



# Optic Disc, Cup and Fovea Detection from Retinal Images Using U-Net++ with EfficientNet Encoder

Ravi Kamble, Pranab Samanta, and Nitin Singhal<sup>(✉)</sup>

AIRAMATRIX Pvt. Ltd., Mumbai, India  
nitin.singhal@airamatrix.com

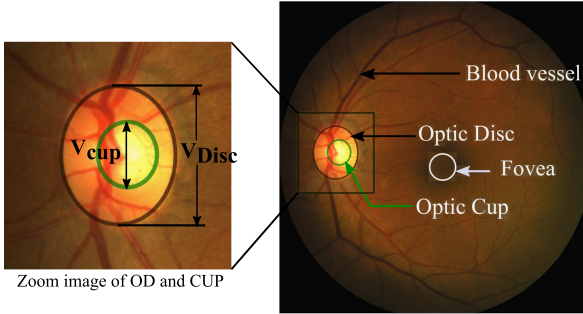
**Abstract.** The accurate detection of retinal structures like an optic disc (OD), cup, and fovea is crucial for the analysis of Age-related Macular Degeneration (AMD), Glaucoma, and other retinal conditions. Most segmentation methods rely on separate detection of these retinal structures due to which a combined analysis for computer-aided ophthalmic diagnosis and screening is challenging. To address this issue, the paper introduces an approach incorporating OD, cup, and fovea analysis together. The paper presents a novel method for the detection of OD with a cup and fovea using modified U-Net++ architecture with the EfficientNet-B4 model as a backbone. The extracted features from the EfficientNet are utilized using skip connections in U-Net++ for precise segmentation. Datasets from ADAM and REFUGE challenges are used for evaluating the performance. The proposed method achieved a success rate of 94.74% and 95.73% dice value for OD segmentation on ADAM and REFUGE data, respectively. For fovea detection, the average Euclidean distance of 26.17 pixels is achieved for the ADAM dataset. The proposed method stood first for OD detection and segmentation tasks in ISBI ADAM 2020 challenge.

**Keywords:** Optic disc segmentation · Fovea localization · Convolutional neural network · Age-related macular degeneration · Glaucoma

## 1 Introduction

Early detection and screening of retinal diseases such as glaucoma and age-related macular degeneration (AMD) play a vital role in reducing vision loss [20]. Glaucoma is a disease that damages the eye's optic nerve which can lead to permanent vision impairment [7]. AMD is the leading cause of blindness due to the presence of macular drusen in people older than 65 years. The occurrence of lesions in the macula of the eye causes loss of central vision. Hence, the optic disc (OD), optic cup (OC), and the fovea are the most important retinal landmarks in ophthalmic imaging and diagnosis [18]. OD is the yellowish vertical oval region where the nerve fibers and blood vessels merge in the retina. The optic cup

is the brightest area in the optic disc region shown in Fig. 1. The cup to disc ratio (CDR) is one of the important markers for the diagnosis of Glaucoma. For AMD detection, macular region analysis is important for early signs of the disease. Macula is the functional center of the retina. Accurate detection of these retinal landmarks can greatly improve diagnostic efficiency.



**Fig. 1.** Sample image from the REFUGE dataset showing retinal structures like an optic disc, cup, fovea, and blood vessels.

In recent years, several methods have been proposed for retinal structure detection. Most of the literature uses retinal features like the variation in intensities, texture and appearance for detecting OD and OC [3, 8, 13, 16]. The past few years have seen significant progress with deep learning approaches for OD with cup segmentation. In [5], an encoder-decoder network with deep residual structure and recursive learning mechanism is proposed for robust OD localization. An end-to-end region-based CNN for joint optic disc and cup segmentation (Joint-RCNN) is proposed in [6]. For automatic glaucoma screening, a Disc-aware Ensemble Network (DENet) is reported in [4], which integrates the local disc region with global information from the whole fundus image. Also, for fovea localization, many researchers have used different CNN models for the visibility of the macular region and fovea localization [1, 14]. A two-stage deep learning framework for accurate segmentation of the fovea in retinal color fundus images is presented in [14]. Recently, a simpler and more effective fovea localization algorithm based on the Faster R-CNN and physiological prior structure are presented in [21]. However, most of these methods treated either the disc with cup or the disc with fovea as two individual segmentation task.

Although many approaches have contributed work in OD, OC, and fovea segmentation, very few methods have considered all these tasks together. Since these retinal structures are spatially correlated to each other, there are advantages in combined detection and segmentation. The presence of retinal lesions in the macular region often occludes fovea, which is difficult to detect individually without using any spatial context. The fuzzy boundary of OC is often difficult to distinguish from OD and make this task quite challenging without any spatial prior.

To overcome these issues, this paper proposes a two-stage approach for segmenting retinal structures using the modified U-Net++ model with EfficientNet encoder. Our approach is free from the prior knowledge of retinal vessels. The major contributions of this work are summarized as follows:

1. We propose a two-stage approach for combined optic disc, cup, and fovea segmentation. In the first stage, a combined OD and fovea detection is performed, while in the next stage, the OD region is extracted and used for optic cup detection.
2. The proposed method uses EfficientNet-B4 encoder with modified U-Net++ architecture. The re-designed skip connection of U-Net++ and uses of concurrent channel and spatial excitation block in the decoder significantly improve the model performance. Also, the extracted features from EfficientNet show effective representations of retinal structures.
3. Our method evaluate on four different datasets including REFUGE [9], ADAM [22], IDRiD [10] and Drishti-GS [15]. Also, we have tested different variants of CNN models with extensive experimentation in comparison with state-of-the-art methods.

The rest of this paper has been organized as follows. Section 2 briefly introduces the proposed method for OD with cup and fovea segmentation. Section 3 describes the experimental results and analysis with discussion. In Sect. 4, we conclude the paper with ideas for future directions.

## 2 Methodology

### 2.1 Dataset

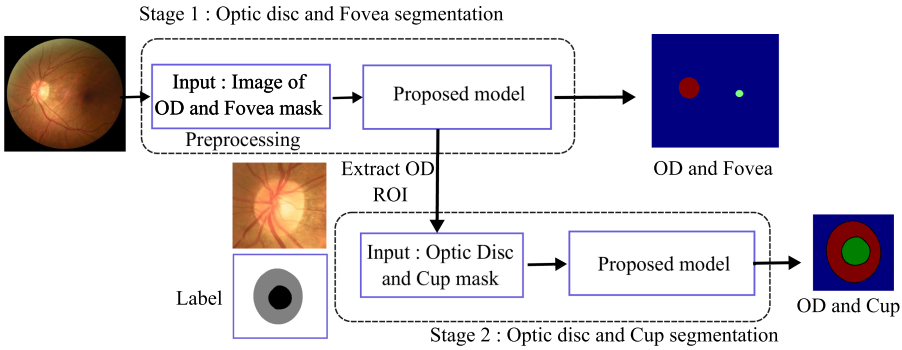
Four different dataset used in this paper namely, i-challenge ADAM [22] and REFUGE [9], Drishti-GS [15] and IDRiD [10]. The training, validation, and test set data comprised 400 images each in both the i-challenge dataset. The training images in ADAM and REFUGE were available in different sizes of  $2124 \times 2056$ ,  $1634 \times 1634$  and  $1444 \times 1444$ . In the first stage of the method, the ADAM dataset is used for OD with fovea segmentation. The REFUGE dataset is used in the second stage for OD with cup segmentation. Also, we empirically validated our approach on Drishti-GS [15] and IDRiD [10] test images for comparison with state-of-the-art methods.

**Pre-processing.** The main objective of the preprocessing module is to prepare the combined OD and fovea data for the segmentation. The fovea x,y center coordinates have been provided for both ADAM and REFUGE dataset. Further, we have created a fovea image with 50 pixels of the circular mask using the fovea coordinates and then combined them with OD mask images. The whole dataset images have been resized into  $512 \times 512$ . We utilized data augmentation including image blur, rotation, vertical and horizontal flip. Finally, with the augmentation factor of 5 on 800 images, the total 4000 images has been used for

the model development. Post this, we performed data normalization on all images by using mean subtraction with dividing standard deviation. These preprocessed images are provided as input to the proposed model.

## 2.2 Proposed Method

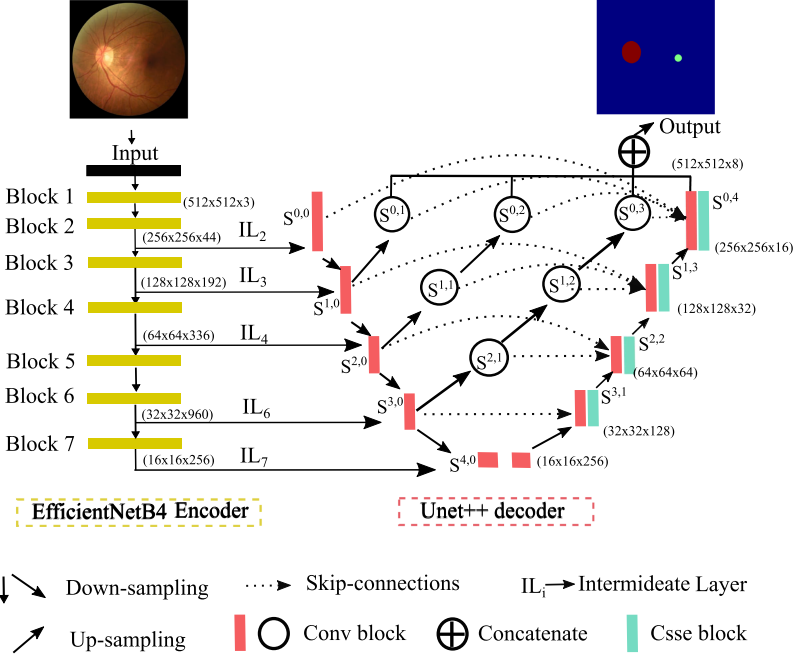
In this section, we provide an overview of the proposed method for the detection of OD, OC, and fovea. The two-stage approach consisted of combined detection of OD and fovea in the first stage. After that, the disc ROI of images is obtained by cropping a sub-image with the size of  $512 \times 512$  based on the center of the detected OD mask. Fine detection of OD boundaries with the optic cup is performed in the second stage. Our proposed method employs recent CNN models for the accurate detection of potential retinal structures from color fundus images. The block diagram of our approach is shown in Fig. 2.



**Fig. 2.** Propose two-stage flow diagram for optic disc, cup, and fovea segmentation. In stage one: Combine OD and Fovea detection. In stage two: The extracted OD region of interest (ROI) given input to the model for OD with cup detection.

**Model Architecture.** The proposed architecture consists of two parts namely encoder and decoder. Since, the U-Net++ nested skip pathways gives corresponding semantically rich feature maps. We have utilized the U-Net++ architecture of varying depths whose decoders are connected at the same resolution via a re-designed skip pathway [24]. Using a progression of skip pathways among decoder and encoder block, U-Net++ showed great success in segmentation tasks. In the recent deep learning era, EfficientNet has perform better in the ImageNet dataset for the classification task as compared to recent state-of-the-art backbone. Hence, we have explored the use of an EfficientNet [17] as an encoder for feature extraction with U-Net++ architecture as the baseline model. Due to the availability of a dense connection in U-Net++, every node in the decoder is represented in the feature maps and is aggregated from the previous and intermediate layers from the encoder. However, the dense connection of U-Net++

creates a larger size of the feature map because of concatenating similar features from different skip pathways. Hence, the number of trainable parameters in the existing U-Net++ model is high with more computational complexity. Therefore, we redesigned the skip pathways without loss of any information in the modified U-Net++. The accumulated feature maps denoted by  $s^{i,j}$  is calculated from Eq. (1).

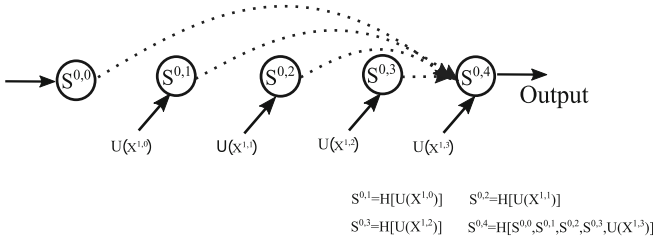


**Fig. 3.** Model architecture using Efficientnet-B4 as encoder with modified U-Net++ decoder.

$$s^{i,j} = \begin{cases} H(D(s^{i-1,j})), & j = 0 \\ H\left(\left[\begin{matrix} s^{i,k} \\ j=0 \end{matrix}\right]^{j-1}, U(s^{i+1,j-1})\right), & j > 0 \end{cases} \quad (1)$$

where,  $H(\cdot)$  is convolution operation,  $D(\cdot)$  and  $U(\cdot)$  denotes a down-sampling layer and an up-sampling layer respectively. Here,  $s^{i,j}$  represents the stack of feature maps which is also output from previous node  $S^{i,j}$ , where  $i$  and  $j$  are the downsample and convolution layer with the skip connection. The final segmented image obtained using concatenated the node output  $S^{0,1}$ ,  $S^{0,2}$ ,  $S^{0,3}$  and  $S^{0,4}$  of model which shown in Fig. 3.

The backbone of U-Net++ is the EfficientNet model, pre-trained on ImageNet, which proficiently separates various essential retinal anatomical structures. The principle block of EfficientNet is mobile reversed bottleneck convolutional (MBConv), which comprises of depthwise separable convolutional layers



**Fig. 4.** The basic block of proposed model with re-designed skip connections of U-Net++.

(DWConv). The model utilizes four DWConv layers and an ordinary convolutional layer with stride  $2 \times 2$  to down-sampling input size from  $512 \times 512$  to  $16 \times 16$ . The intermediate feature maps from five blocks of EfficientNet as  $IL_2$ ,  $IL_3$ ,  $IL_4$ ,  $IL_6$  and  $IL_7$  were extracted at different scales from encoder. We redesign the skip connections of U-Net++ to reduce the complexity of the baseline model as shown in Fig. 4. Also, the use of concurrent squeeze and spatial excitation (CSSE) block in decoder improves the performance [12]. At each intermediate layer level, all concatenated feature maps are merged on the ultimate node on that level. Finally, the concatenation layer combines all feature maps from transposed convolutional layers at the previous and the corresponding layer in the encoding pathway.

### 3 Results and Discussion

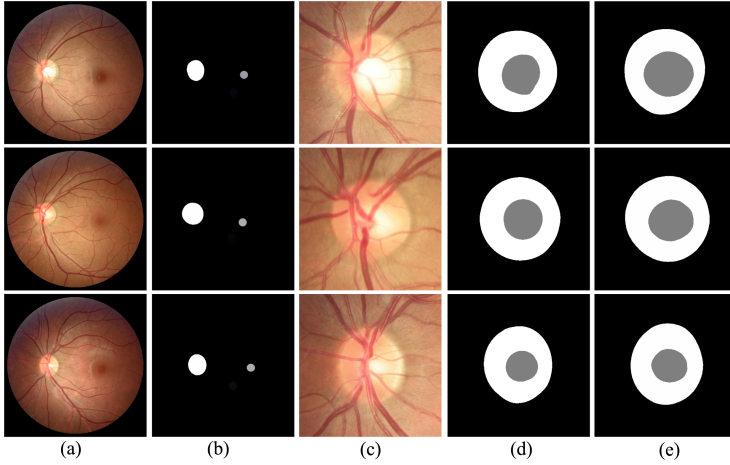
In this section, we first introduce the experimental setup and implementation details. We then provide experimental results with discussion in detail.

#### 3.1 Experimental Set-up

All the experiments were carried on resized images of  $512 \times 512$  pixels. We validated our proposed method on four datasets ADAM [22], REFUGE [9], Drishti-GS [15] and IDRiD [10]. In the experimental setup, the network was initialized with pre-trained weights on the ImageNet classification data. The model was trained using the adam optimizer with the learning rate of 0.0001, momentum was set to 0.95 and the batch size of 4 for 800 epochs. We have evaluated the hyperparameters of our method using the validation set, including learning rates, batch sizes, training epochs, and so on. The model was trained using Keras deep learning framework with an NVIDIA TITAN-RTX (24 GB) GPU.

#### 3.2 Results and Discussion

The dice coefficients (DI) and mean intersection over union (mIoU) use to evaluate the segmentation performance of the method. For OD segmentation, the



**Fig. 5.** Results on REFUGE validation dataset (a) Original image (b) Optic disc and fovea segmentation (c) Extracted OD patch (d) Ground truth of OD and OC (e) Predicted output of OD and cup segmentation.

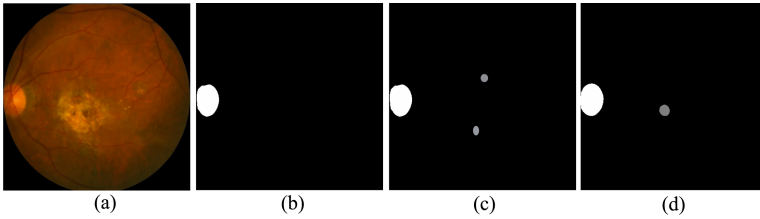
obtained dice is 0.9622 and 0.9474 on validation and test set of ADAM data. For OC segmentation, the obtained dice is 0.8816 and 0.8762 on validation and test set of REFUGE data. The segmentation results on the REFUGE validation dataset are shown in Fig. 5. We have detected both OD with fovea jointly and then localize the fovea center accurately from the fovea image. In the context of fovea localization, the fovea mask was prepared from the given x, y center coordinates. Finally, the best possible fovea location was found by calculating the centroid of the segmented fovea mask. The proposed method achieved the top rank for OD detection and segmentation task on the ADAM challenge testing dataset shown in Table 1.

**Table 1.** Results of top five performing teams for OD segmentation and fovea localization in ISBI ADAM challenge 2020 [22].

Team Name	$DI_{OD}$	F1-score	ED	Semi-final	Final	Score	Rank
Aira matrix	0.9474	0.9861	26.17	5	2	2.6	1
Xxlzt	0.9485	0.9912	284.03	10	1	2.8	2
Forbidden fruit	0.9420	0.9912	19.70	4	3	3.2	3
WWW	0.9445	0.9793	36.35	8	4	4.8	4
Vuno Eye Team	0.9370	0.9893	<b>18.55</b>	2	6	5.2	5

The average Euclidean distance between the predicted and ground truth for fovea localization is 30.23 and 26.17 pixels on IDRiD and ADAM test data.

We further validate the method on a test dataset from REFUGE and Drishti-GS for OD and OC segmentation. The performance comparison with different state-of-the-art methods on the REFUGE and Drishti-GS dataset are shown in Table 2. The EfficientNet-B4 feature extractor using the proposed model can able to detect the fovea despite of lesion present in the macular region. The accurate fovea segmentation results on the retinal image with the macular lesion are shown in Fig. 6. In addition, our method does not use any prior knowledge of vessel information for the detection of these retinal structures. Therefore, reduces the computational load compared to other approaches.



**Fig. 6.** Results of fovea detection in lesion image using different models on ADAM validation dataset (a) Original image (b) U-Net++ result with missing fovea detection (c) EfficientNet-U-Net++ result with false detection (d) Proposed model result with accurate fovea detection.

**Table 2.** The comparison of  $DI_{disc}$ ,  $DI_{cup}$ , and mIoU performance with state-of-the-art methods on the REFUGE and Drishti-GS dataset.

Method	REFUGE [9]			Drishti-GS [15]		
	$DI_{disc}$	$DI_{cup}$	mIoU	$DI_{disc}$	$DI_{cup}$	mIoU
Fu et al. [4]	0.9359	0.8648	0.8402	0.9658	0.8860	0.8588
Chen et al. [2]	0.9401	0.8674	0.8436	0.9655	0.8896	0.8594
Zhang et al. [23]	0.9529	0.8912	0.8670	0.9752	0.9314	0.8792
Wang et al. [19]	0.9460	0.8750	–	0.9740	0.9010	–
<b>Proposed</b>	<b>0.9573</b>	<b>0.8762</b>	<b>0.8725</b>	<b>0.9784</b>	<b>0.9381</b>	<b>0.8820</b>

**Ablation Study.** Recently published U-Net++ network showed the best performance over the vanilla U-Net [11]. The U-Net++ gives dense skip pathways to improve the performance [24]. However, theoretically dense skip pathways carry redundant features through the different skip connection and also increases the computational cost. Therefore, we redesigned the dense skip connections. From Table 3, the experimentation shows that the modified network outperformed over the U-Net++ and vanilla U-Net. Further, we introduce a heavy feature extractor namely EfficientNet-B4 [17] in encoder instead of vanilla encoder. We have



**Table 3.** The performance of different segmentation models and our proposed method for OD, OC, and fovea detection on the REFUGE test dataset. The baseline model is U-Net encoder with redesigned skip pathways.

Model	$DI_{disc}$	$DI_{cup}$	$IoU_{disc}$	$IoU_{cup}$	ED (pixels)	Parameters
U-Net [11]	0.9332	0.8537	0.8872	0.7683	46.32	7.7 M
U-Net++ [24]	0.9476	0.8537	0.8723	0.7668	43.81	21.3 M
Proposed-1	0.9479	0.8573	0.8766	0.7810	44.14	14.1M
Proposed-2	0.9535	0.8616	0.8814	0.8072	36.32	10.2 M
Proposed-3	<b>0.9573</b>	<b>0.8762</b>	<b>0.8847</b>	<b>0.8128</b>	<b>35.18</b>	<b>11.4 M</b>

Proposed-1 : U-Net encoder + skip pathways, Proposed-2 : EfficientNet + skip pathways, Proposed-3: EfficientNet + skip pathways + CSSE

trained all the models using similar hyperparameter setting. The performance of the proposed network is better than the existing models as shown in Table 3. In summary, our experiments gives the more accurate segmentation for the combined analysis of retinal structures.

## 4 Conclusion

In this paper, we have proposed a novel two-stage method for the detection of the optic disc with cup and fovea from fundus images. We have proposed a modified U-Net++ architecture with the EfficientNet-B4 model as a backbone for segmenting retinal structures. The redesigned skip connections of U-Net++ architecture reduces the computational requirements compared to the baseline model. We also performed extensive experiments on four public retinal fundus image datasets to demonstrate the effectiveness of our approach. We achieved the better results for the OD and OC with dice of 0.9573 and 0.8762 on REFUGE dataset. The proposed method is considered the top rank solution for optic disc detection and segmentation task in the ADAM challenge with dice of 0.9474. In the future, our approach can effectively make an impact on the retinal anatomical structure detection problem.

## References

1. Alais, R., Dokládal, P., Erginay, A., Figliuzzi, B., Decenci re, E.: Fast macula detection and application to retinal image quality assessment. *Biomed. Signal Process. Control* **55**, 101567 (2020)
2. Chen, H., Qi, X., Yu, L., Heng, P.A.: DCAN: deep contour-aware networks for accurate gland segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2487–2496 (2016)
3. Cheng, J., Yin, F., Wong, D.W.K., Tao, D., Liu, J.: Sparse dissimilarity-constrained coding for glaucoma screening. *IEEE Trans. Biomed. Eng.* **62**(5), 1395–1403 (2015)
4. Fu, H., et al.: Disc-aware ensemble network for glaucoma screening from fundus image. *IEEE Trans. Med. Imaging* **37**(11), 2493–2501 (2018)

5. Jiang, S., Chen, Z., Li, A., Wang, Y.: Robust optic disc localization by large scale learning. In: Fu, H., Garvin, M.K., MacGillivray, T., Xu, Y., Zheng, Y. (eds.) OMIA 2019. LNCS, vol. 11855, pp. 95–103. Springer, Cham (2019). [https://doi.org/10.1007/978-3-030-32956-3\\_12](https://doi.org/10.1007/978-3-030-32956-3_12)
6. Jiang, Y., et al.: JointRCNN: a region-based convolutional neural network for optic disc and cup segmentation. *IEEE Trans. Biomed. Eng.* **67**(2), 335–343 (2020)
7. Li, L., et al.: A large-scale database and a CNN model for attention-based glaucoma detection. *IEEE Trans. Med. Imaging* **39**(2), 413–424 (2020)
8. Mendonça, A.M., Melo, T., Araújo, T., Campilho, A.: Optic disc and fovea detection in color eye fundus images. In: Campilho, A., Karray, F., Wang, Z. (eds.) ICIAR 2020. LNCS, vol. 12132, pp. 332–343. Springer, Cham (2020). [https://doi.org/10.1007/978-3-030-50516-5\\_29](https://doi.org/10.1007/978-3-030-50516-5_29)
9. Orlando, J.I., et al.: Refugee challenge: a unified framework for evaluating automated methods for glaucoma assessment from fundus photographs. *Med. Image Anal.* **59**, 101570 (2020)
10. Porwal, P., et al.: Indian diabetic retinopathy image dataset (IDRiD): a database for diabetic retinopathy screening research. *Data* **3**(3), 25 (2018)
11. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)
12. Roy, A.G., Navab, N., Wachinger, C.: Concurrent spatial and channel ‘squeeze & excitation’ in fully convolutional networks. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11070, pp. 421–429. Springer, Cham (2018). [https://doi.org/10.1007/978-3-030-00928-1\\_48](https://doi.org/10.1007/978-3-030-00928-1_48)
13. Roychowdhury, S., Koozekanani, D.D., Kuchinka, S.N., Parhi, K.K.: Optic disc boundary and vessel origin segmentation of fundus images. *IEEE J. Biomed. Health Inf.* **20**(6), 1562–1574 (2016)
14. Sedai, S., Tennakoon, R., Roy, P., Cao, K., Garnavi, R.: Multi-stage segmentation of the fovea in retinal fundus images using fully convolutional neural networks. In: 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), pp. 1083–1086 (2017)
15. 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: 2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI), pp. 53–56. IEEE (2014)
16. Soares, I., Castelo-Branco, M., Pinheiro, A.M.G.: Optic disc localization in retinal images based on cumulative sum fields. *IEEE J. Biomed. Health Inf.* **20**(2), 574–585 (2016)
17. Tan, M., Le, Q.V.: Efficientnet: rethinking model scaling for convolutional neural networks. arXiv preprint [arXiv:1905.11946](https://arxiv.org/abs/1905.11946) (2019)
18. Ting, D.S.W., et al.: Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. *Jama* **318**(22), 2211–2223 (2017)
19. 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 Trans. Med. Imaging* **38**(11), 2485–2495 (2019)
20. Wong, W.L., et al.: Global prevalence of age-related macular degeneration and disease burden projection for 2020 and 2040: a systematic review and meta-analysis. *Lancet Global Health* **2**(2), e106–e116 (2014)

21. Wu, J., et al.: Fovea localization in fundus photographs by faster R-CNN with physiological prior. In: Fu, H., Garvin, M.K., MacGillivray, T., Xu, Y., Zheng, Y. (eds.) OMIA 2019. LNCS, vol. 11855, pp. 156–164. Springer, Cham (2019). [https://doi.org/10.1007/978-3-030-32956-3\\_19](https://doi.org/10.1007/978-3-030-32956-3_19)
22. Fu, H., et al.: Adam: automatic detection challenge on age-related macular degeneration (2020). <https://doi.org/10.21227/dt4f-rt59>
23. 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: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11764, pp. 442–450. Springer, Cham (2019). [https://doi.org/10.1007/978-3-030-32239-7\\_49](https://doi.org/10.1007/978-3-030-32239-7_49)
24. Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N., Liang, J.: UNet++: a nested U-Net architecture for medical image segmentation. In: Stoyanov, D., et al. (eds.) DLMIA/ML-CDS -2018. LNCS, vol. 11045, pp. 3–11. Springer, Cham (2018). [https://doi.org/10.1007/978-3-030-00889-5\\_1](https://doi.org/10.1007/978-3-030-00889-5_1)