Optic Disc Segmentation using Multiscale Low to High Resolution Convolutional Neural Networks

Dhruv Mohan

***Abstract*—This paper presents a framework for robust optic disc(OD) segmentation using Convolutional Neural Network(CNN). OD is an important anatomical landmark in the fundus image used for the diagnosis of ophthalmic pathologies. With the objective of unsupervised early and robust detection of diseases like Glaucoma and Diabetic Retinopathy we introduce two CNNs P-Net and Fine-Net which are arranged in a sequence to generate the OD segmentation map. P-Net generates a low-resolution(256x256) segmentation map which is upscaled and along with the input image is fed to Fine-Net which generates a high resolution segmentation map(1024x1024). The framework is trained jointly on three publicly available datasets, MESSIDOR, DRIONS-DB1 and DRISHTI-GS1. The proposed framework generalizes well as it performs reliably on test images with significant variability. For experimental evaluation we perform 5 fold cross validation and achieve OD localization in 99.4% of cases, moreover for OD segmentation we achieve an average Dice coefficient and IoU of 0.966 and 0.934 respectively.**

Keywords*—*Convolutional neural networks, Fundus, Optic Disc segmentation, Multi-scale

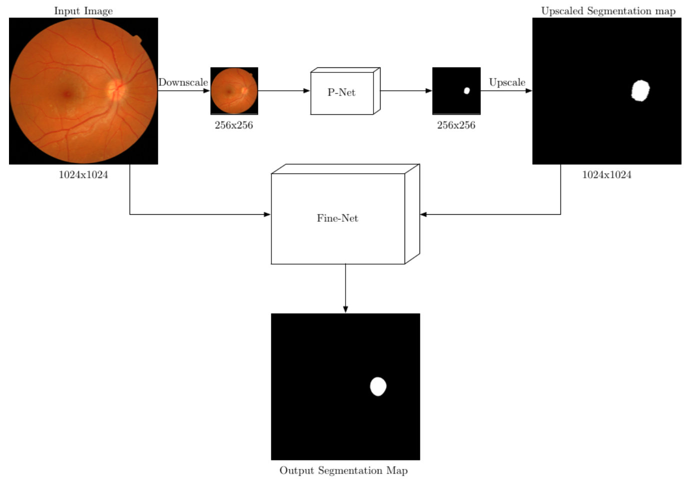


Fig. 1: Overall Architecture, Input image(1024x1024) is downscaled by a factor of 4(256x256) which serves as an input to P-net,output of P-Net is upscaled by a factor of 4 and added to the input image before being sent to Fine-Net

I. Introduction

Diabetic Retinopathy and Glucoma are common causes of blindness. It is estimated that by 2020 about 79 million people in the world are likely to suffer from glaucoma [1] and 93 million people are afflicted by diabetic retinopathy [2]. Diabetic retinopathy refers to retinal changes that occur in patients with diabetes mellitus. These changes affect the small blood vessels of the retina and can lead to vision loss. Glucoma is characterized by damage to the optic nerve due to excessive intraocular pressure. Early diagnosis of Glaucoma and Diabetic retinopathy is critical to preserving sight and preventing further degradation in vision.

OD is characterized as a bright elliptical or circular region in a coloured fundus image. Being a landmark in the fundus image OD localization serves help in registration of anatomical optical features such as the macula [3]. OD centre also serves as an initialize to vessel tracking algorithms [4]. Segmentation of the OD helps eliminate false responses associated with exudates in diabetic retinopathy. Optic cup area to OD area (cup to disc ratio) is an important parameter used for the diagnosis of glaucoma [5], [6].

A. Background

Prior work on OD segmentation can be broadly categorized into classical methods which rely on morphological algorithms, deformable models and template-based methods and recent methods which leverage machine learning for segmentation. Morales et al. [7] generate a greyscale image via PCA. The authors reason that PCA combines the most significant components of the RBG slices into a single image. Blood vessels are in-painted to extract the OD boundary more precisely. Stochastic watershed transformation is applied to the in-painted image and the OD contour is estimated as a circle. Jun Cheng et al. [8] propose a superpixel classification based method where the histogram and centre-surrounding statistics are employed to binarize each super pixel as Disc/No Disc. A circular Hough transform is used to model disc boundary in [9], [10]. However, since the OD is slightly elliptical, ellipse fitting method such as the one introduced in [11] perform better at OD segmentation. Datt et al. [12] extend the active contour without edges model [13] and propose a region based active contour model which uses a local energy function and level-set representation to segment the OD. The method introduced in Lowell et al. [14] works in two stages. First a a correlational filter is employed for OD localization where the peak of the filter output coincides with the OD center. Then, for segmentation the authors use Hus Circular Deformable model where the global model is replaced by an elliptical model and vector gradients and energy functions are used for fast nonlinear optimization. Lowell et al. also introduce a temporal lock algorithm for proper initialization of Hus circular Deformable model. Another method based on deformable models is introduced in Xu et al. [15]concerning morphological based algorithms in literature we have Eswaran et al. [16] and Welfer et al. [17] In both approaches OD boundary is detected via watershed transform. The authors also reason that green channel of an RGB image contains good contrast between the background and the bright retinal components. [17] uses the vascular tree in the fundus image is used to locate the OD where as [16] uses an averaging filter, contrast stretching and minima imposition to pre-process the image.

The onset of Deep Convolutional Neural Networks(CNN) has revolutionized computer vision. Initially demonstrating their efficacy for classification tasks [18] CNN have branched out and successfully applied to object detection [19], semantic segmentation [20] and image generation [21]. For the task on optic disc segmentation authors of [22] use a modified UNET to generate a segmentation map. Maninis et al. [23] propose an architecture based on VGG16 with specialised layers for retinal vasculature segmentation and optic disc segmentation. Zilly et al [24] introduce entropy sampling to train their CNN architecture which is used for feature extraction.

The method proposed in the paper introduces a two stage multiscale approach for segmentation. Figure 1 presents the entire pipeline of our method. We introduce specialized networks for each stage; P-Net and Fine-Net. The first stage involves the generation of a segmentation map via P-Net by downscaling the input image by a factor k. The generated map is then upscaled by k and added to the input image, this serves as an input to Fine-Net which generated the final high-resolution segmentation map. In all figures r is the dilation rate, c is the number of filters in a convolution layer, s is the stride. Default r and s is 1. Feature maps are max-pooled with a pooling rate of 2 and upscaled bilinearly.

II. Method

A. P-Net

While testing we observed that introduction of a low-resolution segmentation map prior as an input to a model decreased convergence time and increased the mean IoU score. To facilitate the generation of the segmentation map prior we introduce P-Net(fig3). The design is motivated by the need to be computationally inexpensive while maintaining high performance. ResNets [25] introduce parameter free shortcut connections to enable easier optimization of the network. Since each layer has it's own weights ResNets have a substantially large number of parameters. Huang et al. [26] Introduce paramerically efficient information flow by connecting the l^{th} layer to all subsequent layers and giving it the concatenated feature maps of all preceding layers as input

$$x_l = \mathcal{H}_l(c_{l-1}) \tag{1}$$

where x_l is the output feature map and c_{l-1} is the concatenated tensor of feature maps from layers 0 to $l-1$. Concatenating feature maps leads to an increased variance which in-turn allows narrow networks will less filters achieve similar performance to their wider counterparts. On the ImageNet ILSVRC-2012 dataset [27] DenseNet with half as many parameters as ResNet (20M vs 44M) achieves similar top-1 classification.

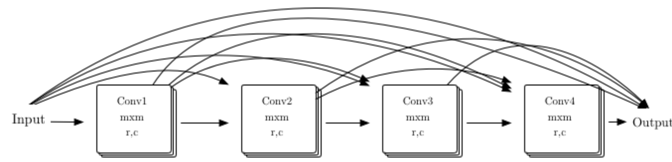


Fig. 2: Dense Atrous Block(DB)

Pooling operations such as max-pooling allow CNN to learn invariance to local image transformations. This learnt invariance benefits classification tasks but is undesirable for spacio-critical tasks like semantic segmentation. Chen et al. [28] bypass pooling layers and lift the spacial resolution constrains by introducing Atrous convolutions which allows spacial resolution control during feature extraction. The Atrous convolution between input f, kernel k and dilatation rate r can be defined as:

$$y_{i,j} = \sum_{u=1}^{m} \sum_{v=1}^{n} k_{u,v} f_{i-r.u,j-r.v} \tag{2}$$

Input to P-net is fed to 3x3 convolutional layer which is subsequently max-pooled with a stride of 2x2 and passed through another 3x3 convolutional layer. Activations of this convolutional layer is sent to parallel set of dense atrous blocks(DB) visualized in figure 2 each configured with a different dilation rate. Activations of each DB are concatenated. The orientation of DBs is similar to Atrous Spatial Pyramid Pooling proposed(ASPP) in [29] where different dilation rates capture multiscale information. Subsequent activations are bilinearly scaled and passed through a 3x3 convolution layer. Finally, we obtain logits via a 1x1 convolution layer.

B. Fine-Net

We design Fine-net(fig. 6) with an emphasis on localization accuracy and model it after Full Resolution Residual Networks(FRRN) [30] where the Pohlen et al. propose using two streams for data propagation through the network. One stream in maintained at full resolution and the other is set up as an encoder-decoder. In doing so FRRN combine multi-scale context with pixel level accuracy. Full resolution residual units(FRRU) can represented as:

$$p_{out} = \mathcal{G}([\mathcal{D}(r_{in}, t), p_{in}]; \mathcal{W}_g) \tag{3a}$$

$$r_{out} = \mathcal{U}(p_{out}, \mathcal{W}_u) + r_{in} \tag{3b}$$

where r_{in} and p_{in} are inputs from the residual and pooling stream. $\mathcal{D}$ is a downscaling function with scaling rate t. $\mathcal{W}_g$ are the parameters of function $\mathcal{G}$ which takes concatenated pooling stream and downscaled residual stream as input. $\mathcal{U}$ is an upscaling function with parameters $\mathcal{W}_u$ and p_{out} as input. r_{out} and p_{out} are the outputs for the residual and pooling stream respectively.

Using a sequence of FRRU [30] presented two architectures FRRN-A with four maxpool-unpool pairs and FRRN-B with

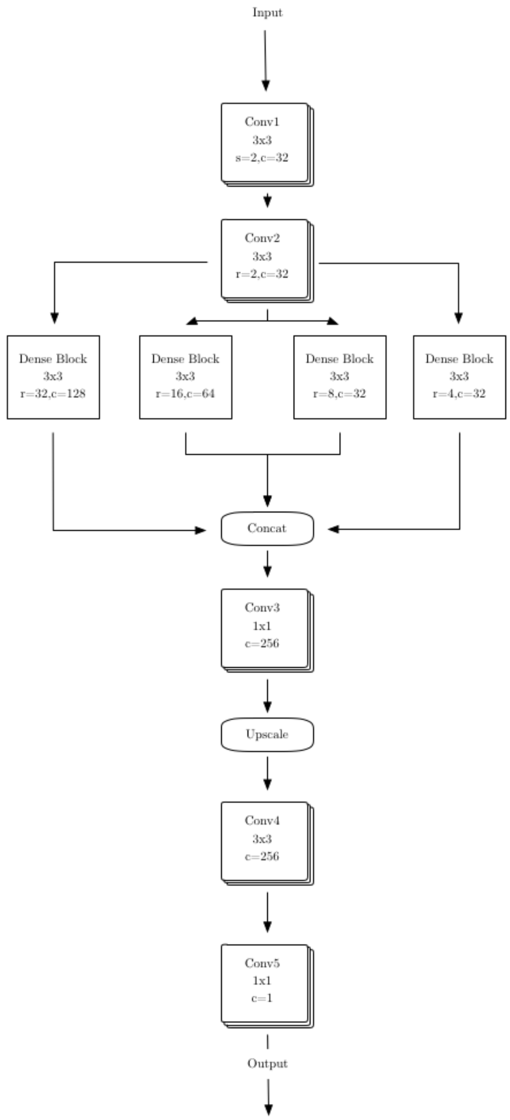


Fig. 3: P-net

five maxpool-unpool pairs. Both architectures are too memory intensive for our objective. To alleviate the intrinsic memory intensive disposition of FRRN without sacrificing performance we introduce Atrous convolutions in FRRU(fig 4.) and propose Fine-Net where we encapsulate 6 serially connected FRRU blocks with increasing dilation rates and base channels within a maxpool-upscale pair.

C. Training

The GT images have an inherent unbalanced class distribution. To introduce class balancing we use bootstrapped cross entropy loss introduced in [31]. Since OD segmentation is essentially a binarization problem we simplify the loss function

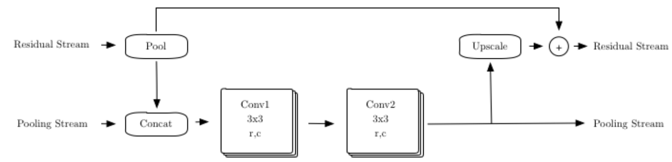


Fig. 4: Full Resolution Residual unit with atrous convolutions

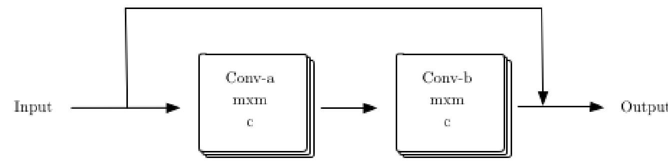


Fig. 5: Residual Unit used in Fine-Net

to a bootstrapped sigmoid cross entropy loss:

$$s = \mathcal{S}(y)$$

$$m_{i,j} = \sum_{i=1}^{m}\sum_{j=1}^{n} 1\{y_{i,j} > t\}$$

$$\mathcal{B} = \frac{\sum(y + y * l * m + \log(1 + \exp(-y)))}{\sum m} \tag{4}$$

We also leverage Dice loss:

$$\mathcal{D} = 1 - \frac{2|s \cap l|}{|s| + |l|} \tag{5}$$

Here y is output of Fine-Net $\mathcal{S}$ is the sigmoid activation function, l is the ground truth label m is a mask generated by thresholding y. The final objective function that both networks minimize is defined as:

$$\mathcal{L} = \lambda_{bce}\mathcal{B} + \lambda_{dice}\mathcal{D} \tag{6}$$

Where λ_{bce} and λ_{dice} are hyper-parameters that control the relative importance $\mathcal{B}$ and $\mathcal{D}$

Both P-Net and Fine-Net are optimized using the ADAM optimizer introduced in Kingma et al. [32] with epsilon=0.1 for 20 epochs with gradient accumulation over 2 iterations. Learning rate of P-Net = 0.001 and Learning rate of Fine-Net = 0.05 both learning rates have a decay of 0.86 every 5 epochs. Train time augmentation in the form of random horizontal mirroring and gamma augmentation is performed to prevent over-fitting.

III. Experimental Validation

We train both P-Net and Fine-Net on three publicly available datasets MESSIDOR [33], Drishti-GS1 [34] and Drions-DB [35]. MESSIDOR dataset contains 1200 eye fundus color images at a resolution of 1440×960, 2240×1488, or 2304×1536. DRIONS-DB contains 110 fundus image at 600×400 pixels and DRISHTI-GS1 has 101 fundus images at 2896 × 1944

Method	MESSIDOR		DRISHTI-GS1		DRIONS-DB	
	Dice	IoU	Dice	IoU	Dice	IoU
Zilly et al. [24]	-	-	0.947	89.5	-	-
Mannis et al. [23]	-	-	-	-	**0.97**	0.88
Sevatoplsky et al. [22]	-	-	-	-	0.94	0.89
Morales et al. [7]	0.8950	0.8228	-	-	0.9084	0.8424
Cheng et al. [8]	-	0.875	0.897	0.93	-	-
Ours:P-Net+Fine-Net	**0.968**	**0.939**	**0.9713**	**0.947**	0.966	**0.935**
Ours: Fine-Net	0.957	0.92	0.964	0.931	0.955	0.914

TABLE I: Dice and IoU scores on MISSIOR DRISHTI-GS1 and DRIONS-DB, - indicates missing datapoint

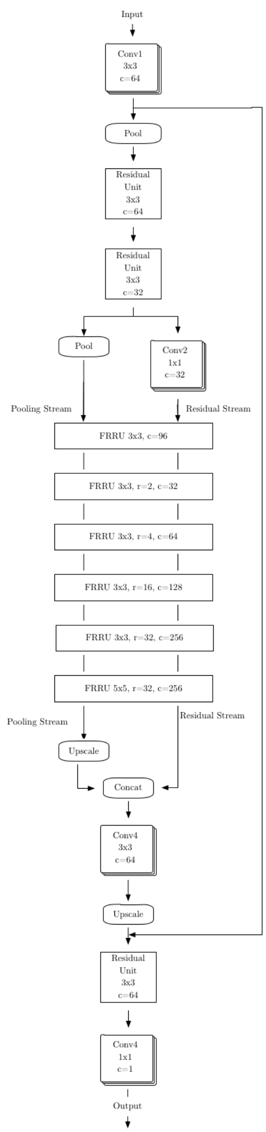
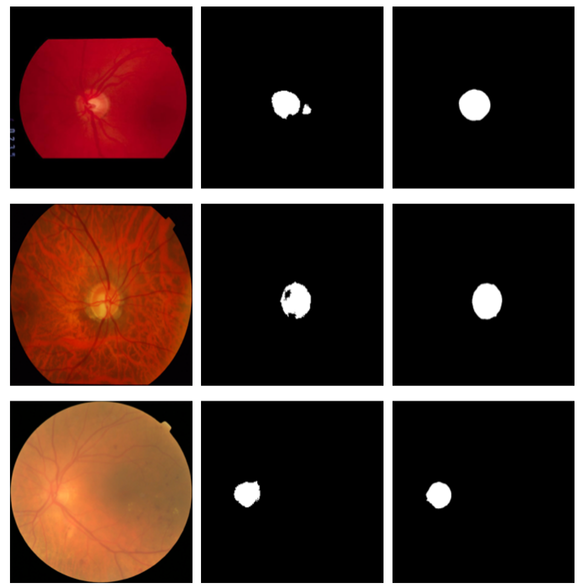
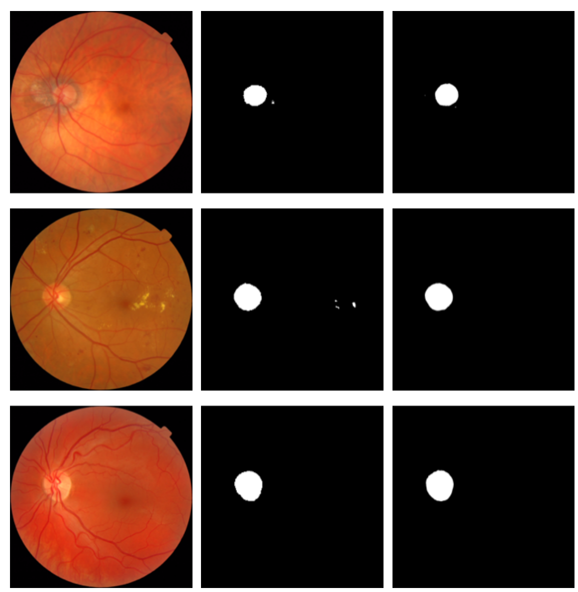


Fig. 6: Fine-net



(a) (b) (c)



(d) (e) (f)

Fig. 7: Results: Input images are in (a) and (d). (b) and (e) are the respective P-Net output and (c), (f) visualize the Fine-Net output

pixels. The expert marked ground truth OD segmentation maps for DRISHTI-GS1 and DRIONS-DB were provided with the dataset whereas for MESSIDOR we obtained the OD segmentation maps from [36]. Images of all three datasets have different resolutions and aspect ratios. To standardize images for training the fundus in MESSIDOR and DRISHTI is cropped via thresholding and the cropped images are scaled to 1024x1024. The images in DRIONS are zero padded to 600x600 and upscaled to 1024x1024. We obtain a combined set of 1411 images, which elicits the need of multi-fold cross validation. Each fold is constructed by sampling T/K images from each dataset. Where T is the number of images in each dataset and K is the total number of folds.

We present the results of optic disc segmentation on MESSIDOR DRISHTI-GS1 and DRIONS-DB after 5 fold cross-validation. Evaluation is performed on Dice score: $\frac{2|P \cap L|}{|P|+|L|}$ and IoU score: $\frac{|P \cap L|}{|P \cup L|}$ where P is the output segmentation map and L is the ground truth. We compare our results against other state of the art CNN based approaches defined in [22]–[24], as well as other non CNN based methods [7], [8].Table 1 summarizes the performance of our model versus others.

IV. Conclusion

In this work we present a two stage multi-resolution architecture for optic disc segmentation from retinal fundus image. To facilitate this architecture we introduced P-Net for the generation of a low resolution segmentation map and Fine-Net which takes the output of P-Net as input and performs high resolution segmentation. The architecture is tested against three publicly available datasets and obtains state of the art results.

References

[1] H A Quigley and A T Broman. The number of people with glaucoma worldwide in 2010 and 2020. *British Journal of Ophthalmology*, 90(3):262–267, 2006.

[2] Joanne W.Y. Yau and Rogers. Global prevalence and major risk factors of diabetic retinopathy. *Diabetes Care*, 35(3):556–564, 2012.

[3] Meindert Niemeijer, Michael Abramoff, and Bram Ginneken. Fast detection of the optic disc and fovea in color fundus photographs. 13:859–870, 12 2009.

[4] Langis Gagnon, Marc Lalonde, M Beaulieu, and Marie Boucher. Procedure to detect anatomical structures in optical fundus images, 07 2001.

[5] J. Liu, D. W. K. Wong, J. H. Lim, H. Li, N. M. Tan, Z. Zhang, T. Y. Wong, and R. Lavanya. *ARGALI: An Automatic Cup-to-Disc Ratio Measurement System for Glaucoma Analysis Using Level-set Image Processing*, pages 559–562. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.

[6] Yuji Hatanaka, Atsushi Noudo, Chisako Muramatsu, Akira Sawada, Takeshi Hara, Tetsuya Yamamoto, and Hiroshi Fujita. *Automatic Measurement of Vertical Cup-to-Disc Ratio on Retinal Fundus Images*, pages 64–72. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010.

[7] S. Morales, V. Naranjo, J. Angulo, and M. Alcaiz. Automatic detection of optic disc based on pca and mathematical morphology. *IEEE Transactions on Medical Imaging*, 32(4):786–796, April 2013.

[8] J. Cheng, J. Liu, Y. Xu, F. Yin, D. W. K. Wong, N. M. Tan, D. Tao, C. Y. Cheng, T. Aung, and T. Y. Wong. Superpixel classification based optic disc and optic cup segmentation for glaucoma screening. *IEEE Transactions on Medical Imaging*, 32(6):1019–1032, June 2013.

[9] Xiaolu Zhu and R. M. Rangayyan. Detection of the optic disc in images of the retina using the hough transform. In *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 3546–3549, Aug 2008.

[10] A. Aquino, M. E. Gegundez-Arias, and D. Marin. Detecting the optic disc boundary in digital fundus images using morphological, edge detection, and feature extraction techniques. *IEEE Transactions on Medical Imaging*, 29(11):1860–1869, Nov 2010.

[11] J. Cheng, J. Liu, D. W. K. Wong, F. Yin, C. Cheung, M. Baskaran, T. Aung, and T. Y. Wong. Automatic optic disc segmentation with peripapillary atrophy elimination. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 6224–6227, Aug 2011.

[12] Gopal Datt Joshi, Rohit Gautam, Jayanthi Sivaswamy, and S. R. Krishnadas. Robust optic disk segmentation from colour retinal images. In *Proceedings of the Seventh Indian Conference on Computer Vision, Graphics and Image Processing*, ICVGIP '10, pages 330–336, New York, NY, USA, 2010. ACM.

[13] T. F. Chan and L. A. Vese. Active contours without edges. *IEEE Transactions on Image Processing*, 10(2):266–277, Feb 2001.

[14] J. Lowell, A. Hunter, D. Steel, A. Basu, R. Ryder, E. Fletcher, and L. Kennedy. Optic nerve head segmentation. *IEEE Transactions on Medical Imaging*, 23(2):256–264, Feb 2004.

[15] Juan Xu, Opas Chutatape, Eric Sung, Ce Zheng, and Paul Chew Tec Kuan. Optic disk feature extraction via modified deformable model technique for glaucoma analysis. *Pattern Recogn.*, 40(7):2063–2076, July 2007.

[16] C. Eswaran, A. W. Reza, and S. Hati. Extraction of the contours of optic disc and exudates based on marker-controlled watershed segmentation. In *2008 International Conference on Computer Science and Information Technology*, pages 719–723, Aug 2008.

[17] Daniel Welfer, Jacob Scharcanski, Cleyson M. Kitamura, Melissa M. Dal Pizzol, Laura W.B. Ludwig, and Diane Ruschel Marinho. Segmentation of the optic disk in color eye fundus images using an adaptive morphological approach. *Computers in Biology and Medicine*, 40(2):124 – 137, 2010.

[18] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc., 2012.

[19] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Region-based convolutional networks for accurate object detection and segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(1):142–158, Jan 2016.

[20] Evan Shelhamer, Jonathan Long, and Trevor Darrell. Fully convolutional networks for semantic segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 39(4):640–651, 2017.

[21] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 2672–2680. Curran Associates, Inc., 2014.

[22] A. Sevastopolsky. Optic disc and cup segmentation methods for glaucoma detection with modification of u-net convolutional neural network. *Pattern Recognition and Image Analysis*, 27(3):618–624, Jul 2017.

[23] Kevis-Kokitsi Maninis, Jordi Pont-Tuset, Pablo Andrés Arbeláez, and Luc Van Gool. Deep retinal image understanding. In *MICCAI*, 2016.

[24] Julian G. Zilly, Joachim M. Buhmann, and Dwarikanath Mahapatra. Boosting convolutional filters with entropy sampling for optic cup and disc image segmentation from fundus images. In *MLMI*, 2015.

[25] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.

[26] Gao Huang, Zhuang Liu, and Kilian Q. Weinberger. Densely connected convolutional networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2261–2269, 2017.

[27] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3):211–252, Dec 2015.

[28] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L. Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 2017.

[29] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. Rethinking atrous convolution for semantic image segmentation. *CoRR*, abs/1706.05587, 2017.

[30] Tobias Pohlen, Alexander Hermans, Markus Mathias, and Bastian Leibe. Full-resolution residual networks for semantic segmentation in street scenes. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3309–3318, 2017.

[31] Zifeng Wu, Chunhua Shen, and Anton van den Hengel. Bridging category-level and instance-level semantic image segmentation. *CoRR*, abs/1605.06885, 2016.

[32] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2014.

[33] Etienne Decencire, Xiwei Zhang, Guy Cazuguel, Bruno Lay, Batrice Cochener, Caroline Trone, Philippe Gain, Richard Ordonez, Pascale Massin, Ali Erginay, Batrice Charton, and Jean-Claude Klein. Feedback on a publicly distributed database: the messidor database. *Image Analysis & Stereology*, 33(3):231–234, August 2014.

[34] Joshi G. D Jain Madhulika Ujjwal Sivaswamy J, Krishnadas K. R and Syed Abbas T. Drishti-gs: Retinal image dataset for optic nerve head(onh) segmentation. *IEEE ISBI*, 2015.

[35] Enrique J. Carmona, Mariano Rincn, Julin Garca-Feijo, and Jos M. Martnez de-la Casa. Identification of the optic nerve head with genetic algorithms. *Artificial Intelligence in Medicine*, 43(3):243 – 259, 2008.

[36] Expert system for early automated detection of dr by analysis of digital retinal images project website univ. huelva. September 2012.