Utilizing Transfer Learning and a Customized Loss Function for Optic Disc Segmentation from Retinal Images

Abdullah Sarhan[1*], Ali Al-Khaz'Aly[1], Adam Gorner[2], Andrew Swift[2], Jon Rokne[1], Reda Alhajj[1,3], and Andrew Crichton[4]

[1] Department of Computer Science, University of Calgary, Calgary, AB, Canada
[2] Cumming School of Medicine, University of Calgary, Canada
[3]
[4] Department of Computer Engineering, Istanbul Medipol University, Istanbul, Turkey
[5] Department of Ophthalmology and Visual Sciences, University of Calgary, Canada

Abstract. Accurate segmentation of the optic disc from a retinal image is vital to extracting retinal features that may be highly correlated with retinal conditions such as glaucoma. In this paper, we propose a deep-learning based approach capable of segmenting the optic disc given a high-precision retinal fundus image. Our approach utilizes a UNET-based model with a VGG16 encoder trained on the ImageNet dataset. This study can be distinguished from other studies in the customization made for the VGG16 model, the diversity of the datasets adopted, the duration of disc segmentation, the loss function utilized, and the number of parameters required to train our model. Our approach was tested on seven publicly available datasets augmented by a dataset from a private clinic that was annotated by two Doctors of Optometry through a web portal built for this purpose. We achieved an accuracy of 99.78% and a Dice coefficient of 94.73% for a disc segmentation from a retinal image in 0.03 seconds. The results obtained from comprehensive experiments demonstrate the robustness of our approach to disc segmentation of retinal images obtained from different sources.

1 Introduction

Sight is one of the most important senses for humans, allowing us to visualize and explore our surroundings. Over the years, several degenerative ocular conditions affecting sight have been identified such as glaucoma and diabetic retinopathy. These conditions can threaten our precious sense of sight by causing irreversible visual-field loss [1]. Glaucoma is the world's second most prominent cause of irreversible vision loss after cataracts, accounting for 12% of annual cases of blindness worldwide [2]. According to one estimate, around 80 million people are currently affected by glaucoma, and around 112 million will be affected by 2024. Approximately 80% of patients do not know they have glaucoma until advanced vision loss occurs [1, 3].

The optic disc is one of the main anatomical structures in the eye which must be monitored and evaluated for progression when glaucoma is suspected [1]. Changes within the optic disc, such as the displacement of vessels or enlargement of the optic cup to optic disc ratio can be used to help determine if glaucoma is present and if there is progression of the disease [4]. These changes occur because of an irreversible decrease in the number of nerve fibres, glial cells and blood vessels.

Several methods have been proposed for disc segmentation. These can be categorized as follows: morphological approaches [5], template based matching approaches [4, 6], adaptive-thresholding based approaches [7], and pixel-classification based approaches [8]. Approaches related to the first three categories mainly fail in the presence of bright objects similar to the ones shown in Fig. 1 [9]. The red arrows in Fig. 1 indicate some bright regions that can be encountered in retinal images that may affect disc segmentation.

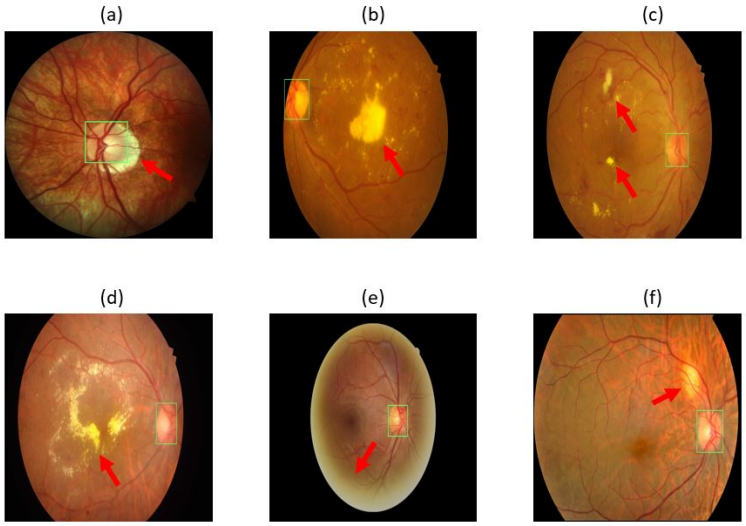

Fig. 1. Images showing various bright regions that can be observed in retinal images. The green box shows the location of the disc and the red arrows indicate other bright regions that may mislead some approaches when segmenting the disc.

With the rise of deep learning comes the potential for achieving high performance when segmenting a retinal image. Researchers have been working to develop models that place each pixel of a retinal image into a specific class during semantic segmentation. However, the performances of these approaches tend to decrease when new datasets emerge with different disc appearances or images with different resolutions. For instance, the deep-learning model M-Net, proposed by [10] performs well on the ORIGA dataset [11], but not on other datasets (e.g. the DRISHTI-GS dataset) [12] as reported by [13]. One cause

might be related to improper handling of the variance between classes when training. This is because the optic disc class may comprise between 2%-10% of an image depending on the angle and resolution of the captured image, whilst the background class would take up the rest of the image. This causes some models to converge toward the background and miss key details related to the disc.

In this paper, we propose a deep-learning approach to disc segmentation from retinal images using the UNET architecture to build the model and the VGG16 convolutional model as our encoder. Given the challenges related to having insufficient annotated disc datasets for deep learning, we adopted the idea of using transfer learning (TL) and image augmentation (IA). Instead of using random weights to initialize our model, we use weights trained on millions of images for semantic segmentation from the Imagenet dataset, which we then fine-tune to match the object we wanted to segment. To handle the issue of imbalanced classes, we use a customized loss function that allows the loss function to penalize more when the wrong classification is made for a pixel related to a disc than that of background. To prove the robustness of our proposed approach in segmenting the disc with various sizes, angles, and orientations, we tested the approach on seven publicly available datasets and one private dataset that we formed.

Our contributions can be listed as follows:

1. We proposed a UNET based deep learning model for disc segmentation that uses VGG16 as the encoder.
2. We demonstrated the effectiveness of using TL and IA for limited data.
3. We handled the issue of imbalanced image classes which may lead to inaccurate results by adopting a weighted loss function.
4. We contributed a new retinal image dataset for disc segmentation (ORDS).
5. We developed an online portal that can be used for annotating disc by multiple contributors.

2 Related Work

Two types of approaches have been developed for disc segmentation: those that locate the optic-disc center but do not segment the disc and those that both locate the disc region and then segment the disc. In this section, we cover approaches that aim to segment the disc rather than just locating it.

Earlier work entails the development of hand-crafted features that rely mainly on the shape of the disc and the intensity of pixels [4, 6, 14]. However, the performance of these hand-crafted approaches is easily affected by the presence of pathological regions and images with different resolutions (Fig. 1). Recently, advancements made in the field of deep learning have opened the door to using deep learning based models in the field of medical-image analysis. Such approaches exhibit superior performance over the hand-crafted ones [1, 15].

Several approaches have been developed for segmenting the optic disc with deep learning: e.g., using an edge-attention guidance network to perform proper

edge detection when segmenting the disc [16], using disc-region localization and then disc segmentation via a pyramidal multi-label network [17], using entropy-driven adversarial-learning models [18], using residual UNet based models [19], and using generative adversarial networks combined with VGG16 and transfer learning [20].

In [8], researchers used an ensemble-learning based convolutional neural-network model to segment the optic disc by first localizing the disc region. Entropy was used to select informative points; then the graph-cut algorithm was used to obtain the final segmentation of the disc. The researchers tested their approach only on the Drishti-GS [12], and RimOnev3[21], datasets. However, they used only 50 images from the Drishti-GS dataset with 40 for training and 10 for testing, even though the Drishti-GS dataset contains 101 images with 50 for training and 51 for testing.

The use of transfer learning when working with deep learning to analyze medical images has been adopted by various studies [22–24]. In the study performed by [20] researchers adopted transfer learning to train their encoder for segmenting the disc when given the whole image without the need for cropping. They used the PASC AL VOC 2012 pretrained weights [25]. They used only the Drishti-GS dataset in both training and testing their model. Moreover, the number of training parameters used by their model, $30.85 * 10^6$, is double that of our approach. In [13] transfer learning for disc segmentation was also used. They started by cropping the disc region using the UNET model developed in [26] initializing their encoder using the weights of the MobileNetV2 [27] trained on the ImageNet dataset [28].

The majority of the developed approaches tend to first locate the disc region and then feed this region into their model to avoid bright regions like the one shown in Fig. 1. For proper segmentation, such approaches are highly dependent on successful localization of this disc region. Moreover, these approaches tend to handle the issue of imbalanced classes, which makes their approach perform differently when new images with different resolutions emerge. In this paper, we show a model with an encoder that uses transfer learning, proper data augmentation, and a customized loss function can segment the optic disc with high precision, giving results comparable to the above-mentioned approaches. In this study, we do not localize the disc prior to performing the segmentation and instead feed the whole image to our model, rather than a specific region of the image.

3 Proposed Method

Our goal is to segment the optic disc given a retinal image. To achieve this, we propose a deep learning model with the same architecture as the UNET model [26]. A pixel matrix I is associated with each retinal image indicating the pixels either belong to the disc and the background. If I_{xy} represents a pixel at location (x, y) in the retinal image, this pixel will have a value of 1 if it belongs to the disc and 0 if it is a background pixel. The model will use these labeled images

and the actual retinal image to produce a new image with the same dimensions, where each pixel has a probability between 0 and 1 inclusive, thus indicating whether this pixel belongs to the disc or not. The closer the value is to 1, the higher the model's confidence is that it belongs to a disc. In this section, we describe the model adopted in this study [6].

3.1 Network Architecture

Instead of creating a new architecture, we adopted the U-Net architecture which consists of an encoder and a decoder. The encoder is responsible for down-sampling the image, and the decoder is responsible for up-sampling the image to provide the final output. In our case, we used the VGG16[29] model as the encoder and built the decoder by using a series of skip connections, convolutional, up-sampling, and activation layers, as shown in Fig. 2.

The original VGG16 model with a down-sampling factor of 32 is customized so that it could be used for semantic segmentation. It contains five down-sampling layers followed by two densely connected layers and a softmax layer for prediction. We removed the two densely connected layers at the end of the original model and replaced them with a single convolutional layer found in the center of our model, as shown in Fig. 2. Doing so reduced the number of parameters used to train the model from 134,327,060 to 16,882,452. Fixing this bottleneck significantly cut down the time and computational power required to train the model without causing any observable changes to the model's predictions.

We also removed the softmax layer and added all of the upsampling and convolutional layers seen on the right half of the model as is needed in image segmentation to regenerate the original image shape, finally our last layer is a sigmoid activation layer which predicts on the feature matrix. The 5 upsampling layers achieve an upsampling factor of 32, allowing the output images to have the same shape as input images, counteracting the data reshaping effects caused by down-sampling layers. The feature map for each convolutional layer is the ReLU [30] activation method, which applies Eq. 1 to each parameter coming out of the layer, thereby removing all negative pixel values.

$$f(x) = max(0, x) \tag{1}$$

Throughout the architecture, there are several instances where we use skip connections. In Fig. 2, the first and second maxpooling layers utilize a skip connection to convolutional layers further down in the pipeline of the model. Whilst the third and fourth maxpooling layers are connected directly to a convolutional layer which attaches either directly or indirectly to the later layers. This was used to shorten the distance between the earlier and later layers. Short connections from early to later layers are useful in preserving high-level information about the positioning of the disc. This is opposed to the low-level pixel-based information which is transferred across the long pipeline of the architecture in

[6] source code: https://github.com/AbdullahSarhan/ACCVDiscSegmentation

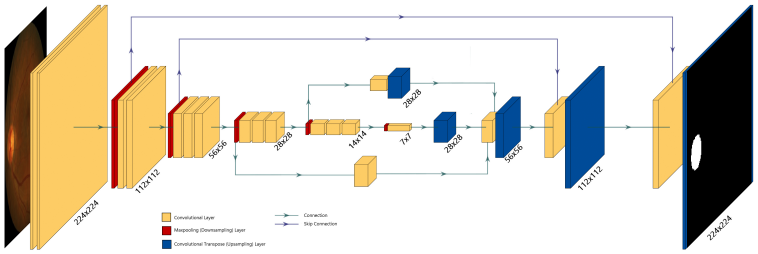

Fig. 2. Architecture of the Customized VGG-16 model Adopted in this Study.

a combination of convolutional and maxpooling/upsampling layers. High-level information tends to be lost as the image gets down-sampled and the shape and structure of the image is changed. Therefore, we maintain this information by using connections to earlier layers in the model. Better results were observed when using instead of not using these connections.

3.2 Transfer Learning

To handle the challenge faced in the field of medical imaging of not having enough datasets to train a deep-learning model, we used an approach referred to as transfer learning. As discussed in Section 2, such approaches can alleviate the issues caused by insufficient training data by using weights generated by training on millions of images [23]. In our study, we adopted the weights generated when training the VGG16 model on the ImageNet dataset [28] which contains around 14 million labeled images. We thus provided a diverse set of images that the model had been exposed to.

By using transfer learning, we could reduce the problem of over-fitting caused when training on limited images and improve the overall performance of the model. Using the ImageNet weights, we initialized the weights of the encoder network component, and other layers were randomly initialized using a Gaussian distribution. We then trained our model using a mini-batch gradient to tune the weights of the whole network. When training, we realized when using transfer learning that the model converged faster than without transfer learning.

In addition to transfer learning, we applied random augmentation to each image by randomly applying any of the following: horizontal shifting, vertical shifting, rotation within a range of 360 degrees, horizontal flipping, vertical flipping, or any combination of the above. We tested the evaluation effectiveness of data augmentation with and without transfer learning.

3.3 Loss Function

During the training of the network, we decided whether a model had improved on the value returned from the loss function by running it on validation data.

Initially, we adopted the binary cross-entropy function (BCE), as shown in Eq. 2 where N is the number of all pixels, y_i is the label of that pixel (0 for background and 1 for the disc), $p(y_i)$ is the predicted probability that the pixel belongs to the disc and $p(y_i)$ is the predicted probability of being a background pixel. Note that the BCE can penalize both false positives and false negatives when working with foreground and background classification.

$$BCE = -\frac{1}{N}\sum_{i=1}^{N} y_i.\log(p(y_i)) + (1-y_i).\log(1-(p(y_i))) \tag{2}$$

For any given retinal image, the disc will be only occupy a small region of the image (usually 2-10%), with the large majority of the image being background, i.e. 90% or more. Using this loss function alone would therefore not be sufficient for a precise disc segmentation output. This is because the BCE will be biased toward the background and hence, the disc will not be properly segmented. Thus, it may give an accuracy of 90%, which may be misleading. To bypass this issue, we decided also to use the Jaccard distance. The Jaccard distance measures how dissimilar two sets of data are. The Jaccard loss function is defined as:

$$L_j = 1 - \frac{|Y_d \cap \hat{Y_d}|}{|Y_d \cup \hat{Y_d}|} = 1 - \frac{\sum_{d\in Y_d}(1 \wedge \hat{y_d})}{|Y_d| + \sum_{b\in Y_b}(0 \vee \hat{y_b})} \tag{3}$$

where Y_d and Y_b represent the ground truth of the disc and background respectively. $\hat{Y_d}$ and $\hat{Y_b}$ represent the predicted disc and background pixels. $|Y_d|$ $|\hat{Y_b}|$ represents the cardinality of the disc Y_d and background $\hat{Y_b}$ respectively with $\hat{y_d} \in \hat{Y_d}$ and $\hat{y_b} \in \hat{Y_b}$. Since $\hat{Y_d}$ and $\hat{Y_b}$ are both probabilities, and their value will always be between 0 and 1, we can approximate this loss function as shown in Eq. 4 and the model will then be updated by Eq. 5 where j represents the the jth pixel of the input image and $\hat{y_j}$ represents the predicted value for that pixel.

$$\tilde{L_j} = 1 - \frac{\sum_{d\in Y_d} min(1, \hat{y_d})}{|Y_d| + \sum_{b\in Y_b} max(0, \hat{y_b})} = 1 - \frac{\sum_{d\in Y_d} \hat{y_d}}{|Y_d| + \sum_{b\in Y_b} \hat{y_b}} \tag{4}$$

$$L_j y_i \begin{cases} -\frac{1}{|Y_d| + \sum_{b\in Y_b} \hat{y_b}} & for \quad i \in Y_d \\ -\frac{\sum_{d\in Y_d} \hat{y_d}}{|Y_d| + \sum_{b\in Y_b} \hat{y_b}} & for \quad i \in Y_b \end{cases} \tag{5}$$

Given the Jaccard loss function, we are able to balance the emphasis the model gives to each of the classes: namely, the disc class and background class. In this, we combine BCE with Jaccard to optimize the results. We realize that, when both are combined, the model can converge faster than it can when using only the Jaccard while still achieving better results than BCE or Jaccard alone. Hence, our final loss function is:

$$Loss = BCE + L_j \tag{6}$$

3.4 Implementation Details

To implement this model we used a windows machine with a NVIDIA GeForce 2060 RTX with 6 GB dedicated GDDR6 memory and 8GB of shared random access memory which the GPU is free to use as necessary. We used the Python language to implement the proposed approach using Keras with TensorFlow back-end.

Training was performed using the NAdam optimizer [31] function with learning rate set to 0.0001, $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 10^{-8}$, and batch size of 4 images. During training, three callbacks were used. First, the model checkpoints would save the model whenever a smaller value was returned on validation data from the custom loss function when comparing to the value at the last checkpoint. Secondly, the learning rate was reduced by a factor of 0.5 whenever 25 epochs passed without any improvement in the validation loss values. Finally, the training was stopped if 100 epochs passed without any improvement.

4 ORDS Dataset

Datasets obtained from different resources had to be used in order to evaluate the reliability and applicability of the proposed method. One of the issues faced when working with disc segmentation is the lack of diverse datasets. To augment the available data sets we decided to contribute a new dataset obtained from a private clinic, annotated by two experts in this field. In this section, we discuss the new data collected.

The ORDS dataset, our new dataset [7], was obtained from a private clinic in Calgary, and the disc was annotated by two Doctors of Optometry. We built a customized web portal to help optometrists trace the disc[8]. Each optometrist was assigned a username and password to log into the portal and view the assigned images. Upon successful login, a user can navigate to the tracing page and start tracing, as shown in Fig.3. Both optometrists traced the same set of images; and hence each image received two annotations for the disc. In total, 135 images were annotated. On the tracing page, a list of images is presented; the user can click on any image, and a pop-up dialogue will appear. Once the pop-up model appears, the user can start tracing; an erase option is presented should the user wish to erase any of the tracing. Once the tracing is done, the user can click on the submit button, which will allow the storage of tracing information on a dedicated server. Users have the option either to trace the whole disc at once or in steps. Upon successful submission of the tracing, the traced image will be eliminated from the list of images on the tracing page.

5 Experimental Results

We evaluated our methods on eight different datasets which allows us to evaluate our approach when discs with different sizes, orientations, and resolutions are

[7] https://github.com/AbdullahSarhan/ACCVDiscSegmentation
[8] Link and login credentials can be provided upon request

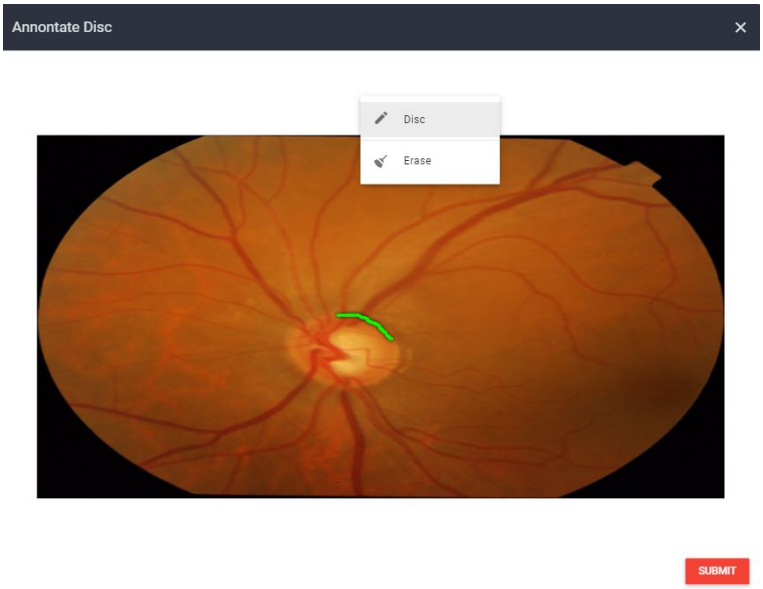

Fig. 3. Web portal showing images assigned to the Optometrists along with the tracing form utilized for disc tracing.

fed to our model. In this section, we discuss the datasets adopted, experiments conducted, and compare the performance of our model with other approaches.

5.1 Datasets

To verify the robustness of our method, we tested our approach on seven publicly available datasets and our dataset. Table 1 provides an overview of these datasets along with the machines used to capture these images, including our new dataset. These datasets contain information regarding multiple retinal conditions: namely, glaucoma and diabetic retinopathy. Moreover, retinal images that belong to these datasets were acquired at different angles and resolutions, as can be seen in Fig. 4. For datasets that contained multiple annotations, including our new dataset which had two expert tracings, we used the average of the tracings when training and evaluating our model, which is the common technique used in such scenarios [21, 12]. In total 1,442 images were used for training and 705 were used for testing. To test our model, we had to have the data split into training and testing portions. The model could only see the training images, and we checked the performance by evaluating the model's predictions on the test images and comparing it to labels. Doing so makes it fair to compare our model with other approaches, as they would test their approach on the same test images we are using. However, not all datasets are split in this manner. For some datasets, we had to do the splitting with 75% of the dataset used for training

Table 1. Dataset properties and machines used to capture their images.

Dataset	Images		Dimensions	Machine
	Train	Test		
Drishti-GS [12]	50	51	2049*1757	-
Refuge [32]	400	400	2124*2056	Zeiss Visucam 500
IDRID [33]	54	27	4288*2848	Kowa VX-10 alpha
Rim_r3 [21]	128	31	1072*712	Nidek AFC-210
BinRushed [34]	147	35	2376*1584	Canon CR2 non-mydriatic
Magrebia [34]	52	11	2743*1936	Topcon TRC 50DX mydriatic
Messidor [34]	365	92	2240*1488	Topcon TRC NW6 non-mydriatic
ORDS	110	25	1444*1444	Zeiss, Visucam 200

and 25% for testing, selection done randomly. We did this split for the Messidor, ORDS, BinRushed, Magrebia, and RimOneV3 datasets. Note that the annotation used for Messidor is different from that used by other approaches,(e.g. [35]) as the annotation used by such studies is not available anymore. We used the one provided by [34]. The testing images for all datasets are provided with our code so that other researchers can make fair comparisons to ours, and thereby standardize the images these comparisons are made on. Fig. 4 shows the performance of our model on a test image from each dataset, each dataset being different in terms of angle and resolution.

Doing this allows other researchers to compare their approaches by standardizing the set of test images without using the leave-one-out strategy ,[36], which would be time-consuming due to the number of training experiments that must be conducted for each dataset. For instance, if we have a dataset with 200 images then we need to train our model 200 times each time using 199 images and test on the excluded image; we would have to train 200 models and average the test results of them. Images found across different datasets and even within a single

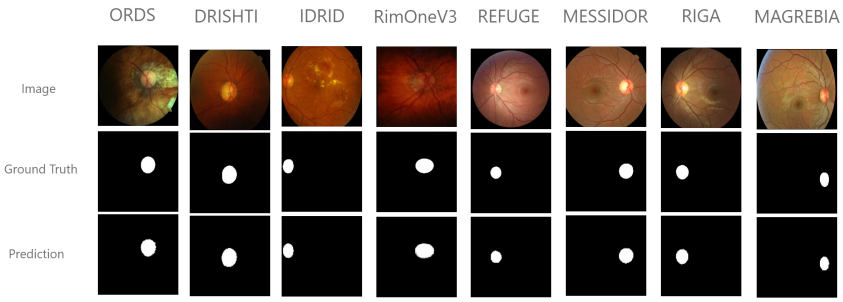

Fig. 4. A sample image from each dataset used in our study. The first row shows the actual retinal image while the second and third shows the related ground truth and prediction made by our model respectively. The name of the dataset to which each image belongs is written at the top of its column.

dataset can be extremely inconsistent in shape, size of optic disc region, and pixel values. Therefore, a general rule is applied to preprocess all images before they are passed to the model. They are first resized to 224*224 pixels, normalized so that all pixel values are within the range (0,1), and finally, undergo binary thresholding for disc ground-truth images.

5.2 Evaluation Methods

We used four evaluation methods to evaluate and compare our approach: namely, accuracy (Acc):$\frac{TP+TN}{TP+FP+TN+FN}$, dice coefficient (DC):$2 * \frac{Area(A \cap B)}{Area(A)+Area(B)}$, sensitivity (Sen): $\frac{TP}{TP+FN}$, and intersection over union (IoU):$\frac{Area(A \cap B)}{Area(A) \cup Area(B)}$. Moreover, we also show the time required by our approach to segment the disc and compare it with information obtained by other approaches (when applicable).

5.3 Effectiveness of TL and IA

To test the impact of using transfer learning (TL) and image augmentation (IA) when training our model we conducted a series of experiments and then evaluated the model obtained using the test images for all datasets. In this section we show the overall performance without showing performance on each dataset. Note that in all these experiments we used the loss function defined in Eq. 6.

We first checked the performance of the model without transfer learning by randomly initializing weights using Gaussian distribution, which needed 128 epochs to finish training. Then we did an experiment using data augmentation also without transfer learning and this needed 141 epochs. The third and fourth experiment using TL but with and without IA and they needed 184 and 207 epochs to finish training respectively. The evaluation results for each of these experiments are shown in Table 2. The results obtained show that using TL and IA together achieve the best results especially for DC and IoU values, which really reflect how precisely the disc is segmented along with it being slightly faster than the other models.

Table 2. Performance comparison of proposed method with and without using transfer learning (TL) and/or image augmentation (IA).

Experiment	Acc	DC	Sen	IoU	Time(s)
No TL and No IA	99.68	92.41	97.01	86.41	0.0317
IA with no TL	99.74	93.80	96.18	88.59	0.0366
TL with No IA	99.72	93.41	97.13	87.94	0.0308
TL with IA	99.78	94.73	96.26	90.13	0.0306

5.4 Effectiveness of Loss Functions

A well known loss function for binary classification is the binary cross entropy loss function. This loss function works great when the classes in the image are balanced. However in our case, the object we are trying to segment represents 10% or less of the total image area of the image. Hence, we decided to use the Jaccard distance approach as noted earlier.

We conducted three experiments to test which configuration would achieve the best results. First, we trained our model using the BCE loss function alone, which is a built-in loss function in the keras library, second, we trained using only the Jaccard loss function and finally, we trained using a combination of both loss functions. The results obtained are shown in Table 3. We realized there is slight improvement in performance when we combine both loss functions compared to using either one of them alone. We also realized that using Jaccard alone achieved better results than BCE but it took 516 epochs to finish training compared to 210 epochs when using BCE alone and 374 epochs when combining both. Note that in all these experiments we used TL and IA.

Table 3. Performance of the model across different loss functions.

Loss Function	Acc	DC	Sen	IoU	Time(s)
BCE	99.75	94.01	94.89	88.82	0.0358
Jaccard Distance	99.76	94.03	95.72	88.85	0.0329
Jaccard Distance+BCE	99.78	94.73	96.26	90.13	0.0306

5.5 Comparing with Other Approaches

To evaluate our proposed method we compared with approaches which were tested on some of the same datasets we used, as shown in Table 4. Unfortunately, these approaches did not evaluate using all available datasets and hence when comparing we split our results per dataset to be able to do a fair comparison. We achieved an overall average accuracy of 99.78%, DC of 94.73%, Sensitivity of 96.26% and IoU of 90.13%. Our approach outperformed other approaches tested on some of the online publicly available datasets as shown in Table 4 except two approaches for some of the dataset they used. Further, we achieved a prediction time that is the best among the current state of the art approaches with average segmentation time is 0.03s.

For the Refuge dataset we achieved better results than the ones reported by [17] and [16] yet we achieved slightly lower than the values reported by [13] whom reported achieving 96% where we achieved 94.09%. However, we achieved better than them in the RimOneV3 dataset and Drishti-GS. Note that they first localize a region of interest and then segment the disc whereas in our case we directly segment the disc from the whole retinal image without first localizing the region the disc is located in.

Table 4. Performance comparison of proposed method on optic disc segmentation.

Method	Dataset	performance metrics				Time(s)
		Acc	DC	Sen	IoU	
2***Wang et al. [18]**	**RimOnev3**	-	89.80	-	-	-
	Drishti-GS	-	96.40	-	-	-
PM-Net [17]	**Refuge**	97.90	-	-	-	-
2***ET-Net [16]**	**Refuge**	-	92.29	-	86.70	-
	Drishti-GS	-	93.14	-	87.90	-
2***Thakur et al. [35]**	**RimOneV3**	94.84	93.00	-	-	38.66
	Drishti-GS	93.23	92.00	-	-	-
GAN-VGG16 [20]	**Drishti-GS**	-	**97.10**	-	-	1
ResUNet [19]	**IDRID**	-	86.50	-	-	-
pOSAL [13]	**Refuge**	-	**96.00**	-	-	-
	Drishti-GS	-	96.50	-	-	-
	RimOneV3	-	86.50	-	-	-
9*Proposed Approach	**Drishti-GS**	**99.79**	96.50	**97.54**	**93.18**	**0.03**
	IDRID	**99.80**	**95.39**	**96.94**	**91.30**	**0.12**
	RimOneV3	**99.50**	**94.91**	**96.11**	**90.44**	**0.03**
	Refuge	**99.80**	94.09	**95.77**	**89.00**	**0.02**
	BinRushed	**99.82**	**95.57**	**96.97**	**91.53**	**0.03**
	Magrebia	**99.80**	**96.18**	**95.58**	**92.68**	**0.04**
	Messidor	**99.83**	**96.16**	**97.18**	**92.62**	**0.03**
	ORDS	**99.50**	**93.58**	**96.83**	**88.25**	**0.03**

For the Drishti-GS dataset our model performed better than [18, 16, 35], the same as [13], and slightly lower than [20]. However, in [20] they only trained and tested their approach one the Refuge dataset, which is not enough to show how well their system work on images from multiple sources. Moreover, their model requires $30.85 * 10^6$ parameters which is almost double what our model requires.

Our model achieved better results than the approaches mentioned above for the IDRID and RimOneV3 datasets. For the dataset provided by [34] they are still a new dataset and up to our knowledge there is no study with published testing images that we can use to compare the performance of our model with. To ensure continuity of this research and allow researchers to be able to perform fair comparison we will publish all test images used to evaluate our model in our supporting material. We also publish both the training history log and our model which was tested on in Table 4. In general our model demonstrated high performance segmenting the disc for images obtained from different resources with different angles of the disc and resolutions, including challenging ones as shown in Fig. 1 (check supplementary material for more images).

5.6 Leave One Out Experiment

Clinics may capture images with different resolutions and angles. To verify the robustness of our model on images that it was not trained on, that may have different characteristics than what it was trained on, we conducted 8 experiments

where a model was trained on all datasets except for one which was used for evaluation. The results obtained for each dataset are showing in Table 5. This table shows that for instance, when the model is trained on all datasets except for Refuge, it will achieve a DC value of 92.51% which is slightly less than when using the cross training which is 94.09%. The results seem consistent in that our model can effectively segment the disc, except for the RimOneV3 dataset. This is likely because this dataset only provides the images of the area surrounding the disc.

Table 5. Performance of the model when being trained on all datasets except for the one being evaluated on.

Dataset	Acc	DC	Sen	IoU	Time(s)
Drishti-GS	99.76	95.94	96.82	92.25	0.07
IDRID	99.66	92.47	97.83	86.11	0.03
RimOneV3	98.23	80.00	89.49	70.00	0.03
Refuge	99.75	92.51	96.34	86.42	0.02
BinRushed	99.74	95.01	92.67	90.54	0.03
Magrebia	99.76	95.68	93.77	91.74	0.04
Messidor	99.66	94.11	97.83	90.00	0.03
ORDS	99.29	89.29	86.00	81.51	0.03

6 Conclusion and Future Work

In this paper, we proposed a deep learning based approach for disc segmentation where we proved the effectiveness of transfer learning, image augmentation, and a customized loss function. Our approach achieved state of the art performance on disc segmentation when compared to other modern approaches. We also contribute a new dataset the can be used by researchers for improving disc segmentation. This will help researchers testing their approaches on images obtained from various sources with diverse data. Our new dataset was annotated by two doctors of optometry using an online portal we built for the annotation task.

As for future work, we would like to expand our approach to include glaucoma detection by analyzing the disc region. Using the cup/disc alone is not always an indicator for glaucoma and hence we need to analyze the disc region and make an assessment. Moreover, we would like also to expand our portal to be used for educational and research purposes where people can share and annotate the datasets. Further, we would like to improve our dataset to include annotation for other anatomical objects in the retina such as peripapillary atrophy and exudates.

References

1. Sarhan, A., Rokne, J., Alhajj, R.: Glaucoma detection using image processing techniques: A literature review. Computerized Medical Imaging and Graphics **78** (2019) 101657
2. Fu, H., Xu, Y., Lin, S., Zhang, X., Wong, D.W.K., Liu, J., Frangi, A.F., Baskaran, M., Aung, T.: Segmentation and quantification for angle-closure glaucoma assessment in anterior segment oct. IEEE Transactions on Medical Imaging (2017) 1930 – 1938
3. Tham, Y.C., Li, X., Wong, T.Y., Quigley, H.A., Aung, T., Cheng, C.Y.: Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. Ophthalmology **121** (2014) 2081–2090
4. Issac, A., Sarathi, M.P., Dutta, M.K.: An adaptive threshold based image processing technique for improved glaucoma detection and classification. Computer Methods and Programs in Biomedicine **122** (2015) 229–244
5. Panda, R., Puhan, N., Panda, G.: Robust and accurate optic disk localization using vessel symmetry line measure in fundus images. Biocybernetics and Biomedical Engineering **37** (2017) 466–476
6. Sun, J., Luan, F., Wu, H.: Optic disc segmentation by balloon snake with texture from color fundus image. International Journal of Biomedical Imaging **2015** (2015)
7. De La Fuente-Arriaga, J.A., Felipe-Riverón, E.M., Garduño-Calderón, E.: Application of vascular bundle displacement in the optic disc for glaucoma detection using fundus images. Computers in Biology and Medicine **47** (2014) 27–35
8. Zilly, J., Buhmann, J.M., Mahapatra, D.: Glaucoma detection using entropy sampling and ensemble learning for automatic optic cup and disc segmentation. Computerized Medical Imaging and Graphics **55** (2017) 28–41
9. Sarhan, A., Rokne, J., Alhajj, R.: Approaches for early detection of glaucoma using retinal images: A performance analysis. In: Data Management and Analysis. Springer (2020) 213–238
10. Fu, H., Cheng, J., Xu, Y., Wong, D.W.K., Liu, J., Cao, X.: Joint optic disc and cup segmentation based on multi-label deep network and polar transformation. IEEE Transactions on Medical Imaging **37** (2018) 1597–1605
11. Zhang, Z., Yin, F.S., Liu, J., Wong, W.K., Tan, N.M., Lee, B.H., Cheng, J., Wong, T.Y.: Origa-light: An online retinal fundus image database for glaucoma analysis and research. In: 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology, IEEE (2010) 3065–3068
12. Sivaswamy, J., Krishnadas, S., Joshi, G.D., Jain, M., Tabish, A.U.S.: Drishti-gs: Retinal image dataset for optic nerve head (onh) segmentation. In: International Symposium on Biomedical Imaging (ISBI), IEEE (2014) 53–56
13. Wang, S., Yu, L., Yang, X., Fu, C.W., Heng, P.A.: Patch-based output space adversarial learning for joint optic disc and cup segmentation. IEEE Transactions on Medical Imaging **38** (2019) 2485–2495
14. Mohamed, N.A., Zulkifley, M.A., Zaki, W.M.D.W., Hussain, A.: An automated glaucoma screening system using cup-to-disc ratio via simple linear iterative clustering superpixel approach. Biomedical Signal Processing and Control **53** (2019) 101454
15. Shen, D., Wu, G., Suk, H.I.: Deep learning in medical image analysis. Annual Review of Biomedical Engineering **19** (2017) 221–248
16. Zhang, Z., Fu, H., Dai, H., Shen, J., Pang, Y., Shao, L.: Et-net: A generic edge-attention guidance network for medical image segmentation. In: International

Conference on Medical image computing and computer-assisted intervention (MICCAI), Springer (2019) 442–450
17. Yin, P., Wu, Q., Xu, Y., Min, H., Yang, M., Zhang, Y., Tan, M.: Pm-net: Pyramid multi-label network for joint optic disc and cup segmentation. In: International Conference on Medical image computing and computer-assisted intervention (MICCAI), Springer (2019) 129–137
18. Wang, S., Yu, L., Li, K., Yang, X., Fu, C.W., Heng, P.A.: Boundary and entropy-driven adversarial learning for fundus image segmentation. In: International Conference on Medical image computing and computer-assisted intervention (MICCAI), Springer (2019) 102–110
19. Baid, U., Baheti, B., Dutande, P., Talbar, S.: Detection of pathological myopia and optic disc segmentation with deep convolutional neural networks. In: TENCON 2019-2019 IEEE Region 10 Conference (TENCON), IEEE (2019) 1345–1350
20. Jiang, Y., Tan, N., Peng, T.: Optic disc and cup segmentation based on deep convolutional generative adversarial networks. IEEE Access **7** (2019) 64483–64493
21. Pena-Betancor, C., Gonzalez-Hernandez, M., Fumero-Batista, F., Sigut, J., Medina-Mesa, E., Alayon, S., de la Rosa, M.G.: Estimation of the relative amount of hemoglobin in the cup and neuroretinal rim using stereoscopic color fundus images. Investigative Ophthalmology & Visual Science **56** (2015) 1562–1568
22. Shin, H.C., Roth, H.R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., Summers, R.M.: Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning. IEEE Transactions on Medical Imaging **35** (2016) 1285–1298
23. Pan, S.J., Yang, Q.: A survey on transfer learning. IEEE Transactions on knowledge and Data Engineering **22** (2009) 1345–1359
24. Karri, S.P.K., Chakraborty, D., Chatterjee, J.: Transfer learning based classification of optical coherence tomography images with diabetic macular edema and dry age-related macular degeneration. Biomedical Optics Express **8** (2017) 579–592
25. Everingham, M., Van Gool, L., Williams, C.K., Winn, J., Zisserman, A.: The pascal visual object classes (voc) challenge. International Journal of Computer Vision **88** (2010) 303–338
26. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention (MICCAI), Springer (2015) 234–241
27. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.C.: Mobilenetv2: Inverted residuals and linear bottlenecks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2018) 4510–4520
28. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al.: Imagenet large scale visual recognition challenge. International Journal of Computer Vision **115** (2015) 211–252
29. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
30. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
31. Dozat, T.: Incorporating nesterov momentum into adam. (2016)
32. Orlando, J.I., Fu, H., Breda, J.B., van Keer, K., Bathula, D.R., Diaz-Pinto, A., Fang, R., Heng, P.A., Kim, J., Lee, J., et al.: Refuge challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs. Medical Image Analysis **59** (2020) 101570

33. Porwal, P., Pachade, S., Kokare, M., Deshmukh, G., Son, J., Bae, W., Liu, L., Wang, J., Liu, X., Gao, L., et al.: Idrid: Diabetic retinopathy–segmentation and grading challenge. Medical Image Analysis **59** (2020) 101561
34. Almazroa, A., Alodhayb, S., Osman, E., Ramadan, E., Hummadi, M., Dlaim, M., Alkatee, M., Raahemifar, K., Lakshminarayanan, V.: Retinal fundus images for glaucoma analysis: the riga dataset. In: Medical Imaging 2018: Imaging Informatics for Healthcare, Research, and Applications. Volume 10579., International Society for Optics and Photonics (2018) 105790B
35. Thakur, N., Juneja, M.: Optic disc and optic cup segmentation from retinal images using hybrid approach. Expert Systems with Applications **127** (2019) 308–322
36. Wang, X., Jiang, X., Ren, J.: Blood vessel segmentation from fundus image by a cascade classification framework. Pattern Recognition **88** (2019) 331–341