REFUGE Competition Report

Jaemin Son[1], Jeyoung Kim[2]

[1]VUNO Inc., Seoul, Korea, [2]Gachon University, Gyeonggi-do, Republic of Korea

1 Glaucoma Classification

CAVEAT: **Details of the method such as network architecture, loss function and subdivision of the fundus are also explained in detail in our accepted paper at OMIA 2018.**

We first assigned labels to public datasets (kaggle[1], messidor [3], IDRiD [7]) with a CNN that was trained with in-house dataset (The paper is to appear at OMIA2018). Then, we retrained a CNN with the same architecture from the scratch. The given dataset was used only for validation purpose. Source code that we used is available[2].

1.1 In-house dataset

We used in-house dataset for training the model that assigns labels to the public dataset. The data collection was approved by the institutional review board at Seoul National Bundang Hospital (SNUBH) (IRB No. B-1508-312-107) and conducted in accordance with the tenets of the Declaration of Helsinki. Macula-centered retinal fundus images were collected at the Seoul National University Bundang Hospital, from June 1st, 2003 to June 30th, 2016 at the health screening center and ophthalmology outpatient clinic. Various fundus cameras (Kowa nonmyd7, Kowa VX-10, Canon CF60Uvi, GENESIS D portable, Canon CR6-45NM, Kowa VX-10a) were used. 57 licensed ophthalmologists including 16 certified retina specialists, 9 certified glaucoma specialists, and 3 certified cornea specialists assessed the images.

1.2 Network Architecture

Fig. 1 shows our network architecture which consists of residual layer (feature maps after residual unit [5]), reduction layer (feature maps after 3×3 conv with stride 2, batch-norm, ReLU), average pooling layer, atrous pyramid pooling layer [2] and 1×1 conv (depth=1) layer.

We also exploit the fact that the progress of glaucoma accompanies with Retinal Nerve Fiber Layer (RNFL) defect and changes in the optic disc which can be measured by cup-to-disc ratio. RNFL defect and glaucomatous disc changes occur at specific locations in the fundus. Therefore, by localizing the centers of

[1] `https://www.kaggle.com/c/diabetic-retinopathy-detection`
[2] `https://bitbucket.org/woalsdnd/refuge/src`

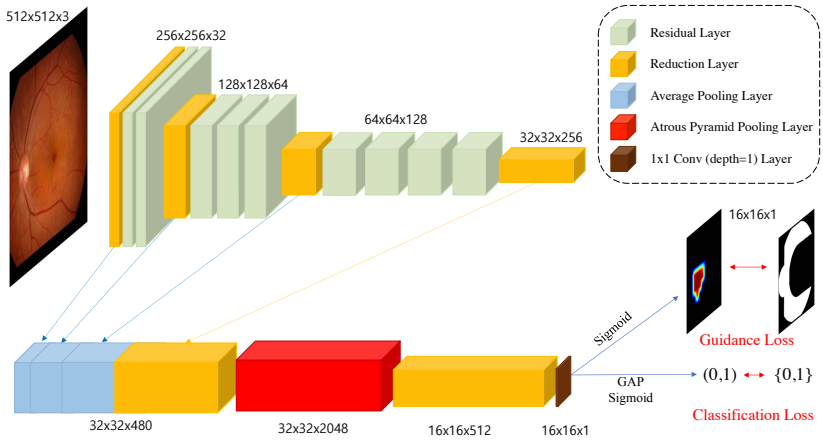

Fig. 1: Proposed network architecture for glaucoma classification.

optic disc and fovea, it is possible to separate the fundus into several subregions and mask the subregions that relate to glaucoma. We used the localization models that took 2nd place at ISBI IDRiD challenge[3] in the optic disc and fovea detection subtasks (team name - VRT). An example of subregions in the fundus and a pair of a fundus image and its mask are shown in Fig. 2.

1.3 Loss Function

For a fundoscopic image ($I \in R^{W_I \times H_I}$), the existence of a target finding in the image I is encoded to $y_{true} \in \{0, 1\}$ and the probability of the existence $y_{pred} \in (0, 1)$ is output from the network. When k images are given as a mini-batch, binary cross entropy for classification loss in Fig. 1 is given by

$$L_{class}(\mathbf{y}_{true}, \mathbf{y}_{pred}) = \frac{1}{k} \sum_{i=1}^{k} \left[-y^i_{true} \log y^i_{pred} - (1 - y^i_{true}) \log(1 - y^i_{pred}) \right] \quad (1)$$

where $\mathbf{y}_{true} = \{y^1_{true}, ..., y^k_{true}\}$ and $\mathbf{y}_{pred} = \{y^1_{pred}, ..., y^k_{pred}\}$.

When the last feature maps are size of $W_F \times H_F$, a region mask for a target finding ($M \in \{0, 1\}^{W_F \times H_F}$) is given as label and the activation map ($A \in (0, 1)^{W_F \times H_F}$) is generated from the network. With a mini-batch of size k, guidance loss in Fig. 1 is given by

$$L_{guide}(\mathbf{A}, \mathbf{M}) = \frac{1}{kW_F H_F} \sum_{i=1}^{k} \sum_{l=1}^{W_F H_F} (1 - m^i_l) \log(\max(a^i_l, \epsilon)) \quad (2)$$

[3] `https://idrid.grand-challenge.org/leaderboard/`

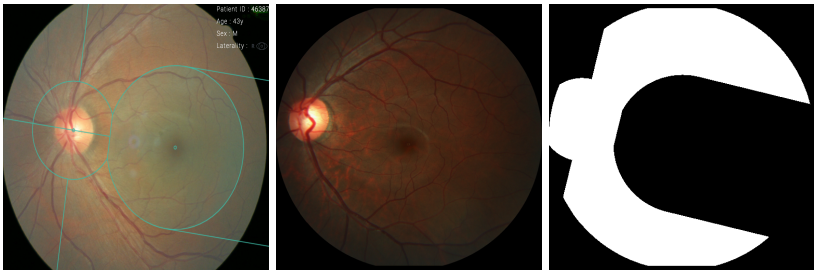

Fig. 2: Example of subregions in the fundus and a pair of a fundus image and the mask for glaucomatous findings (RNFL defect and glaucomatous disc changes). The fundus is divided into 8 regions. When the distance between the optic disc and fovea is D, circles are drawn at the centers of the optic disc and fovea with the radius of $\frac{2}{5}D$ and $\frac{2}{3}D$ and the intersections of the two circles are connected with a line segment. Then, a half-line passing through the optic disc and fovea (L) cuts the circle of the optic disc in half and two half-lines parallel to L and tangent to the circle of fovea are drawn in a direction away from the optic disc. Finally, a line perpendicular to L is drawn to pass through the center of the optic disc. Masks for glaucoma are generated to mark up optic disc areas and upper and lower temporal areas.

where $\mathbf{A} = \{A^1, ..., A^k\}$ and $\mathbf{M} = \{M^1, ..., M^k\}$ and m_l^i and a_l^i are values at lth pixel in M^i and A^i for $l = 1, ..., W_F H_F$. Note that $\epsilon > 0$ is added inside the logarithm for numerical stability when $a_l^i \approx 0$. In experiments, we used $\epsilon = 10^{-3}$. In a nutshell, the guidance loss suppresses any activation (a_l^i) in regions where the value of the mask is 0 ($m_l^i = 0$) and has no effect for activation inside the mask ($m_l^i = 1$).

Then, total loss is given by combining the classification loss and the guidance loss,

$$L_{total} = L_{class}(\mathbf{y}_{true}, \mathbf{y}_{pred}) + \lambda L_{guide}(\mathbf{A}, \mathbf{M}) \tag{3}$$

where λ balances two objective functions and we used $\lambda = 1$

1.4 Assigning labels to public datasets

We ran the trained network to assign labels to public datasets. The network assigned value between 0 and 1 indicating the possibility of glaucoma to 89,803 images. Then, images with the value below 0.1 were labeled as non-glaucoma (82,608) and images with the value above 0.5 (3,000) were labeled as glaucoma.

1.5 Training with the Generated Labels

Total 85,608 images (3,000 positive and 82,608 negative) were used to train a CNN from the scratch with the REFUGE dataset as a validation set. A network that yielded the highest AUROC on the REFUGE dataset was chosen.

We used the same architecture in the paper (to appear at OMIA2018). Original color fundoscopic images were resized to 512×512 for the network input. The resized images were randomly augmented by affine transformation (flip, scaling, rotation, translation) and random re-scaling of the intensity. Weights and biases were initialized with Xavier initialization [4]. As an optimizer, we used SGD with *nestrov* momentum 0.9 and decaying learning rate starting from 10^{-3} to 10^{-4}. Batch size was set to 32. We oversampled the positive images so that roughly a minibatch maintains the ratio of 1:1 for positive and negative images. The network were trained for 50 epochs and stopped as AUROC did not improve.

Unlike the paper, we standardized the image by subtracting mean and dividing by standard deviation. When computing mean and std, pixels that have an intensity less than 10 were excluded from the statistics. The choice of standardization instead of normalization yielded slightly high AUROC ($\approx 0.3\%$) in the REFUGE training set. The network yielded AUROC of 0.9523 on the training dataset.

To boost the performance further, we have ensembled three external networks that were trained with SNUBH dataset that detects (1) glaucomatous disc change, (2) RNFL defect and (3) glaucoma suspect. An image was augmented to 10 inputs and the prediction results of them were averaged into a single value for each of 4 networks (1 public + 3 external). Then, the max was chosen among two prediction values for RNFL defect and glaucoma disc change and averaged with the result from the original network and the value for glaucoma suspect.

2 Optic Disc and Cup Segmentation

Our network is in essence similar to U-NET [6] that can be separated into encoder and decoder parts. We segment the disc and the cup separately from networks that have a different architecture. The characteristic aspect of our models is that another CNN takes a vessel segmentation mask [8] and generates a coarse mask that approximately indicates the position of the disc or the cup. The coarse mask is concatenated to the bottleneck layer of a CNN that takes a fundus image as input and generates the final segmentation mask. The idea is similar to 2nd place at ISBI IDRiD challenge[4] in the optic disc segmentation. However, we did hyperparameter-search and selected the best architecture. After selecting the architecture, we trained a network and picked up the network that yields high dice coefficient on the reserved validation dataset (g0006.bmp, n0009.bmp, n0049.bmp, n0125.bmp, n0159.bmp, n0298.bmp, n0003.bmp, n0040.bmp, n0110.bmp, n0144.bmp, n0271.bmp, n0354.bmp). The reserved validation dataset included hard examples.

3 Disc Segmentation

We have used IDRiD dataset [7], riga dataset [1] and the given training dataset. The final network architecture is shown in table 1. We trained with the loss

[4] `https://idrid.grand-challenge.org/leaderboard/`

$L_{total} = L_{Main} + \lambda * L_{vessel}$ where L_{Main}, L_{vessel} are pixel-wise binary cross entropy for the main branch and the vessel branch.

We have done grid search to choose an architecture by only changing (1) depth of the U-NET (2) the number of filters for conv6 at the vessel branch (3) lambda value in the loss $L_{total} = L_{Main} + \lambda * L_{vessel}$. Other variables were fixed when the effect of one variable was experimented. In our experiments, $\lambda = 2$ was the best. The depth was chosen from $\{3, 4, 5\}$ and the number of filters for conv6 was chosen from $\{255, 510, 768, 1023\}$ and the λ was chosen among $\{0.1, 0.5, 1, 2\}$.

The network outputs values ranging from 0 to 1 for each pixel. The threshold was found to maximize dice coefficient in the reserved validation dataset which was 0.5. The cup was considered as the disc and any holes and non-convex areas were filled in after the thresholding. The final result scored dice coefficient of 0.9490 on the reserved validation dataset.

4 Cup Segmentation

We have used riga dataset [1] and the given training dataset. The final network architecture is shown in table 2.

Similar to the disc segmentation, grid search was done for each variable which is (1) the depth of U-Net, (2) the number of filters for conv6 at the vessel branch (3) lambda value in the loss $L_{total} = L_{Main} + \lambda * L_{vessel}$. In our experiments, $\lambda = 0.1$ was the best for the cup. The depth was chosen from $\{3, 4, 5\}$ and the number of filters for conv6 was chosen from $\{256, 512, 768, 1024\}$ and the λ was chosen among $\{0.1, 0.5, 1, 2\}$.

The network outputs values ranging from 0 to 1 for each pixel. The threshold was found to maximize dice coefficient in the reserved validation dataset which was 0.6575. After the thresholding, any holes and non-convex areas were filled in. The network yielded dice coefficient of 0.8039 on the reserved validation dataset.

5 Fovea Localization

We used the fovea localization model that took 2nd place at ISBI IDRiD challenge[5] in fovea detection subtasks (team name - VRT) which is in preparation for publication. In a nutshell, one branch takes a vessel image as input and output low resolution segmentation which is concatenated to the bottleneck layer of U-Net that takes a fundus image as input. Only two minor changes were made. One is that no loss was applied when L1 loss becomes below 0.015 (9.6 pixels in 640×640 image) for the vessel branch. Second is that a coefficient that balances losses for the vessel branch and the image branch was chosen through random search.

The given dataset was used solely for validation and a model that yielded the lowest L1 loss was chosen. The result was 7.2 pixels in 640×640 image.

[5] `https://idrid.grand-challenge.org/leaderboard/`

Table 1: Details of model architecture for OD segmentation. Concatenation is described with brackets. Conv 6 for Vessel network indicates atrous pyramid pooling with dilation rate of 1,2,4 with output filter number of 170 for each. *up* means up-scaling of feature maps by scale of 2.

Block	**Main**		**Vessel**	
	Operation	Output size	Operation	Output size
Input	fundus	(640,640,3)	vessel	(640,640,1)
conv 1	$\left\{\begin{matrix}3\times3\text{ conv}\\ \text{ReLU}\end{matrix}\right\}\times 2$	(640, 640, 32)	$\left\{\begin{matrix}3\times3\text{ conv}\\ \text{ReLU}\end{matrix}\right\}\times 2$	(640, 640, 16)
pool 1	2×2 maxpool	(320, 320, 32)	2×2 maxpool	(320, 320, 16)
conv 2	$\left\{\begin{matrix}3\times3\text{ conv}\\ \text{ReLU}\end{matrix}\right\}\times 2$	(320, 320, 64)	$\left\{\begin{matrix}3\times3\text{ conv}\\ \text{ReLU}\end{matrix}\right\}\times 2$	(320, 320, 32)
pool 2	2×2 maxpool	(160, 160, 64)	2×2 maxpool	(160, 160, 32)
conv 3	$\left\{\begin{matrix}3\times3\text{ conv}\\ \text{ReLU}\end{matrix}\right\}\times 2$	(160, 160, 128)	$\left\{\begin{matrix}3\times3\text{ conv}\\ \text{ReLU}\end{matrix}\right\}\times 2$	(160, 160, 64)
pool 3	2×2 maxpool	(80, 80, 128)	2×2 maxpool	(80, 80, 64)
conv 4	$\left\{\begin{matrix}3\times3\text{ conv}\\ \text{ReLU}\end{matrix}\right\}\times 2$	(80, 80, 256)	$\left\{\begin{matrix}3\times3\text{ conv}\\ \text{ReLU}\end{matrix}\right\}\times 2$	(80, 80, 128)
pool 4	2×2 maxpool	(40, 40, 256)	2×2 maxpool	(40, 40, 128)
conv 5	$\left\{\begin{matrix}3\times3\text{ conv}\\ \text{ReLU}\end{matrix}\right\}\times 2$	(40, 40, 512)	$\left\{\begin{matrix}3\times3\text{ conv}\\ \text{ReLU}\end{matrix}\right\}\times 2$	(40, 40, 256)
pool 5	2×2 maxpool	(20, 20, 512)	2×2 maxpool	(20, 20, 256)
conv 6	$\left\{\begin{matrix}3\times3\text{ conv}\\ \text{ReLU}\end{matrix}\right\}\times 2$	(20, 20, 512)	$\left[\begin{matrix}3\times3\text{ conv (1)}\\ 3\times3\text{ conv (2)}\\ 3\times3\text{ conv (4)}\end{matrix}\right]$	(20, 20, 510)
concat 1	**[conv6 (Main), conv6 (Vessel)]**			
concat 2	[up(concat 1), conv5]			
conv 7	$\left\{\begin{matrix}3\times3\text{ conv}\\ \text{ReLU}\end{matrix}\right\}\times 2$	(40, 40, 512)		
concat 3	[up(conv 7), conv4]			
conv 8	$\left\{\begin{matrix}3\times3\text{ conv}\\ \text{ReLU}\end{matrix}\right\}\times 2$	(80,80,256)		
concat 4	[up(conv 8), conv3]			
conv 9	$\left\{\begin{matrix}3\times3\text{ conv}\\ \text{ReLU}\end{matrix}\right\}\times 2$	(160, 160, 128)		
concat 5	[up(conv 9), conv2]			
conv 10	$\left\{\begin{matrix}3\times3\text{ conv}\\ \text{ReLU}\end{matrix}\right\}\times 2$	(320, 320, 64)		
concat 6	[up(conv 10), conv1]			
conv 11	$\left\{\begin{matrix}3\times3\text{ conv}\\ \text{ReLU}\end{matrix}\right\}\times 2$	(640, 640, 32)		
output	$\left\{\begin{matrix}1\times1\text{ conv}\\ \text{sigmoid}\end{matrix}\right\}\times 1$	(640, 640, 1)	$\left\{\begin{matrix}1\times1\text{ conv}\\ \text{sigmoid}\end{matrix}\right\}\times 1$	(20, 20, 1)

Table 2: Details of model architecture for cup segmentation. Concatenation is described with brackets. Conv 6 for Vessel network indicates atrous pyramid pooling with dilation rate of 1,2,4,8 with output filter number of 192 for each. *up* means up-scaling of feature maps by scale of 2.

Block	**Main**		**Vessel**	
	Operation	Output size	Operation	Output size
Input	fundus	(640,640,3)	vessel	(640,640,1)
conv 1	$\left\{\begin{matrix}3\times3\text{ conv}\\ \text{ReLU}\end{matrix}\right\}\times 2$	(640, 640, 32)	$\left\{\begin{matrix}3\times3\text{ conv}\\ \text{ReLU}\end{matrix}\right\}\times 2$	(640, 640, 16)
pool 1	2×2 maxpool	(320, 320, 32)	2×2 maxpool	(320, 320, 16)
conv 2	$\left\{\begin{matrix}3\times3\text{ conv}\\ \text{ReLU}\end{matrix}\right\}\times 2$	(320, 320, 64)	$\left\{\begin{matrix}3\times3\text{ conv}\\ \text{ReLU}\end{matrix}\right\}\times 2$	(320, 320, 32)
pool 2	2×2 maxpool	(160, 160, 64)	2×2 maxpool	(160, 160, 32)
conv 3	$\left\{\begin{matrix}3\times3\text{ conv}\\ \text{ReLU}\end{matrix}\right\}\times 2$	(160, 160, 128)	$\left\{\begin{matrix}3\times3\text{ conv}\\ \text{ReLU}\end{matrix}\right\}\times 2$	(160, 160, 64)
pool 3	2×2 maxpool	(80, 80, 128)	2×2 maxpool	(80, 80, 64)
conv 4	$\left\{\begin{matrix}3\times3\text{ conv}\\ \text{ReLU}\end{matrix}\right\}\times 2$	(80, 80, 256)	$\left\{\begin{matrix}3\times3\text{ conv}\\ \text{ReLU}\end{matrix}\right\}\times 2$	(80, 80, 128)
pool 4	2×2 maxpool	(40, 40, 256)	2×2 maxpool	(40, 40, 128)
conv 5	$\left\{\begin{matrix}3\times3\text{ conv}\\ \text{ReLU}\end{matrix}\right\}\times 2$	(40, 40, 512)	$\left[\begin{matrix}3\times3\text{ conv (1)}\\ 3\times3\text{ conv (2)}\\ 3\times3\text{ conv (4)}\\ 3\times3\text{ conv (8)}\end{matrix}\right]$	(40, 40, 768)
concat 1	**[conv5 (Main), conv5 (Vessel)]**			
concat 2	[up(concat 1), conv4]			
conv 6	$\left\{\begin{matrix}3\times3\text{ conv}\\ \text{ReLU}\end{matrix}\right\}\times 2$	(80,80,256)		
concat 4	[up(conv 7), conv3]			
conv 8	$\left\{\begin{matrix}3\times3\text{ conv}\\ \text{ReLU}\end{matrix}\right\}\times 2$	(160, 160, 128)		
concat 5	[up(conv 8), conv2]			
conv 9	$\left\{\begin{matrix}3\times3\text{ conv}\\ \text{ReLU}\end{matrix}\right\}\times 2$	(320, 320, 64)		
concat 6	[up(conv 9), conv1]			
conv 10	$\left\{\begin{matrix}3\times3\text{ conv}\\ \text{ReLU}\end{matrix}\right\}\times 2$	(640, 640, 32)		
output	$\left\{\begin{matrix}1\times1\text{ conv}\\ \text{sigmoid}\end{matrix}\right\}\times 1$	(640, 640, 1)	$\left\{\begin{matrix}1\times1\text{ conv}\\ \text{sigmoid}\end{matrix}\right\}\times 1$	(20, 20, 1)

References

1. 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. vol. 10579, p. 105790B. International Society for Optics and Photonics (2018)
2. Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L.: Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. arXiv preprint arXiv:1606.00915 (2016)
3. Decencire, E., Zhang, X., Cazuguel, G., Lay, B., Cochener, B., Trone, C., Gain, P., Ordonez, R., Massin, P., Erginay, A., Charton, B., Klein, J.C.: Feedback on a publicly distributed database: the messidor database. Image Analysis & Stereology 33(3), 231–234 (Aug 2014), http://www.ias-iss.org/ojs/IAS/article/view/1155
4. Glorot, X., Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks. In: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics. pp. 249–256 (2010)
5. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
6. 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. pp. 234–241. Springer (2015)
7. Sahasrabuddhe, P.P.S.P.R.K.M.K.G.D.V., Meriaudeau, F.: Indian diabetic retinopathy image dataset (idrid) (2018), http://dx.doi.org/10.21227/H25W98
8. Son, J., Park, S.J., Jung, K.H.: Retinal vessel segmentation in fundoscopic images with generative adversarial networks. arXiv preprint arXiv:1706.09318 (2017)