora C, Mitra K (eds) Computer Vision Applications, Springer, pp 14 – 27

Singh VK, Rashwan H, Akram F, Pandey N, Sarker MMK, Saleh A, Abdulwahab S, Maaroof N, Romani S, Puig D (2018) Retinal Optic Disc Segmentation using Conditional Generative Adversarial Network. arXiv:1806.03905 [cs.CV]

Sivaswamy J, Krishnadas SR, Joshi GD, Jain M, Ujjwal, Syed Tabish A (2014) Drishti-GS: Retinal Image Dataset for Optic Nerve Head (ONH) Segmentation. In: Proc. IEEE International Symposium on Biomedical Imaging (ISBI)

Trucco E, MacGillivray T, Xu Y (2019) Computational Retinal Image Analysis: Tools, Applications and Perspectives. Mara Conner

Veena HN, Muruganandham A, Senthil Kumaran T (2022) A Novel Optic Disc and Optic Cup Segmentation Technique to Diagnose Glaucoma using Deep Learning Convolutional Neural Network over Retinal Fundus Images. Journal of the King Saud University - Computer and Information Sciences 34:6187 – 6198

Wang S, Yu L, Yang X, Fu C, Heng P (2019) Patch-based Output Space Adversarial Learning for Joint Optic Disc and Cup Segmentation. IEEE Transactions

on Medical Imaging 38(11):2485 – 2495
Wang Y, Yu X, Wu C (2022) An Efficient Hierarchical Optic Disc and Cup Segmentation Network Combined with Multi-Task Learning and Adversarial Learning. Journal of Digital Imaging 35
Zahoor MN, Fraz MM (2017) Fast optic disc segmentation in retina using polar transform. IEEE Access 5:12293–12300
Zhang Z, Fu H, Dai H, Shen J, Pang Y, Shao L (2019) Et-net: A Generic Edge-Attention Guidance Network for Medical Image Segmentation. CoRR abs/1907.10936
Zou B, Liu Q, Yue K, Chen Z, Chen J, Zhao G (2019) Saliency-based segmentation of optic disc in retinal images. Chinese Journal of Electronics 28(1):71 – 75