Research Article

Optic Disc and Optic Cup Segmentation for Glaucoma Detection from Blur Retinal Images Using Improved Mask-RCNN

Tahira Nazir ⓘ,[1] Aun Irtaza ⓘ,[1] and Valery Starovoitov ⓘ[2]

[1]*Department of Computer Science, University of Engineering and Technology, Taxila 47050, Pakistan*
[2]*United Institute of Informatics Problems of the National Academy of Sciences of Belarus, Minsk 220012, Belarus*

Correspondence should be addressed to Tahira Nazir; tahira.nazir77@gmail.com

Glaucoma is a fatal eye disease that harms the optic disc (OD) and optic cup (OC) and results into blindness in progressed phases. Because of slow progress, the disease exhibits a small number of symptoms in the initial stages, therefore causing the disease identification to be a complicated task. So, a fully automatic framework is mandatory, which can support the screening process and increase the chances of disease detection in the early stages. In this paper, we deal with the localization and segmentation of the OD and OC for glaucoma detection from blur retinal images. We have presented a novel method that is Densenet-77-based Mask-RCNN to overcome the challenges of the glaucoma detection. Initially, we have performed the data augmentation step together with adding blurriness in samples to increase the diversity of data. Then, we have generated the annotations from ground-truth (GT) images. After that, the Densenet-77 framework is employed at the feature extraction level of Mask-RCNN to compute the deep key points. Finally, the calculated features are used to localize and segment the OD and OC by the custom Mask-RCNN model. For performance evaluation, we have used the ORIGA dataset that is publicly available. Furthermore, we have performed cross-dataset validation on the HRF database to show the robustness of the presented framework. The presented framework has achieved an average precision, recall, *F*-measure, and IOU as 0.965, 0.963, 0.97, and 0.972, respectively. The proposed method achieved remarkable performance in terms of both efficiency and effectiveness as compared to the latest techniques under the presence of blurring, noise, and light variations.

1. Introduction

Glaucoma harms the optic nerve (ON) because of the imbalance of intraocular pressure within the eye. The affected nerve fibers result in deterioration of the retinal layer and give rise to the enlarged OD, that is, the part of the retina, and the OC is the main portion of the OD. Glaucoma is typically analysed by attaining the medical history of patients, determining intraocular pressure (IOP), conducting visual field loss tests, and manual assessment of OD employing ophthalmoscopy to investigate the shape and color of the ON [1]. The cup-to-disc ratio (CDR) is one of the key structural image cues reflected for glaucoma identification. The CDR compares the diameter of OC with the diameter of OD; less than 0.5 CDR considers the normal value [2]. So, timely detection of disease can avoid blindness

[3]. Hence, clustering of the malicious area is not only advantageous for additional rigorous medical analysis by the ophthalmologist but also useful for designing the automated systems for disease categorization [4]. Initially, experts identify eye abnormalities through the manual examination of the glaucoma regions, by calculating the CDR, diameter, and boundaries variations [5]. However, due to the lack of available experts, timely identification of the eye abnormality is typically delayed [6], whereas early detection and treatment of the disease can save the victim from complete blindness. To tackle with mentioned challenges, the research community is targeting disease identification via Computer-Aided Diagnosis (CAD) based solutions.

In research, deep learning (DL) based approaches [3, 4, 7–20] have been utilized to identify glaucoma signs from the retinal images. In [21], an end-to-end RCNN

method for joint OD and OC segmentation was proposed. In joint-RCNN, OD and OC proposal networks were used to create bounding box (BB) proposals for OD and OC, respectively. The presented technique is computationally complex because it utilizes two distinct RCNNs to calculate the BBs of ROI regions. Therefore, a more reliable technique is required which can detect glaucoma affected region efficiently. In [22], a region-based pixel density calculation method was used for OD localization. Afterward, OD segmentation was performed through the Circular Hough Transform method. The procedure is efficient and robust to the segmentation of OD; however, its recognition performance is disturbed over the images having pathological distractions. In [3], the authors adapted DenseNet into a U-Net shaped framework for OD and OC segmentation. The method was comprised of three major phases, (i) preprocessing, (ii) FC-DenseNet model designing, and (iii) segmentation of OD and OC. In the first, the green channel was extracted from RGB images; after that, OD region within two OD diameters has been collected, which were utilized for the model training. In the second phase, the model has been built which was composed of three blocks, that is, dense and transition down and up. In the final phase, refinement was performed for the extraction of OD and OC through Softmax operation. The performance of the method [3] was evaluated over five different datasets and has achieved good results with a short testing time. However, the method [3] has some shortcomings: (i) calculation of OD centre being dependent on GT data, (ii) high training time, and (iii) training being done on small set. In [18], an eighteen-layer CNN architecture was proposed for glaucoma localization, which has two main components: (i) convolutional and max-pooling layer phase (ii) and fully connected layer phase. The method has evaluated 1426 images and achieved an accuracy of 98.13%. However, the method in [18] degrades performance on unseen samples and may not detect glaucoma at early stages.

In [15], Lu et al. presented a weekly and semisupervised segmentation method based on the Modified U-Net model for OD segmentation. Initially, the GrabCut technique was employed for the generation of the GTs. The U-Net model was improved by minimizing the original U-shape structure by adding a 2-dimensional convolutional layer at the end of the convolutional layer. This method needs a smaller amount of training, however, indicating less accuracy than other methods due to the lack of GTs. Elangovan et al. [23] have proposed the approach for glaucoma identification based on CNN which was consisted of 18 layers. The technique has different phases: preprocessing, key points computation, and classification. Initially, image resizing and data augmentation were performed; furthermore, rotation augmentation was applied to enhance the number of samples. Features were extracted through CNN which has four convolutional, two pooling, and a fully connected layer. For performance evaluation of the method, different datasets were used, namely, ORIGA, DRISHTI-GS1, RIM-ONE2, LAG, and ACRIMA. In [24], authors have presented the attention-based CNN (AG-CNN) technique for glaucoma recognition. In this paper [24], the authors have created a new database called

large-scale attention-based glaucoma, which has a total of 11760 retinal images. All images were marked with negative or positive glaucoma. The AG-CNN method was comprised of two main stages; in the first phase, the attention prediction subnet was used to learn the ROI of glaucoma and then predicted the attention map. Secondly, the predicted map was utilized in the localized region, and then the feature map of this subnet was visualized to locate the pathological region. Lastly, the located region was merged with the anticipated attention to combining the input and subnet of glaucoma key points, for computing the binary labels of glaucoma. The method in [24] shows good performance and reduces the redundancy of fundus images; however, the method depends on the attention prediction subnet.

Existing techniques perform well over the standard datasets but not generalized well to real-world scenarios. The main reasons for performance degradation are the occurrence of blurring, noise, and light variations during the image capturing process, while the standard datasets are acquired in the control environment. In this work, our main motivation is to propose such techniques that can localize and segment the fundus samples under the presence of such factors. We have selected standard datasets like ORIGA and HRF databases which contain light variations and noise effects but lack the presence of blurriness. So, in this work, we have added blurriness in samples of mentioned datasets and proposed a novel technique, namely, Densenet-77 based [25] customized Mask-RCNN to detect and segment the OC and OD from fundus samples. The following are the main contributions of our work:

(1) The proposed method can precisely segment the OD and OC for glaucoma diagnosis from retinal images under the presence of blurring, noise, and light variations in input images.

(2) We have created the annotations which are essential for the training of the proposed model because available datasets do not have a BB and mask GTs.

(3) Accurate localization and segmentation of OD and OC due to effective region proposal network of Custom Mask-RCNN as it works in an end-to-end manner.

(4) Extensive results perform over challenging dataset ORIGA to show the robustness of the presented framework. Moreover, we have performed cross-dataset validation over the HRF database to demonstrate the generalization power of our technique to real-world scenarios.

2. Materials and Methods

The retinal images collected from different clinics can contain various artifacts like blurring, noise, out-of-focus images, or light variations, which must be removed to enhance the segmentation performance of the system. In our paper, we have employed the feature level set technique for correcting the bias field and applied the median filter to minimize the noisy effects from retinal images.

2.1. Preprocessing. The augmentation step is employed to increase the image samples in terms of data diversity. For this purpose, the input images are rotated at the angles of 0°, 90°, 180°, and 270° degrees, and Gaussian blur [26] is used over them to add blurriness.

Furthermore, we have generated the annotations for OC and OD regions. The GT mask along the retinal image is needed to detect glaucoma regions, that is, OD and OC for the training procedure. We used the VGG Image Annotator [27] tool to create a polygon mask for every image. Figure 1 presents a sample of images and related mask images. The annotations are saved in a JSON file that contains the set of polygon points for OD and OC regions. This file is utilized to generate a mask image related to each retinal image.

2.2. Localization and Segmentation of OD and OC Using Custom Mask-RCNN. Our objective is the automated detection and segmentation of OD and OC from fundus images with complicated backgrounds and under the presence of postprocessing operations without any human involvement. We aimed to identify glaucoma affected and nonaffected areas from a given sample by utilizing the Mask-RCNN [28] approach. The introduced approach (as shown in Figure 2) comprises the following steps: (1) key points computation, (2) region proposal network (RPN), (3) region of interest (ROI) classifier and bounding box regressor (BBR), and (4) OD and OC segmentation. The comprehensive explanation of all steps is described in the following.

2.3. Features Extraction. In our approach, we have used DenseNet-77 at the feature extraction level of the Mask-RCNN to compute the key points from a given sample. Utilizing DenseNet-77 for features computation exhibits an improvement in both the segmentation accuracy and computational complexity. The starting layers compute low-level key points from the images, that is, edge and corner information, and the deep layers calculate high-level key points, that is, structure and chrominance information. The extracted feature map is more enhanced through the FPN that calculates the key points with improved object representation at diverse scales for the RPN module.

DenseNet [25] model is the advanced or improved form of Resnet, where the current layer belongs to all other layers. DenseNet contains the set of dense blocks, which remain consecutively linked with each other by using the extra convolutional and pooling layers among consecutive dense blocks. DenseNet can present the complex transformations which result in improving the issue of the absence of the target's position information for the top-level key points to some degree. DenseNet reduces the total parameters which makes them cost-effective. Furthermore, it supports the calculation of key points and encourages them to recycle, which makes them more suitable for region classification in retinal images. So, in this paper, we have employed the DenseNet-77 as a feature extractor for Mask-RCNN. The explanation of the DenseNet-77 model is shown in Figure 3. It also signifies the query sample size to be accommodated before computing key points from the allocated layer. The complete flow or description of the proposed method is presented in Algorithm 1.

The DenseNet-77 has two potential differences from traditional DenseNet: (i) it has a smaller number of parameters than the actual model and (ii) the layers within all dense block are adjusted to overcome with the computational complexity. Table 1 presents the detail of the training parameters for the Custom CenterNet.

2.4. Region Proposal Network. The calculated feature map from the previous step is passed as input to the RPN module to produce ROIs. Our work has used a 3×3 convolutional layer to scan the input sample by a sliding window to produce appropriate anchors that show the BB with varying scales and dispersed over the whole input sample. RPN module generates almost 20 k anchors of varying scales and dimensions which relate to each other to cover the entire image. A classifier is employed to decide whether an anchor holds the object or background (fg/bg). The BBR produces BBes according to the set intersection-over-union (IoU) value. Precisely, if the IoU value for an anchor is greater than 0.7 holding a GT box, then it is categorized positive; otherwise, it is marked as negative. The RPN module may generate overlapped areas; therefore, a nonmaximum suppression technique is used to keep the regions with the highest foreground score and discard the remaining insignificant parts. The final RoIs are passed to the succeeding step for performing classification.

2.5. ROI Classification and Bounding Box Regression. This module accepts two types of inputs which are the introduced RoI and feature map from previous steps. In contrast to the RPN module, this part is deeper and assigned a specific class to RoIs like glaucoma or nonglaucoma and improves the location of BB. The main objective of the BBR is to improve the location and dimension of the BB to correctly capture the glaucoma region. Typically, the margins of ROI do not overlap with the granularity of the feature map because of the reason that the computed feature map is shrunk k times from the actual image size. For resizing the feature maps, the ROIAlign layer is utilized to compute fixed-length key points vectors for random-sized candidate areas. For resizing, the ROIAlign layer employs the bilinear interpolation to evade misalignment problems that occurred in the ROI pooling layer which utilizes the quantization process.

2.6. Segmentation Mask. This module accepts positive marked ROIs by the ROI classifier as input and computes the segmentation mask with the dimension of 28×28 shown by floating values that hold more details as compared to binary masks. The GT masks are resized to 28×28 to compute the loss using the identified mask in the training step, which is later scaled up to match the actual size of the ROI BB to show the final mask.

(a)   (b)
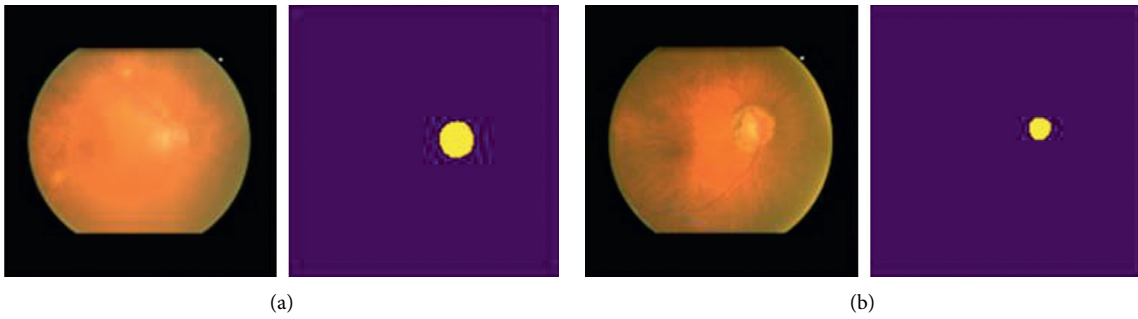
FIGURE 1: Sample original images and corresponding GT masks. (a) Optic disc. (b) Optic cup.
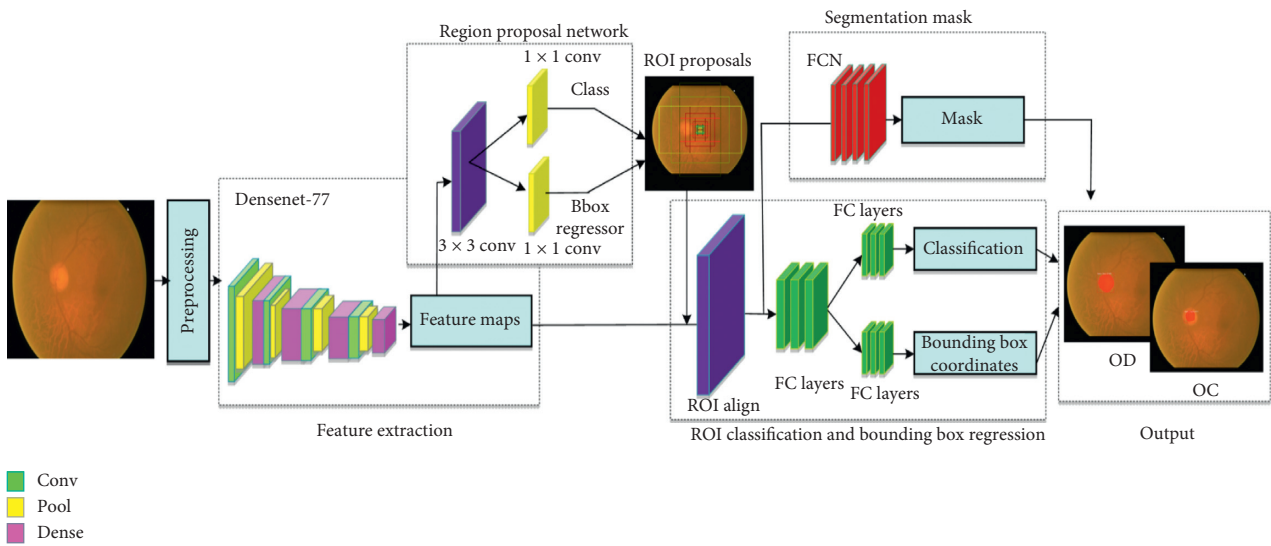

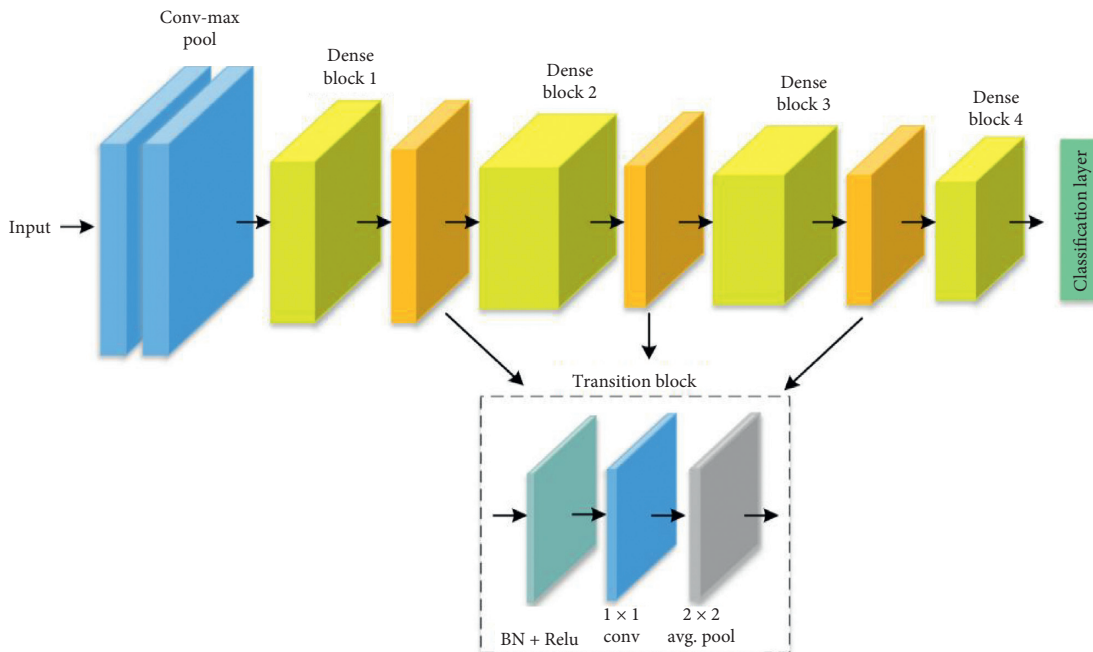FIGURE 2: Architecture of the proposed approach.

FIGURE 3: Structure of DenseNet-77.

```
START
 INPUT: NS, annotation (orientation)
OUTPUT: Localized RoI, CMskDenseNet-77
  NS : Total image samples containing.
  annotation (orientation): Mask coordinates of the glaucoma regions in the retinal image
  Localized RoI : Region placement
  CMskDenseNet-77- : Custom Mask-RCNN network with DenseNe-77 key points
SampleResolution ← [x y]
// Computing Mask
μ← AnchorsComputation (NS, annotation)
// Customized MaskRCNN model
CMskDenseNet-77← DesignCustomDenseNet-77MaskRCNN (SampleResolution, μ)
[ Sr, St] ← database division into train and test section
 // Glaucoma Region recognition from Training part
For each sample f in ⟶Sr
  Compute DenseNet-77 keypoints ⟶ns
End For
Training CMskDenseNet-77over ns, and compute training time t_dense
∂_dense ← PreRegionLoc(ns)
Ap_dense ← Evaluate_AP (DenseNet-77, ∂_dense)
For each sample F in ⟶ St
  (a) compute features by employing trained model ¥⟶βI
  (b) [Mask, objectness_score, classLabel] ←Predict (βI)
  (c) Output sample along with Mask, class
  (d) ∂← [∂ Mask]
End For
Ap_¥← Evaluate framework ¥ using ∂
FINISH.
```

ALGORITHM 1: Steps for OD and OC segmentation with custom Mask-RCNN.

TABLE 1: Hypermeters details.

Framework parameters	Value
Epochs	30
Learning rate	0.001
Batch size	8
Confidence score threshold	0.2
Unmatched threshold	0.5

2.6.1. Multitask Loss.

The presented framework uses a multitask loss L on all sampled ROIs given as follows:

$$L \text{ (MaskRCNN)} = L_{\text{bclass}} + L_{\text{ref}} + L_{\text{smask}}. \tag{1}$$

Here L_{bclass}, L_{ref}, and L_{smask} demonstrate the box class labels estimation loss, BB refinement loss, and segmentation mask prediction loss, respectively. L_{bclass} presents the log loss of the two categories (glaucoma/nonglaucoma), given as follows:

$$L_{\text{bclass}} (P_t, l) = -\log[P_t \, l + (1 - P_t)(1 - l)]. \tag{2}$$

L_{bclass} is the log loss of the binary classification, where P_t presents the target prediction probability of whether the anchor t holds glaucoma and l shows the gt label. There are about 20 k anchors generated of distinct scales and sizes that correspond with each other to cover the image. If an anchor has intersection over union (IoU) higher than 0.5 with a ground-truth (GT) box, it is classified as a positive anchor;

otherwise, it is negative. If several anchors overlap too much, we keep the one with the highest foreground score and discard the rest (referred to as nonmax suppression). Moreover, the value of l is 1 for true-marked anchors and 0 otherwise. The BB regression loss is given as follows:

$$L_{\text{lref}}\left(c_j, c_j^*\right) = \sum_{j \varepsilon \{x, y, w, h\}} \text{smooth}_{L1}\left(c_j - c_j^*\right), \tag{3}$$

where

$$\text{smooth}_{L1}(x) = \begin{cases} 0.5x^2, & \text{if } |x| < 1 \\ |x| - 0.5, & \text{otherwise.} \end{cases} \tag{4}$$

Here, vector c_j is presenting four dimensions of the estimated BB, and c_j^* is showing the dimensions of gt relating to the true-marked anchors. The smooth-$L1$ function is a robust $L1$ loss which is prone to outliers as compared to $L2$ loss. When regression targets are unbounded, training $L2$ loss leads to a gradient explosion and requires a carefully tuned learning rate. During the training of Mask-RCNN, the average cross-entropy loss is used which is calculated as follows:

$$L_{\text{mask}} = -\frac{1}{N^2} \sum_{1 \le x, y \le N} \left[p_{xy} \log V_{xy}^k + \left(1 - p_{xy}\right)\log\left(1 - V_{xy}^k\right) \right], \tag{5}$$

where p_{xy} is the pixel value at the location (x, y) in a *gt* mask of size $N \times N$ and for the same pixel, V_{xy}^k is presenting its estimated value in the mask obtained for class k ($k = 1$ for glaucoma region and 0 for nonglaucoma region) [28].

3. Results and Discussion

We have implemented the model using Keras and TensorFlow libraries with DenseNet-77 and FPN for feature extraction. We initialized the model using pretrained weights obtained from the COCO dataset and employed transfer learning to fine-tune the model on retinal datasets for OD and OC segmentation. For experimentation, we used a 70–30 ratio that is randomly divided into training (70%) and test (30%) sets.

3.1. Dataset. The evaluation experiments of the system were performed on the ORIGA "Online Retinal Fundus Image Database for Glaucoma Analysis" dataset [29]. The details of dataset are presented in Table 2. The dataset have a total of 650 images in which 168 are glaucomatous samples and the remaining 482 are nonglaucomatous samples and gathered from the "Eye Research Institute, Singapore." In each image, OD and OC regions are marked by experts using a vertical and nonrotated ellipse. The sample images are shown in Figure 4.

3.2. Evaluation Parameters. The proposed method is assessed by employing the intersection over union (IOU) as described in Figure 5. *A* shows the GT rectangle, and *B* denotes the estimated rectangle with ROI regions.

The first decision for the region is identified when the value of IOU is greater than 0.5; otherwise, it is not recognized. The average precision (AP) is mostly employed in evaluating the precision of object detectors, that is, R–CNN, SSD, and YOLO. The geometrical explanation of precision is shown in Figure 6. In our framework of the detection of glaucoma regions, AP depends on the idea of IOU [30].

3.3. Results. This section presented the details of results achieved after performing the experiments over diverse samples with light, color, region sizes variations, and the presence of blurring. For OD, to show the detection accuracy of the presented framework, the visual results are reported in Figure 7. It can be observed from the results that the proposed method can accurately localize the OD regions from the healthy areas despite discontinuous or blurry boundaries and artifacts in fundus images. Moreover, the Mask-RCNN method can precisely segment the OD regions by overcoming the challenges of location, shape, and size.

Furthermore, the visual results for OC segmented regions are shown in Figure 8. From the reported results, it can be visualized that our method can accurately localize and segment the OC regions under the different conditions due to a representative set of features extraction by DenseNet-77 and segmentation power of Mask-RCNN. However, its localization and segmentation power may slightly decrease for

samples with intense color variations which results in color-matching with healthy regions.

The proposed method can accurately recognize the OD and OC with an average accuracy of 0.965 on the ORIGA dataset. Moreover, the proposed technique can precisely segment the OD and OC by overcoming the challenges of blurriness and variations in location, size, and shape.

To further understand the performance of our method, we have used the evaluation parameters i.e., accuracy, precision, recall, *F*-measure, and IOU. Table 3 demonstrates the results or proposed approach. We can observe that the presented framework has achieved an average precision, recall, *F*-measure, and IOU as 0.965, 0.963, 0.97, and 0.972, respectively. Moreover, the confusion matrix of the proposed approach is presented in Figure 9.

3.4. DenseNet-77 Framework Evaluation. We performed an analysis to evaluate the robustness of the DenseNet-77 framework for eye disease detection by comparing it with other DL approaches. To accomplish this, the accuracy of the introduced Mask-RCNN with DenseNet-77 is compared with other base models, that is, Inception-v4 [31], VGG-16 [32], ResNet-101 [33], ResNet-152 [33], and DenseNet-121 [34].

Table 4 shows the comparative analysis of the presented method with other frameworks in both the aspect of model parameters and detection accuracy. The results of this comparative analysis indicate that the custom Mask-RCNN with DenseNet-77 works better than the Inception-v4, VGG-16, ResNet-50, ResNet-101, ResNet-152, and DenseNet-121. Moreover, from Table 4, it can be seen that VGG-16 has the highest model parameters, whereas ResNet-152 is the most expensive approach in terms of execution time. On the contrary, the presented framework with the DenseNet-77 model is economically most efficient and took only 1067 seconds for execution. The main reason for the efficient performance of DenseNet-77 is having a shallow architecture that employs efficient reuse of framework parameters without using redundant key point maps. Such structure of DenseNet-77 results in the extensively minimum number of framework parameters, whereas the comparative techniques suffer from high economical cost and unable to show efficient classification performance for the samples with noise, blurring, scale, and angle variations. Therefore, the presented technique better tackles the issues of comparative models by introducing a robust network for feature extraction and shows complicated transformations perfectly, leading to enhanced detection accuracy in postprocessing attacks as well. From the conducted analysis, it can be summarized that our customize Mask-RCNN with DenseNet-77 framework exhibits better performance than the other deep learning models in both terms of accuracy and efficacy.

3.5. Evaluation of the Custom Mask-RCNN Model. In this section, we have compared the performance of the introduced methodology with other region-based segmentation methods, that is, RCNN and Faster-RCNN over the ORIGA database, and results are reported in Figure 10. The RCNN is computationally complex as it randomly generates region

TABLE 2: Dataset details.

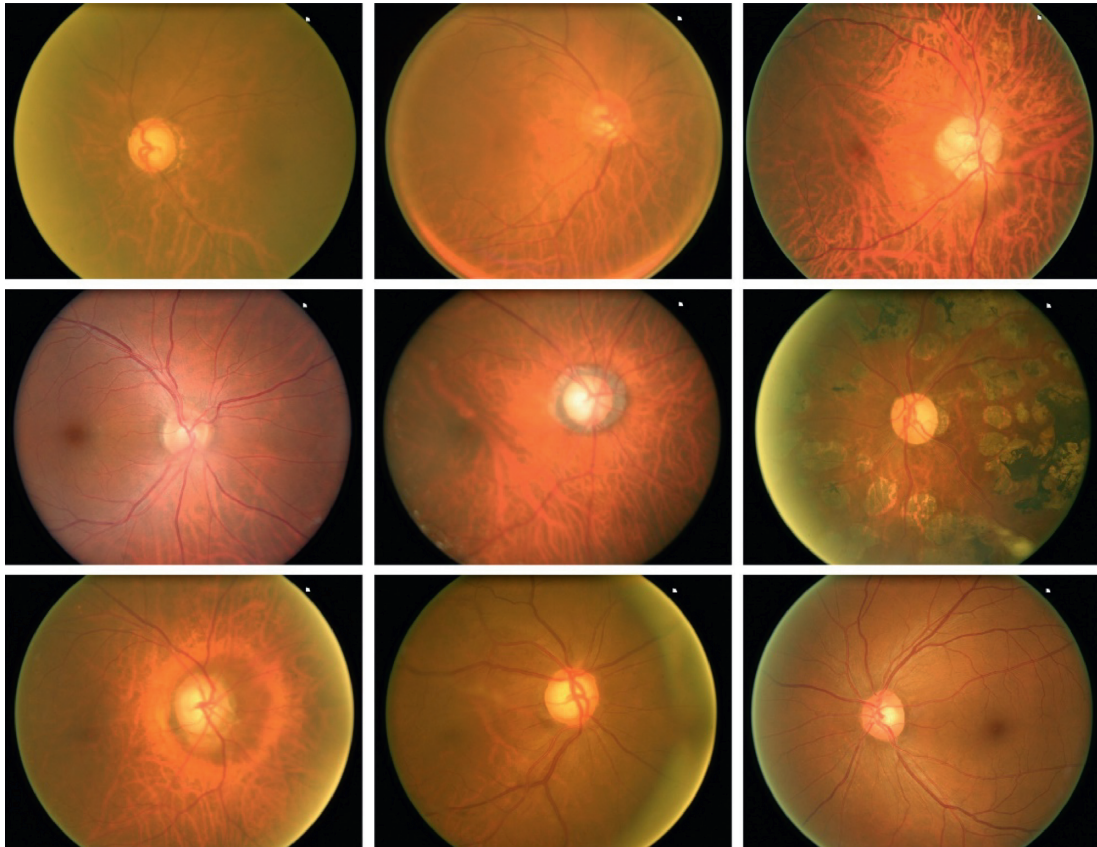| Attribute	Value
Total images	650
Glaucoma images	168
Normal images	482
Resolution	3072×2048
Ground truths	OD and OC regions

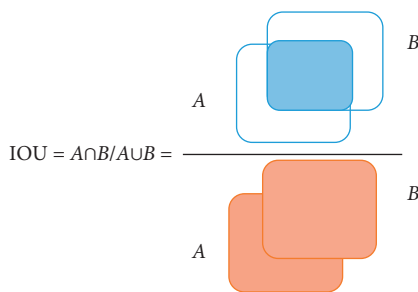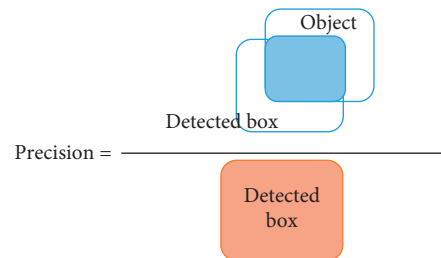FIGURE 4: Sample images of the ORIGA dataset.

FIGURE 5: IOU Venn diagram.

FIGURE 6: Geometrical representation of precision.

proposals (2000 per image) and uses a selective search algorithm for classification. The Faster-RCNN automatically extracts the region proposals using the RPN and shares the convolutional layer among class and BB network to reduce the computational cost. The traditional Mask-RCNN offers an added advantage over Faster-RCNN by providing an automated segmentation mask as well but is unable to capture the robust set of features under the postprocessing attacks. Therefore, the presented DenseNet-77 based Mask-RCNN performs well in comparison to traditional Mask-RCNN as DenseNet can capture the complex transformations with more accuracy which results in better automated segmentation and localization of

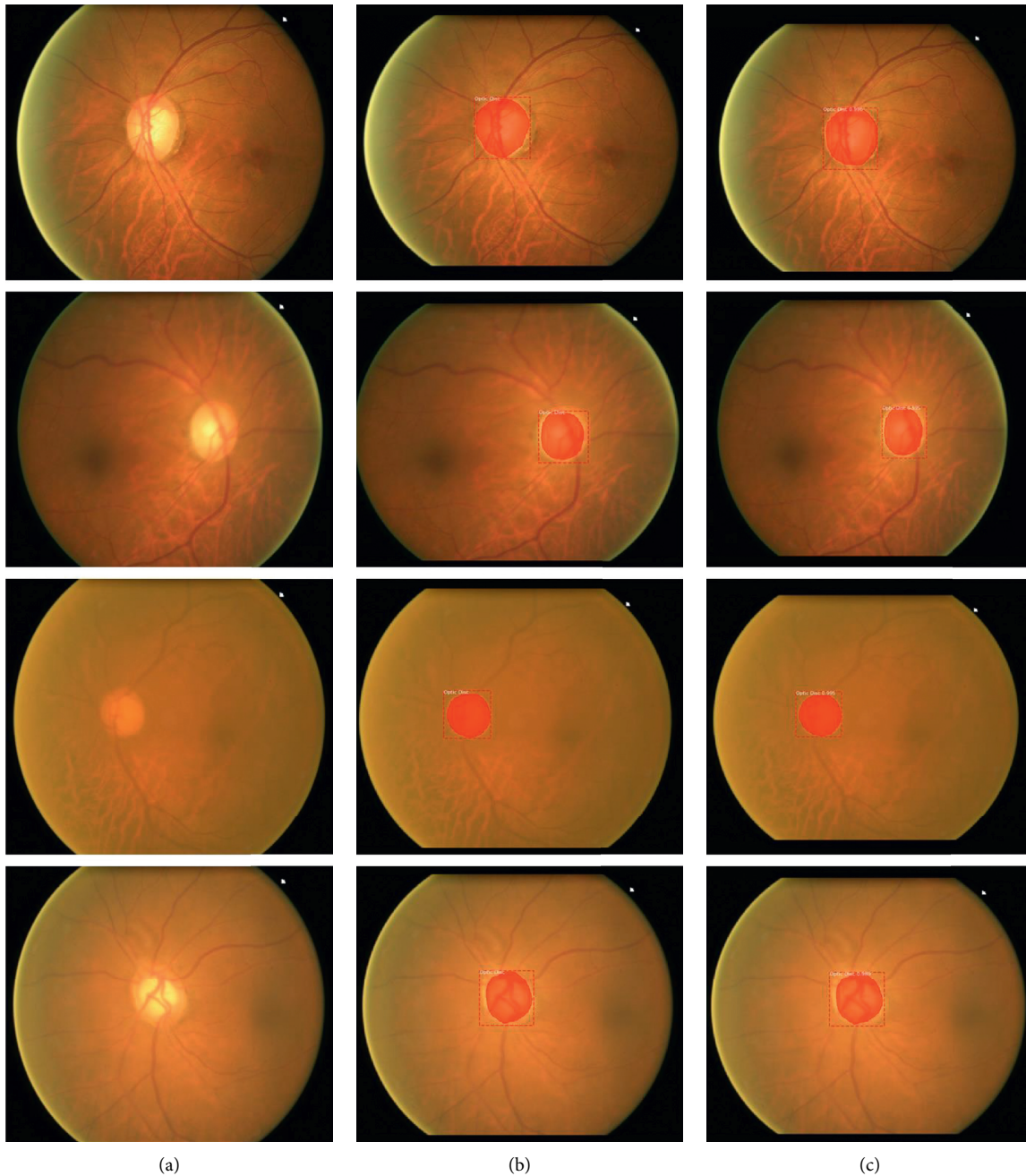|		
(a)	(b)	(c)

FIGURE 7: Visual results of OD segmentation. (a) Input images. (b) Annotated images. (c) Output images.

glaucoma regions. Moreover, our model is easier to train and adds a very small overhead over Mask-RCNN.

3.6. Comparative Analysis. Here, we have compared the performance of our model with the existing approaches over the ORIGA dataset. The proposed technique uses deep features that are more discriminating and reliable and provide a more effective representation of glaucoma regions over other methods. For performance evaluation, we evaluate our approach against the work of Bajwa et al. [1], Jiang et al. [21], Xu et al. [35], and Fu et al. [8]. These techniques are capable of detecting glaucoma from retinal images. However, they require intense training and exhibit lower accuracy for training samples with the class imbalance problem. The comparison results are presented in Table 5. Our framework has acquired the highest average precision, recall, and AUC, that is, 0.965, 0.963, and 0.96, respectively, that signifies the reliability of the proposed method in comparison with other methods. Unlike these methods, our model performs segmentation on the localized ROIs, which limits the space of segmentation and uses the ROIAlign layer which ultimately improves the accuracy of the final segmentation result.
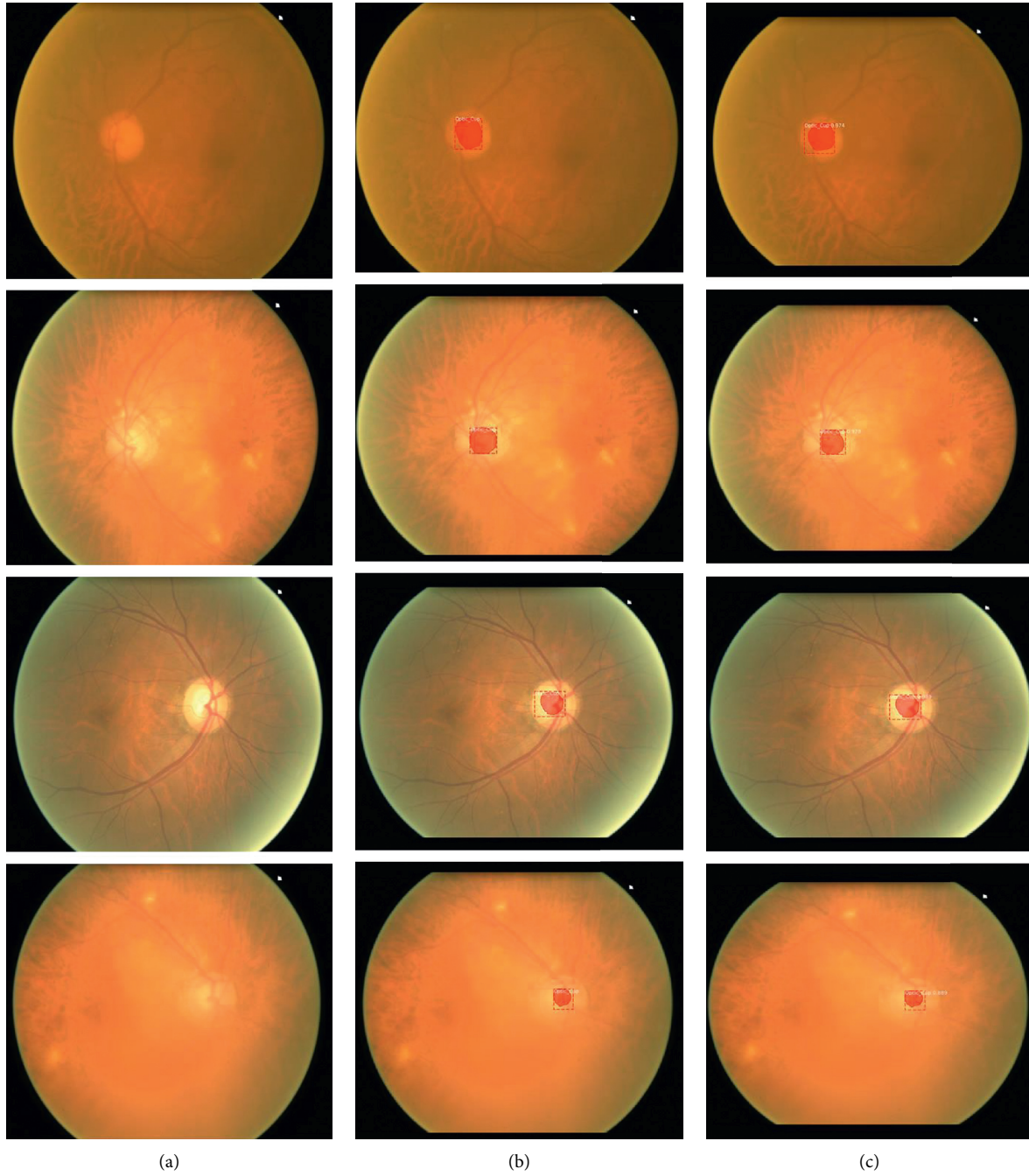
FIGURE 8: Visualization results of OC segmentation. (a) Input images. (b) Annotated images. (c) Output images.

TABLE 3: Proposed method results.

Method	Accuracy	Precision	Recall	F-measure	IOU
OD	0.979	0.959	0.969	0.953	0.981
OC	0.951	0.971	0.957	0.987	0.963
Average	0.965	0.965	0.963	0.970	0.972

3.7. Cross-Dataset Validation. To further evaluate the performance of the proposed method, we trained our method on the ORIGA dataset, and testing is performed on the HRF dataset [36]. The dataset contains 45 retinal images in which

15 images are healthy, 15 images are affected with diabetic retinopathy, and 15 images are affected by glaucoma.

We have plotted the box plot for evaluation of the cross dataset in Figure 11; the accuracy of the test and train is
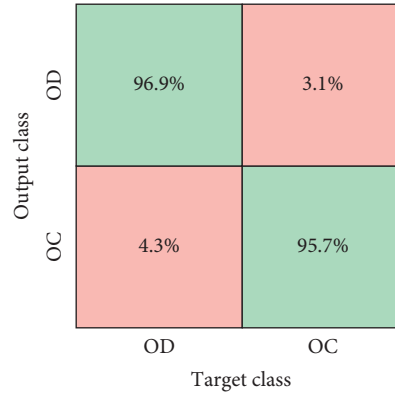
FIGURE 9: Confusion matrix.

TABLE 4: Comparative analysis of the proposed approach with base models.

| Parameters	Inception-V4	VGG-16	ResNet-101	ResNet-152	DenseNet-121	DenseNet-77
Total parameters (million)	41.2	119.6	42.5	58.5	7.1	6.2
Training loss	0.0102	0.5069	$4.1611e^{-04}$	$2.4844e^{-04}$	$5.6427e^{-04}$	$6.442e^{-04}$
Test loss	0.0686	0.6055	0.02082	0.0246	0.0159	0.0085
Training accuracy	99.74%	83.86%	99.99%	100%	100%	100%
Test accuracy	98.08%	81.83%	99.66%	99.59%	99.75%	99.983%
Processing time (s)	4042	1051	2766	4366	2165	1067
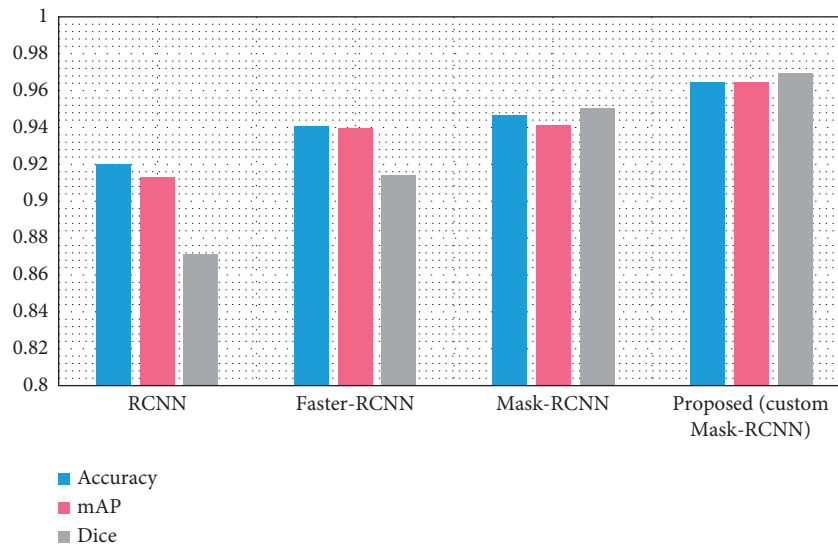
FIGURE 10: Comparison with other RCNN methods.

TABLE 5: Comparison with other techniques.

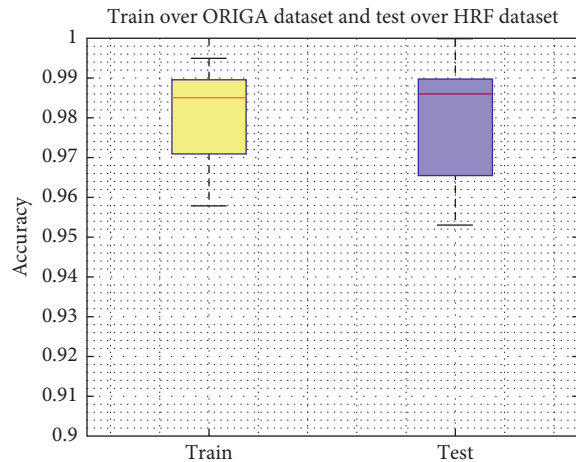| Method	Recall	Precision	AUC
Bajwa et al. [1]	0.71	—	0.860
Jiang et al. [21]	—	0.937	0.854
Xu et al. [35]	0.58	—	0.830
Fu et al. [8]	0.84	0.92	0.910
Proposed	0.963	0.965	0.970

Figure 11: Cross-dataset validation results.

spreading across the number line into quartiles, median, whisker, and outliers. According to the figure, we achieved an average accuracy of 98% for training and 97.7% for testing which exhibits that our proposed work outperforms the unknown samples as well. Therefore, it can be concluded that the introduced framework is robust to OD and OC localization and segmentation.

4. Conclusions

In this paper, we presented a deep learning technique to customize Mask-RCNN for precise and automated segmentation of OD and OC from the retinal images. We introduce the DenseNet-77 model at the feature computation level of Mask-RCNN to compute the more diverse key points which assist in accurately localizing the OD and OC regions under the various sample conditions. We have tested our framework over a challenging database, namely, ORIGA, and performed cross-dataset validation on the HRF database to show its robustness. The results exhibit that improved Mask-RCNN can compute deep features with effective representation of glaucoma regions over existing systems and serves as a new automated tool for diagnostic purposes. Moreover, both the qualitative and quantitative results show that Custom Mask-RCNN works better than the base framework. Although our approach has presented better OD and OC detection accuracy, however, it can be further enhanced by the inclusion of other latest DL-based techniques like EfficientNet. Furthermore, we plan to extend our work to other medical abnormalities.

Data Availability

Data sharing is not applicable to this article as authors have used publicly available datasets, whose details are included in the Experimental Results section of this article.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

References

[1] M. N. Bajwa, M. I. Malik, S. A. Siddiqui et al., "Two-stage framework for optic disc localization and glaucoma classification in retinal fundus images using deep learning," *BMC Medical Informatics and Decision Making*, vol. 19, no. 1, p. 136, 2019.

[2] N. B. Prakash and D. Selvathi, "An efficient detection system for screening glaucoma InRetinal images," *Biomedical and Pharmacology Journal*, vol. 10, no. 1, pp. 459–465, 2017.

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

[4] X. Chen, Y. Xu, D. W. K. Wong, T. Y. Wong, and J. Liu, "Glaucoma detection based on deep convolutional neural network," in *Proceedings of the 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 715–718, Milan, Italy, August 2015.

[5] J. Martins, J. S. Cardoso, and F. Soares, "Offline computer-aided diagnosis for Glaucoma detection using fundus images targeted at mobile devices," *Computer Methods and Programs in Biomedicine*, vol. 192, Article ID 105341, 2020.

[6] V. S. Mary, E. B. Rajsingh, and G. R. Naik, "Retinal Fundus Image Analysis for Diagnosis of Glaucoma: A Comprehensive Survey," *IEEE Access*, vol. 4, 2016.

[7] T. Nazir, A. Irtaza, A. Javed, H. Malik, D. Hussain, and R. A. Naqvi, "Retinal image analysis for diabetes-based eye disease detection using deep learning," *Applied Sciences*, vol. 10, no. 18, p. 6185, 2020.

[8] H. Fu, J. Cheng, Y. Xu et al., "Disc-aware ensemble network for glaucoma screening from fundus image," *IEEE Transactions on Medical Imaging*, vol. 37, no. 11, pp. 2493–2501, 2018.

[9] J. I. Orlando, E. Prokofyeva, M. del Fresno, and M. B. Blaschko, "Convolutional neural network transfer for automated glaucoma identification," in *Proceedings of the 12th International Symposium on Medical Information Processing and Analysis*, vol. 10160, p. 101600U, Tandil, Argentina, December 2017.

[10] B. Al-Bander, W. Al-Nuaimy, M. A. Al-Taee, and Y. Zheng, "Automated glaucoma diagnosis using deep learning approach," in *Proceedings of the 2017 14th International Multi-Conference on Systems, Signals & Devices (SSD)*, pp. 207–210, Marrakech, Morocco, March 2017.

[11] A. Li, J. Cheng, D. W. K. Wong, and J. Liu, "Integrating holistic and local deep features for glaucoma classification," in *Proceedings of the 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 1328–1331, Orlando, FL, USA, August 2016.

[12] Q. Abbas, "Glaucoma-deep: detection of glaucoma eye disease on retinal fundus images using deep learning," *International Journal of Advanced Computer Science and Applications*, vol. 8, no. 6, pp. 41–45, 2017.

[13] R. Hemelings, B Elen, J. Barbosa-Breda et al., "Accurate prediction of glaucoma from colour fundus images with a convolutional neural network that relies on active and transfer learning," *Acta Ophthalmologica*, vol. 98, no. 1, pp. e94–e100, 2020.

[14] M. D. Abràmoff, Y. Lou, A. Erginay et al., "Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning," *Investigative Opthalmology & Visual Science*, vol. 57, no. 13, pp. 5200–5206, 2016.

[15] Z. Lu and D. Chen, "Weakly supervised and semi-supervised semantic segmentation for optic disc of fundus image," *Symmetry*, vol. 12, no. 1, p. 145, 2020.

[16] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1904–1916, 2015.

[17] X. Chen, Y. Xu, S. Yan, D. W. K. Wong, T. Y. Wong, and J. Liu, "Automatic Feature Learning for Glaucoma Detection Based on Deep Learning," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 669–677, Munich, Germany, October 2015.

[18] U. Raghavendra, H. Fujita, S. V. Bhandary, A. Gudigar, J. H. Tan, and U. R. Acharya, "Deep convolution neural network for accurate diagnosis of glaucoma using digital fundus images," *Information Sciences*, vol. 441, pp. 41–49, 2018.

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

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

[21] Y. Jiang, L. Duan, Z. Gu et al., "JointRCNN: a region-based convolutional neural network for optic disc and cup segmentation," *IEEE Transactions on Biomedical Engineering*, vol. 67, Article ID 2913211, 2019.

[22] R. G. Ramani and J. J. Shanthamalar, "Improved image processing techniques for optic disc segmentation in retinal fundus images," *Biomedical Signal Processing and Control*, vol. 58, Article ID 101832, 2020.

[23] P. Elangovan and M. K. Nath, "Glaucoma assessment from color fundus images using convolutional neural network," *International Journal of Imaging Systems and Technology*, 2020.

[24] L. Li, "A large-scale database and a CNN model for attention-based glaucoma detection," *IEEE Transactions on Medical Imaging*, vol. 39, no. 2, pp. 413–424, 2019.

[25] Y. Wang, H. Li, P. Jia, G. Zhang, T. Wang, and X. Hao, "Multi-scale DenseNets-based aircraft detection from remote sensing images," *Sensors*, vol. 19, no. 23, p. 5270, 2019.

[26] J. Flusser, S. Farokhi, C. Höschl, T. Suk, B. Zitová, and M. Pedone, "Recognition of images degraded by Gaussian blur," *IEEE Transactions on Image Processing*, vol. 25, no. 2, pp. 790–806, 2015.

[27] A. Zisserman, "Vgg image annotator (VIA)," [Online]. Available: https://www.robots.ox.ac.uk/vgg/software/via. [Accessed: 03-May-2020], 2020.

[28] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2961–2969, Venice, Italy, October 2017.

[29] Z. Zhang, "Origa-light: an online retinal fundus image database for glaucoma analysis and research,," in *Proceedings of the 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, pp. 3065–3068, Buenos Aires, Argentina, September 2010.

[30] T. Nazir, A. Irtaza, J. Rashid, M. Nawaz, and T. Mehmood, "Diabetic retinopathy lesions detection using faster-RCNN from retinal images," in *Proceedings of the 2020 First International Conference of Smart Systems and Emerging Technologies (SMARTTECH)*, pp. 38–42, Riyadh, Saudi Arabia, March 2020.

[31] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 31, no. 1, San Francisco, CA, USA, February 2017.

[32] W. Yu, K. Yang, Y. Bai, T. Xiao, H. Yao, and Y. Rui, "Visualizing and comparing AlexNet and VGG using deconvolutional layers," in *Proceedings of the 33 rd International Conference on Machine Learning*, New York, NY, USA, June 2016.

[33] A. Canziani, A. Paszke, and E. Culurciello, "An analysis of deep neural network models for practical applications," 2016, https://arxiv.org/abs/1605.07678.

[34] I. Allaouzi and M. Ben Ahmed, "A novel approach for multi-label chest X-ray classification of common thorax diseases," *IEEE Access*, vol. 7, pp. 64279–64288, 2019.

[35] Y. Xu, S. Lin, D. W. K. Wong, J. Liu, and D. Xu, "Efficient reconstruction-based optic cup localization for glaucoma screening," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 445–452, Nagoya, Japan, September 2013.

[36] A. Budai, R. Bock, A. Maier, J. Hornegger, and G. Michelson, "Robust vessel segmentation in fundus images," *International Journal of Biomedical Imaging*, vol. 2013, Article ID 154860, 11 pages, 2013.