Segmentation and Detection Glaucoma Using Deep Convolutional Neural Network *

Tianjiao Guo[1,2,3], Yun Gu[1,2], Kun Fang[2], Qi Yu[4], and Jie Yang[1,2,3]

[1] Institute of Medical Robotics, Shanghai Jiao Tong University
[2] Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University
[3] School of Biomedical Engineering, Shanghai Jiao Tong University
[4] Shanghai Jiaotong University Eye Institute Reading Center

Abstract. The detection of optic disc (OD), optic cup (OC) and fovea is crucial to automatic diagnosis based on fundus images. In this paper, we proposed an framework that detects the OD, OC and fovea based on deep convolutional neural networks. The original image is firstly preprocessed which is followed by the generation of pseudo labels; These pseudo labels and coordinate labels are then fed into a fully convolutional neural network with residual modules and a convolutional neural network respectively for localization of OD and fovea. The ROI patch including OD and OC region is then generated to the segmentation of OC, polar transformation is then introduced to ROI patch for the segmentation of OD. Glaucoma screening then is achieved by the idea if transfer learning. Experimental results show that our method performs well on online validation set.

Keywords: Optic disc (OD)· Optic cup (OC) · Segmentation · fovea · Glaucoma diagnosis.

1 Introduction

Computer-aided diagnosis of ophthalmic disease is a cheap, efficient and convenient way to early screening of several ophthalmic disease, which can avoid vision loss. The diagnosis of glaucoma is usually based on the shape of optic disc (OD) and optic cup (OC). Other disease diagnosis as diabetic retinopathy can be based on the region around fovea. As a result the detection of OD, OC and fovea is crucial. The paper contains all 3 tasks in REFUGE2 challenge [1]: 1)localization of fovea; 2)segmentation of OD and OC; 3)the screening of glaucoma.

In color fundus image, the optic disc (OD) is a yellowish elliptical region includes a cup-like region named optic cup (OC) and a peripheral region called

* This research was partly supported by National Key R&D. Program of China (No. 2019YFB1311503), Committee of Science and Technology, Shanghai, China (No. 19510711200), Shanghai Sailing Program (No. 20YF1420800), National Natural Science Foundation of China (No. 61661010, 61977046).

neuroretinal rim. [3] In this work, an algorithm based on the combination neural networks is proposed to the detection of OD, OC and fovea, as well as screening glaucoma. The proposed method is composed with three stages including Localization stage, Segmentation stage and Classification stage. For Localization stage, we generate the pseudo labels of fovea and optic disc regions. These pseudo labels as well as the real coordinates are used as supervision to train the multi-task neural networks for object localization. For Segmentation stage, based on the localization result, the region-of-interests (ROI) with OD and OC region is cropped and processed via polar transformation. The original coordinate system and polar transformed ROI patches are fed into convolutional neural networks for the segmentation of OD and OC. For Classification stage, we introduce transfer learning to improve the performance. These process stages are shown in Fig. 4 and Fig. 5

2 Method

2.1 Pre-processing and post-processing

Image Pre-processing The background brightness of the fundus image is estimated by Gaussian filter, which is then subtracted to balance the luminance and enhance the contrast of the whole image. This process can be formulated as follows Eq. 1.

$$I_C(P;\sigma) = \alpha[I(P) - G(P;\sigma) * I(P)] + \gamma \tag{1}$$

Where I is the color fundus image; I_C is the enhanced image; P represents the position of each pixel; $G(P;\sigma)$ is Gaussian filter; σ is the variance of Gaussian filter; * represents the convolution operation; α is the coefficient of contrast enhancement; γ is used to keep most gray-values of pixels in the range of [0, 1]. We empirically set $\alpha = 4$, $\gamma = 0.5$, $\sigma = \frac{r}{30}$, where r denotes the radius of retinal image Field of View (FoV).

Localization fake annotations generation For some dataset including OD segmentation annotation, we calculate the geometric center coordinate of OD annotation area and treat the coordinate as the center of OD P_{OD}. For the other dataset only including the coordinate P_{OD}, we just use the coordinate. All datasets we use for localization task include the coordinate of fovea P_F. The location of the fovea center is about 2.5ρ from the OD center, where ρ is the optic disc diameter [2]. We generate circular binary region locates at P_{OD} as OD region label binary image L_{OD} and circular binary region locates at P_F as fovea region label binary image. The radius of them is 0.5ρ.

Localization result choice The shape index of ouput is defined as follows:

$$SI = \frac{C^2}{S} \tag{2}$$

Where S and C denote the area and the perimeter of a region respectively in prediction binary image I_{pre}. The shape of output with very small or large SI is considered as a failure case. We set a threshold that $SI \in (11, 12.2)$. When $SI \in (11, 12.2)$, we use the center of FCN output region as the localization result and we use the CNN output when $SI \notin (11, 12.2)$.

Polar transformation for segmentation Suppose that an image patch I'_{Cpt} with a size of $2R_i + 1$, and the origin of coordinate is located at the central point. For the polar transformation in digital image, we have

$$I'_P(r, \theta_n) = I'_{Cpt}\left(r\cos\left(\frac{\theta_{max}\theta_n}{N_\theta}\right), r\sin\left(\frac{\theta_{max}\theta_n}{N_\theta}\right)\right) \tag{3}$$

Where I'_P denotes the image after polar transformation (PT). The r and θ_n denote the new coordinate axis index. N_θ denotes the number of sampling points and θ_{max} denotes the total sampling angle. Bilinear interpolation is introduced when $r\cos\left(\frac{\theta_{max}\theta_n}{N_\theta}\right)$ or $r\sin\left(\frac{\theta_{max}\theta_n}{N_\theta}\right)$ is not integer. The more N_θ is set, the less systematic error is generated, however the more computing resource is needed. In our experiments we set N_θ to be 1072 and θ_{max} to be 3π.

Network design We modify the U-shape convolutional network (U-Net) and add shortcut path in basic block. The architecture of our FCN and the basic block of U-Net and our networks are shown in Fig. 1 and Fig. 2 respectively. As for the CNN, we remove the layers behind the first upsampling layer, then connect 3 fully connected (FC) layers. The architecture of CNN last several layers is shown in Fig. 3. The output of CNN here are 4 values, which denote the OD and fovea central coordinates. Note that we set all the dropout rate to be 0.5 in our experiments.

Post process For the segmentation of OD and OC, we use a threshold 0.7 to generate binary segmentation map. Inverse polar transformation is used on the polar transformed segmentation map. Then we introduce the method mentioned in [3], preserve the largest connected area and introduce ellipse fitting to tune the segmentation output.

2.2 Localization stage

Original image is pre-processed by Eq. 1., after crop and resize the image pair to a fixed size, we concatenate the image pair together and feed it into FCN and CNN. We use a threshold 0.7 to decide FCN output region. We select the localization results of two networks according to SI as described above. Dice loss is used as the loss function of FCN and the mean square error loss is used in CNN. The whole localization stage procedure is shown in Fig. 4. The input image size is set to be 256 × 256.

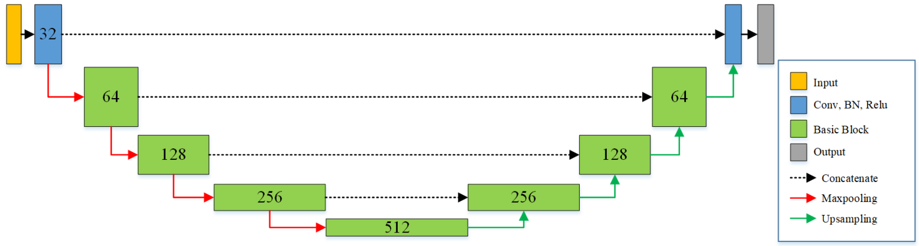

Fig. 1. The architecture of U-shaped FCN we used, the green block denotes the basic block. The output channels of each block is shown on the block.

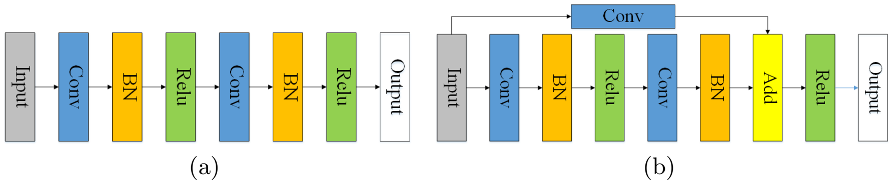

Fig. 2. The architecture of two basic blocks. (a) is the U-net basic block, (b) is our basic block, we add the shortcut path in the block. The convolution kernel of shortcut path is set to be 'size 1 pad 0' and other convolution kernels are set to be 'size 3 pad 1'.

2.3 Segmentation stage

After the localization stage, we extract a square ROI patch which locates at the localization stage output. The length of ROI patch is 2ρ where 2.5ρ is the distance between OD and fovea. Note that the ROI patch here is also the concatenation of original RGB-space image and pre-processed image pair. We introduce polar transformation to transform ROI patch. The Cartesian coordinate as well as polar transformed ROI patches are fed into two FCNs which have the same structure. We introduce CE-Net [4] to the FCN here, only the input channels of first convolutional layer is changed to 6 in order to fit the input pair. The input Cartesian coordinate ROI patch size here is set to be 448 × 448 and polar transformed ROI patch size is set to be 224 × 896.

We expect the output of FCN including 3 maps p_0, p_1 and p_2, p_0 is background map, p_1 is the map including OD mask excluding OC mask, i.e. the neuroretinal rim region, and p_2 includes only OC mask. Some databases we used include the fine annotations of both OD and OC while others only include OD annotations. Suppose that g_0, g_1 and g_2 denote the groundtruth map corresponding to p_0, p_1 and p_2, we set $g_0' = g_0$, $g_1' = g_1 + g_2$, then g_1' will be the OD (including OC area) groundtruth map. Similarly, we also set $p_0' = p_0$, $p_1' = p_1 + p_2$ to be the OD prediction map. We define two loss functions $Loss_{OC}$ and $Loss_{OD}$, $Loss_{OC}$ is defined as follows Eq. 4.

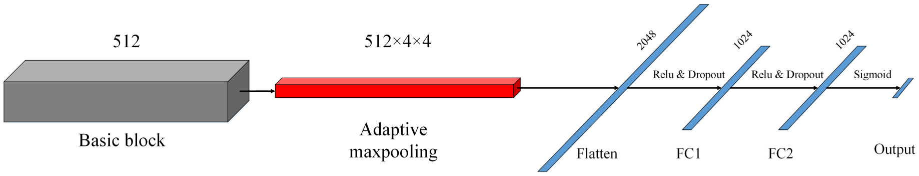

Fig. 3. The last several layers of CNN. The first several layers are the same with the FCN layers before the first upsampling layer. The FC layer size is written on corresponding layer in figure. Note that this structure is also used in Classification stage, the difference is only the output size.

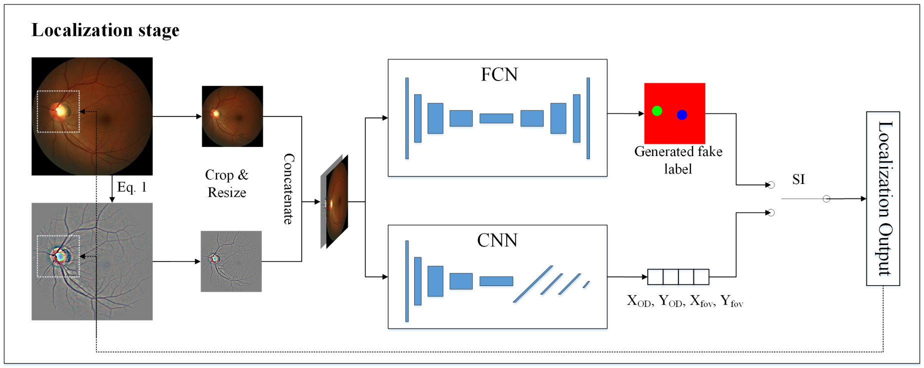

Fig. 4. Whole localization stage procedure. The original and pre-proceed images pair is cropped and resized to a small size, then they are concatenated together and fed into two networks. The FCN predicts 2 circular areas and the CNN predict the coordinate of OD and fovea directly. The localization result is then chosen by SI

$$Loss_{OC} = (C - \sum_{c=0}^{C-1} \frac{\sum p_c g_c}{\sum p_c g_c + \alpha \sum (1-p_c) g_c + \beta \sum p_c (1-g_c)}) \tag{4}$$

Where C denotes total number of classes and equals to 3 here, α and β are the trade-offs of penalties and all set to be 0.5 here. Similarly, $Loss_{OD}$ is defined as follows Eq. 5.

$$Loss_{OD} = (C' - \sum_{c=0}^{C'-1} \frac{\sum p'_c g'_c}{\sum p'_c g'_c + \alpha \sum (1-p'_c) g'_c + \beta \sum p'_c (1-g'_c)}) \tag{5}$$

Where C' equals to 2 here, p'_c and g'_c are defined as above. The total loss $Loss_T$ is the combination of $Loss_{OD}$ and $Loss_{OC}$ as follows Eq. 6.

$$Loss_T = Loss_{OD} + f Loss_{OC} \tag{6}$$

Where $f = \{0, 1\}$ is a binary parameter to decide whether loss is exist. $f = 1$ if and only if OC annotation exists. Cartesian coordinate transformation is used to transform the polar segmentation map, then we preserve the largest connected area and introduce ellipse fitting [3] to tune the segmentation output.

2.4 Classification stage

We introduce the thought of transfer learning in classification task. We change the structure of the FCN to classification CNN in Segmentation stage, remove the Decoder layers, then connect 3 fully connected (FC) layers, and copy the weights of the layers in FCN. ResNet50 is also introduced to train Cartesian coordinate patches. The classification layer of ResNet50 is modified to 1 output dimension. The final result of classification is the ensemble of two network outputs. The input image patch size are 448 and 224 in our classification CNN and ResNet50 respectively. Binary cross entropy loss is used here. The optimizer we used is Adam (learning rate = 1e-4), training epoch is set to be 40, batch size is set to be 24. The Segmentation and Classification stage procedures are shown in Fig. 5.

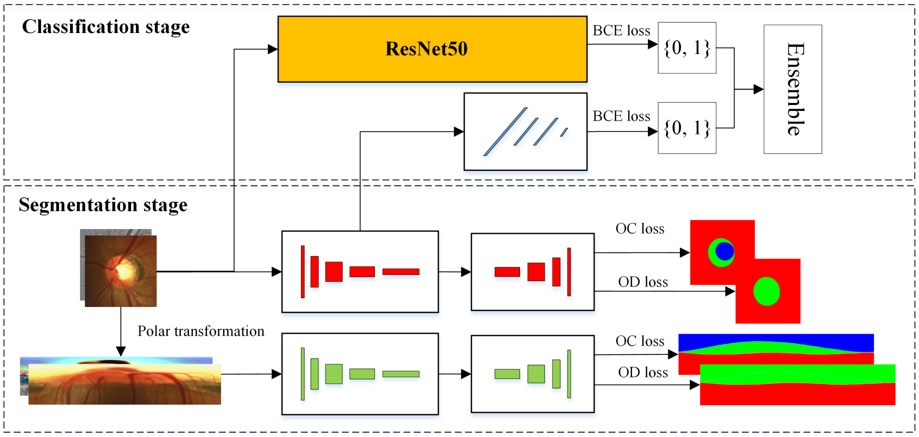

Fig. 5. Segmentation and Classification stages procedure. The original and pre-proceed OD patch pair is proceed by PT. Both Cartesian coordinate patch and transformed patch are used to segmentation training. Two loss functions are defined by Eq. 4 and Eq. 5. The first several layers of the FCN which is trained by Cartesian coordinate patches are copied and 3 FC layers are connected in Classification stage. ResNet50 is also introduced to train Cartesian coordinate patches. The final classification result is the ensemble of our FCN and ResNet50

3 Experiments

3.1 Materials

The database we used includes AMD, DRIONS, IDRID, Messidor, PALM, BinRushed, Drishti-GS, Magrabia, EyePACS, REFUGE and RIM_ONE, which are all public available. The details about these database and usage are listed on Table 1. 'Y' and 'N' denote the annotation of corresponding item IS or IS NOT public available respectively, 'M' denotes the annotation of corresponding item is not public available while we annotate them manually. The last column shows the corresponding stage that the database is used in, 1, 2 and 3 denote Localization, Segmentation and Classification stage respectively.

Table 1. Databases used in this work.

Dataset	OD loc	Fovea loc	OD mask	OC mask	Glaucoma label	Usage in
AMD	N	Y	Y	N	N	1,2
DRIONS	N	N	Y	N	N	2
IDRID(c)	Y	Y	N	N	N	1
IDRID(a)	N	N	Y	N	N	2
Messidor	N	M	Y	Y/N	N	1,2
PALM	N	Y	Y	N	N	1,2
BinRushed	N	N	Y	Y	N	2
Drishti-GS1	Y	N	Y	Y	N	2
Magrabia	N	N	Y	Y	N	2
EyePACS(part)	M	M	N	N	M	1,3
REFUGE	N	Y	Y	Y	Y	1,2,3
RIM_ONE	N	N	Y	Y	Y	2,3

3.2 Data augmentation strategy

Affine Transformation is introduced for data augmentation. Transformation includes random shuffle, shift, rotation, resizing, horizontal flip and shearing. The resizing rate is set to be [0.8, 1.2] and the shift range is [-0.1W, 0.1W], where W is the width of image. The rotation range is [0, 2π) in localization and segmentation stages, and {-$\pi/2$, $\pi/2$, π} in classification stage. Shear is only introduced in segmentation stage with a rate of [-0.2, 0.2]. Note that we utilize affine transformation to augment the image and corresponding annotation before extracting ROI in segmentation stage. The ROI patch center, also used as the polar transformation center, is also shift [-0.1, 0.1] times of the length of the patch, in order to augment polar transformation data. We do not introduce any augmentation strategy during validating and testing.

3.3 Training Details

In the localization and classification stages, the training data is randomly shuffled and split to 4 subsets, 4-fold cross validation is introduced. In the segmentation stage, only the training data including both OD and OC annotation are randomly shuffled and split to 4 subsets, one of the subsets is used as validation set while the rest 3 subsets as well as the training data including only OD annotation are used as training set. Python and Matlab are used as coding language. The deep learning structure is PyTorch. The optimizer we used is Adam (learning rate = 3e-4, other parameters use default values), and the initializer is Random Uniform. The convolution kernel of shortcut path in each basic block is set to be 'size 1 pad 0' and other convolution kernels are set to be 'size 3 pad 1'. The momentum of BatchNormalization layer is 0.99. The Dropout rate in FC layer is 0.5. The batch size in first two stages is 6. Training epoch in first two stages is set to be 300 and the weights which show the lowest loss on the validation set are preserved respectively. The final result is the ensemble output of 4 cross validated models.

3.4 Results

We submit our results to the online evaluation system and the part of leaderboard is listed on Table 2. 'Pami-G' is our team name.

Table 2. Leaderboard of Validation dataset.

Rank	Team	AUC	ED	Cup Dice	Disc Dice	CDR RME
1	MAI	.9840	8.412	.8804	.9662	.0368
2	Pami-G	.9842	9.457	.8654	.9668	.0424
3	cheeron	.9795	10.086	.8743	.9652	.0384
4	PingAn Smart Health	.9853	8.631	.8680	.9615	.0420
5	VUNO EYE TEAM	.9830	8.727	.8697	.9659	.0397

References

1. https://refuge.grand-challenge.org/Home2020//.
2. Baidaa Al-Bander, Waleed Al-Nuaimy, Bryan M Williams, and Yalin Zheng. Multiscale sequential convolutional neural networks for simultaneous detection of fovea and optic disc. *Biomedical Signal Processing and Control*, 40:91–101, 2018.
3. Huazhu Fu, Jun Cheng, Yanwu Xu, Damon Wing Kee Wong, Jiang Liu, and Xiaochun Cao. Joint optic disc and cup segmentation based on multi-label deep network and polar transformation. *IEEE transactions on medical imaging*, 37(7):1597–1605, 2018.
4. Zaiwang Gu, Jun Cheng, Huazhu Fu, Kang Zhou, Huaying Hao, Yitian Zhao, Tianyang Zhang, Shenghua Gao, and Jiang Liu. Ce-net: Context encoder network for 2d medical image segmentation. *IEEE Transactions on Medical Imaging*, pages 1–1, 2019.