Computerized Medical Imaging and Graphics 74 (2019) 61–71

Contents lists available at ScienceDirect

Computerized Medical Imaging and Graphics

journal homepage: www.elsevier.com/locate/compmedimag

Robust optic disc and cup segmentation with deep learning for glaucoma detection

Shuang Yu*, Di Xiao, Shaun Frost, Yogesan Kanagasingam

Australian eHealth Research Center, CSIRO, 147 Underwood Ave, Perth, Australia

ARTICLE INFO

Article history:
Received 29 September 2018
Received in revised form 26 January 2019
Accepted 26 February 2019

Keywords:
Retinal imaging
Glaucoma
Cup disc segmentation
Deep learning

ABSTRACT

Glaucoma is rated as the leading cause of irreversible vision loss worldwide. Early detection of glaucoma is important for providing timely treatment and minimizing the vision loss. In this paper, we developed a robust segmentation method for optic disc and cup segmentation using a modified U-Net architecture, which combines the widely adopted pre-trained ResNet-34 model as encoding layers with classical U-Net decoding layers. The model was trained on the newly available RIGA dataset, and achieved *an average dice value of 97.31% for disc segmentation and 87.61% for cup segmentation*, comparable to that of the experts' performance for optic disc/cup segmentation and Cup-Disc-Ratio (CDR) calculation on a reserved RIGA dataset. When tested on DRISHTI-GS and RIM-ONE dataset without re-training or fine-tuning, the model achieved comparable performance to that of the state-of-the-art in literature. We have also fine-tuned the model on two databases, which achieves an average disc dice value of 97.38% and cup dice value of 88.77% for DRISHTI-GS test set, disc dice of 96.10% and cup dice of 84.45% for RIM-ONE database, which is the state-of-the-art performance on both databases in terms of cup dice and disc dice value. The advantage of the proposed method is the combination of the pre-trained ResNet and U-Net, which avoids training the network from scratch, thereby enabling fast network training with less epochs, thus further avoids over-fitting and achieves robust performance.

1. Introduction

Glaucoma is a chronic disease that gradually damages the optic nerve by degeneration of nerve fibers (Mary et al., 2016). It is rated as the second leading cause for blindness worldwide, as well as the leading cause of irreversible visual loss (Giangiacomo and Coleman, 2009). The number of people being affected by glaucoma is projected to reach 80 million by year 2020 (Giangiacomo and Coleman, 2009; Bourne, 2006). Glaucoma in the early stage is asymptomatic and it is reported that approximately 50% of patients are unaware of the disease (Giangiacomo and Coleman, 2009; Mitchell et al., 1996; Michelson et al., 2008). Progression of glaucoma without diagnosis and proper treatment gradually leads to irreversible vision loss (Almazroa et al., 2015). It is reported that early treatment of glaucoma can decrease the rate of blindness 20 years later by around 50% (Michelson et al., 2008). Therefore, early diagnosis and treatment of glaucoma is essential for the proper management of the disease.

The Cup-to-Disc ratio (CDR) of the optic nerve head, which measures the vertical diameter of the optic cup to that of the disc, is among one of the most commonly adopted screening tests for glaucoma (Kumar et al., 2007). Generally, a CDR value higher than 0.5 is considered as glaucoma suspects and higher CDR value indicates a higher probability of glaucoma (Garway-heath et al., 1998; Sivaswamy et al., 2015). Fig. 1 compares a healthy (Fig. 1(a)) and a glaucomatous optic nerve head (Fig. 1(b)). The segmentation of the optic cup is challenging due to the high density of blood vessels and also the very gradual color intensity changes between the cup and rim.

Since it is a time-consuming and highly subjective task to manually segment optic disc and cup, various algorithms have been proposed for the automatic segmentation. The automated methods can be broadly classified into (1) image processing and hand-crafted feature based methods and (2) deep learning based methods.

Representative research in the image processing based methods includes the thresholding based algorithms and active contour algorithms. Thresholding based methods take advantage of the color intensity difference between the optic disc and cup and generate a binary image from single channel, multi-channel or processed images (Thakur and Juneja, 2018; Joshi et al., 2010; Cheng et al., 2011; Noor et al., 2013; Issac et al., 2015). However, thresholding

* Corresponding author.
E-mail address: Shuang.Yu@csiro.au (S. Yu).

https://doi.org/10.1016/j.compmedimag.2019.02.005
0895-6111/

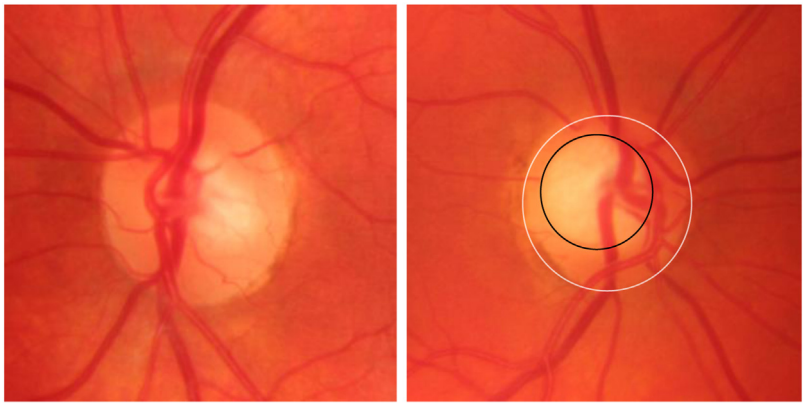

Fig. 1. Representation of healthy and glaucomatous optic nerve head. Circular region marked by black smaller circle is optic cup and white larger circle is optic disc. (A), healthy image; (B), glaucomatous optic disc image.

based methods are not robust enough for images with low contrast or presence of pathologies (Thakur and Juneja, 2018). Active contour based algorithms use a set of points to describe the boundary of optic disc and cup by minimizing the energy function (Mendels et al., 1999; Mary et al., 2015; Kumar et al., 2015; Joshi et al., 2011; Thakur and Juneja, 2018). However, due to the presence of noise and pathologies, the active contour method may get trapped in a local minimum and performance depends highly on the initialization of the contour model. Chen et al. proposed to over-segment the optic disc image into super-pixels and then use hand-crafted features to classify each super-pixel into optic disc or cup (Cheng et al., 2013a, 2013b). Wong et al. proposed a method to automatically detect optic cup boundary by detecting the kinks, i.e., bending of vessels when they cross the cup boundary (Damon et al., 2012). A summary of different algorithms can be approached at the survey papers by Mary et al. (2016) and Thakur and Juneja (2018).

In recently years, there is an increasing trend in adopting deep learning based methods for the optic disc and cup segmentation. Many state-of-the-art performance models are based on deep learning methods (Thakur and Juneja, 2018). Shankaranarayana et al. utilized fully convolutional networks with adversarial training to jointly segment the optic cup and disc (Shankaranarayana et al., 2017). The model achieved intersection over union (IOU) score of 0.961 for disc segmentation on RIM-ONE dataset. Zilly et al. proposed to use ensemble learning based convolutional neural network architecture to segment the optic disc and cup (Zilly et al., 2017). This method achieved state-of-the-art performance on DRISHTI-GS. Sevastopolsky used a modified U-Net for optic disc and cup segmentation (Sevastopolsky, 2017). The model first segment the optic disc, and extract a Region-of-Interest (ROI), based on which further segment the optic cup. Ful et al. proposed an architecture named M-Net combined with polar transformation to segment optic disc and cup simultaneously (Fu et al., 2018). The disc centered region is first passed through a polar transformation before training and prediction with M-Net. The method achieved a balanced accuracy of 0.983 for disc segmentation and 0.930 for cup segmentation on ORIGA dataset. Al-Bander et al. utilized DenseNet incorporated with a fully convolutional network to segment optic disc and cup (Al-Bander and Zheng, 2018). The model was trained on ORIGA dataset and tested on a number of publicly available databases.

However, most deep learning based models either have pre-requirement to prepare or extract the optic disc centered region with pre-processing procedures with assumptions that the algorithm can provide the ROI accurately (Fu et al., 2018; Shankaranarayana et al., 2017; Zilly et al., 2017; Al-Bander and Zheng, 2018), or detect the optic disc and cup separately (Sevastopolsky, 2017). In this paper, the automatic segmentation of optic disc and cup is implemented in a two-stage manner. In the first stage, a single label deep learning model is built and trained to segment the optic disc, which aims to obtain the disc center and diameter via the rough disc segmentation. Then the local region (randomized 1.5–2 diameter) surrounding the detected disc is extracted as Region-of-Interest (ROI) and fed to the second stage. In the second stage, a multi-label modified U-Net model is trained to obtain the refined segmentation of the optic disc and cup simultaneously. We modified the widely adopted U-Net architecture by using pre-trained ResNet34 model as encoder and validated the performance of the modified U-Net model on several publicly available glaucoma datasets. Compared to training the deep learning model from scratch, using pre-trained model weights enables a fast training stage, avoiding over-fitting and achieving robust model performance.

Contributions of this paper include: (1), development and validation of the modified U-Net architecture with pre-trained ResNet model as encoder layers; (2), realization of fully automated accurate segmentation of optic disc and cup for color fundus images, with segmentation performance comparable to that of experts for RIGA dataset; (3), achievement of robust performance when applying the model trained on RIGA to DRISHTI-GS and RIM-ONE V3 database, indicating the robustness and generalization capability of the model.

2. Method and methodology

2.1. Image database

In this research, we mainly adopted three publicly available glaucoma databases that provide the pixel-level segmentation of both optic disc and cup, including retinal fundus images for glaucoma analysis (RIGA) database (Almazroa et al., 2018), DRISHTI-GS1 database (Sivaswamy et al., 2015) and RIM-ONE database (Fumero et al., 2011).

The RIGA database was made publicly available very recently and it contains in total of 750 color fundus images derived from three resources: including 460 images from MESSIDOR database (Topcon TRC NW6 non-mydriatic retinograph, dimension 2304×1536 and 1440×960) (Decenciere et al., 2014), 195 images from Bin Rushed Ophthalmic center in Riyadh, Saudi Arabia (Canon CR2 non-mydriatic digital retinal camera, dimension 2376×1584) and 95 images obtained from Magrabi Eye center (Topcon TRC 50DX mydriatic retinal camera, dimension 2743×1936) (Almazroa et al., 2018). The optic disc and cup boundaries of each image were manually annotated by six experienced ophthalmologists.

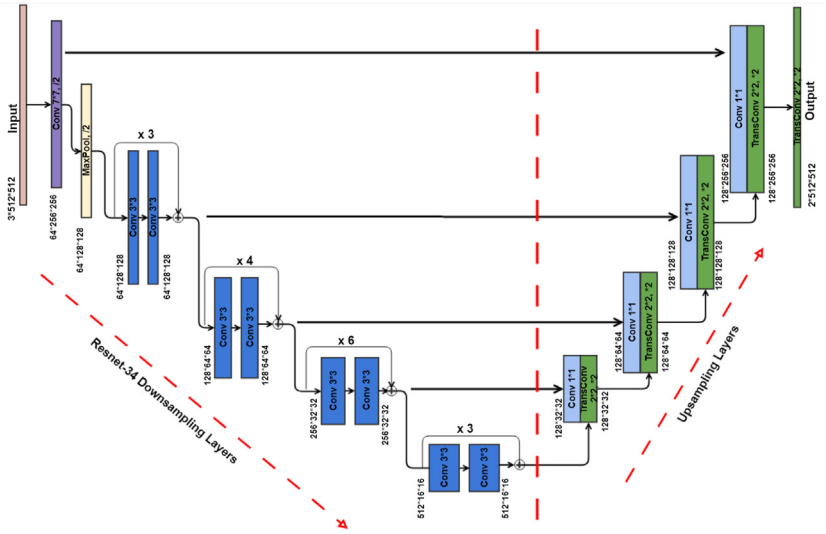

Fig. 2. Deep learning architecture for modified U-Net from Resnet-34.

The DRISHTI-GS1 database contains 101 color fundus images, including 50 training and 51 test images (Sivaswamy et al., 2015). The images have a Filed-of-View (FoV) of 30 degrees with a dimension of 2896 × 1944. Average boundaries of optic disc and cup based on 4 expert manual labels are provided by DRISHTI-GS1.

The RIM-ONE database contains in total of 169 fundus images, including 74 glaucoma or suspects images and 85 healthy images (Fumero et al., 2011). The images were captured using Nidek AFC-210 with a body of a Canon EOS 5D Mark II of 21.1 megapixels. Pixel-wise segmentation of optic disc and cup from two experts are provided by the RIM-ONE database.

2.2. Network architecture

U-Net is a U-shaped convolutional network which was originally developed for biomedical image segmentation (Ronneberger et al., 2015). The classical U-Net architecture is basically composed of down-sampling encoder layers and up-sampling decoder layers. The down-sampling (encoder) steps are consisted of repeated groups of two convolutional layers, followed by rectified linear unit (ReLU) and max pooling layer. Each up-sampling (decoder) layer concatenates the up-sampled (or transposed convolutional) feature maps from previous layer with corresponding feature maps from the down-sampling layers. The U-Net architecture has been widely adopted for segmentation tasks and proved to be capable of efficient learning from a relatively small training samples.

Various adaptations of U-Net architecture has been proposed recently. Fu et al. developed M-Net for the segmentation of optic disc and cup segmentation for color fundus images (Fu et al., 2018). The M-Net is different from U-Net in that M-Net contains multi-scale inputer layer and multiple side-output layer. TernausNet, which uses the pre-trained VGG11 model as encoder section of U-Net, was developed by Iglovikov and Shvets for the Kaggle Carvana challenge (Iglovikov and Shvets, 2018). Chaurasia and Culurciello developed LinkNet that exploits ResNet-18 as encoder. The decoder section of LinkNet is also different from classic U-Net architecture in that residual block is used, instead of concatenation (Chaurasia and Culurciello, 2017).

In our modified U-Net architecture, we employ the widely adopted ResNet-34 model as the down-sampling encoder section of the U-Net architecture, while in the up-sampling decoder section we follow the original U-Net structure, as illustrated in Fig. 2. The ResNet-34 down-sampling section contains a 7 × 7 convolutional layer with 64 filters followed by a max pooling layer, then repeated residual blocks that contains two 3 × 3 convolutional layers with batch normalization and identity shortcut connections (residual connection). To adapt the ResNet-34 model into U-Net architecture, each decoder block is composed by 2 × 2 transposed convolution of the previous layer output with stride of 2, concatenated with 1 × 1 convolution of the corresponding feature maps from the encoder section. The concatenated tensor is then batch normalized before passing to the next decoder block. The final layer of the architecture is transposed convolution with channel number the same as the target classes and output image size the same as the input image.

2.3. System working procedure

As illustrated in Fig. 3, our optic disc and cup segmentation procedure has two main steps. The images are firstly pre-processed with contrast enhancement, followed by resizing to the dimension of 512 × 512. Then, an optic disc segmentation model (single label modified U-Net model) is applied to roughly segment the optic disc.

The purpose of this Step 1 is to obtain the rough center and diameter of optic disc, which determines the Region-of-Interest (ROI) for the second step. We deliberately avoided to place the optic disc in the center of the ROI by randomly selecting four extreme points within the range of 1.5 to 2 times the detected disc diameter surrounding the optic disc. The randomly selected four point are used to determine the upper most, lower most, left most and right most of the extracted ROI.

The ROI is then resized to the dimension of 512 × 512 and passed to the Step 2 for refined optic disc and cup segmentation. The modified U-Net model used in the Step 2 is a multi-label model, which aims to segment the optic disc and cup simultaneously.

After the model prediction, the predicted optic disc and cup segmentations are further processed via a morphological post-processing module. The areas of all the individual segmented blobs are analyzed and only the blobs with the largest area are returned as the final segmentation of optic disc and cup. Cup Disc Ratio (CDR) is calculated using the vertical diameter of optic cup versus that of the optic disc.

2.4. Model training

Both the Step 1 and Step 2 models were trained with the MESSIDOR and BinRushed dataset from the RIGA database, in total of 655

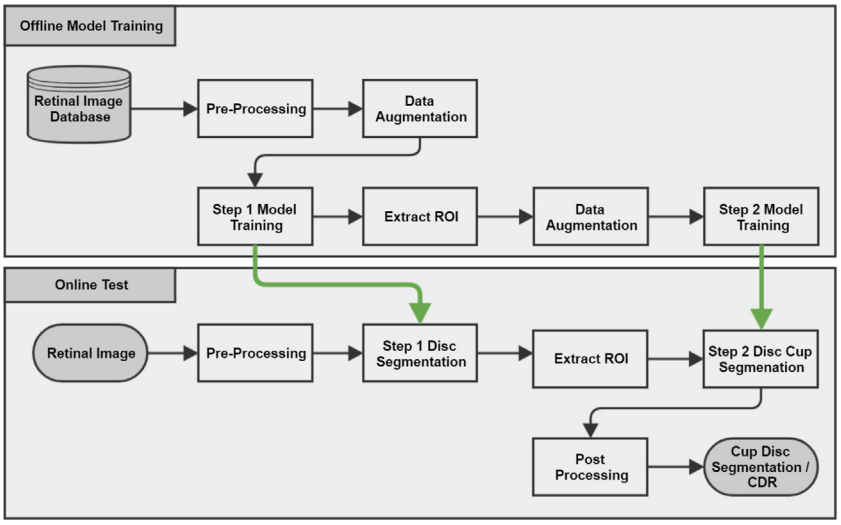

Fig. 3. System work flow.

images. Since there are in total of 6 expert label images provided for each image, we take the pixel label as positive only if it is labeled as positive from more than 3 of the experts, for both disc and cup label images. 15% of the training images were randomly selected and reserved for validation purpose. Images were resized to the dimension of 512×512 before training and testing. Data augmentation was applied on both original images and label images with tricks of rotation, upside down flip and left-right flip for the training dataset.

The segmentation task is considered as a pixel-level classification problem. Therefore, we use the binary cross entropy logistic loss as the loss function, as defined in Eq. (1), where N is the batch number, y_i is the ground truth label and $\hat{y}_i$ is the model prediction. Specially, for the Step 2 multi-label model, the relative weight between cup segmentation loss and disc segmentation loss was set as 2 : 1, since we are trying to stress the segmentation importance of optic cup.

$$\begin{aligned} \mathcal{L} &= -\frac{1}{N}\sum_{i=1}^{N}(y_i \cdot \log\sigma(\hat{y}_i) + (1-y_i)\cdot \log(1-\sigma(\hat{y}_i))) \\ \sigma(x) &= \frac{1}{1+e^{-x}} \end{aligned} \tag{1}$$

Adam optimization with an initial learning rate of 0.05 was used for model training. The learning rate decreases progressively following the optimization strategy proposed by Loshchilov and Hutter (2016). In order to take advantage of the pre-trained weight of ResNet-34, the model was trained in two stages. In the first stage, the weight of the ResNet-34 encoder section was frozen and only the weight of the up-sampling decoder part was optimized. In the second stage, the down-sampling layers were unfrozen with learning rates of $[e^{-4}, e^{-3}, e^{-2}]$ for near input layers, middle layers and near output layers. The purpose of the second stage training is to fine-tune the weight of all layers, with least tuning for ResNet layers and most tuning for decoding layers.

The model was trained on personal computer with Graphics Processing Unit (GPU) of NVIDIA Quadro K2100M, which has a compute capability rated as 3.0 with memory size of 2GB. Limited by the memory capacity of the GPU, the training batch size is set at 2. For both of the Step 1 and Step 2 models, the first stage training with frozen ResNet layers takes 10 epochs and the second stage training with unfrozen ResNet layers takes 20 epochs. It takes around 1 min to train the first stage training and 4 min for the second stage training. Therefore, it takes in total of around 90 min for the model training of Step 1 and Step 2.

3. Results

3.1. Evaluation criteria

In order to evaluate the model's segmentation performance of optic disc and cup, dice coefficient and Jaccard index (i.e., Intersection over Union) are used, which are calculated with Eqs. (2 and 3) respectively. In the equations, Y denotes the ground truth segmentation label and $\hat{Y}$ denotes the predicted segmentation image by the model.

$$D(Y, \hat{Y}) = \frac{2|Y \cap \hat{Y}|}{|Y| + |\hat{Y}|} \tag{2}$$

$$D(Y, \hat{Y}) = \frac{|Y \cap \hat{Y}|}{|Y \cup \hat{Y}|} = \frac{|Y \cap \hat{Y}|}{|Y| + |\hat{Y}| - |Y \cap \hat{Y}|} \tag{3}$$

Apart from this, we have also measured the accuracy of CDR, which is measured using Pearson correlation coefficient (PCC), mean and standard deviation of the absolute difference between the measured values and ground truth values.

3.2. RIGA database

In this research, the Step1 and Step2 models are trained on the MESSIDOR and BinRushed folders from the RIGA dataset, and tested on the Magrabia folder, which contains 95 images. The Step1 disc segmentation achieved an average Dice score of 95.54% and Jaccard score of 91.42%, which was further improved to 96.99% and 94.18% for Dice score and Jaccard score respectively in Step2. The cup segmentation in Step2 achieved an average Dice score of 89.13% and Jaccard score of 81.20%. Comparison between the model segmentation with that of the experts' labeling is shown in Fig 4 . Fig. 5 displays the representative segmentation results for the best and worst performance on the Magrabia Test Set, with cup dice of 0.97 and disc dice of 0.98 for the best performance, and cup dice of 0.87 and disc dice of 0.93 for the worst.

As listed in Table 1, the optic disc segmentation performance is worse than that from the experts 3, 4, 5 and 6, but better than that of experts 1 and 2. For cup segmentation, the model's segmentation performance is worse than that from the experts 1 and 2, but better than that from the experts 3, 4, 5 and 6. Note that all

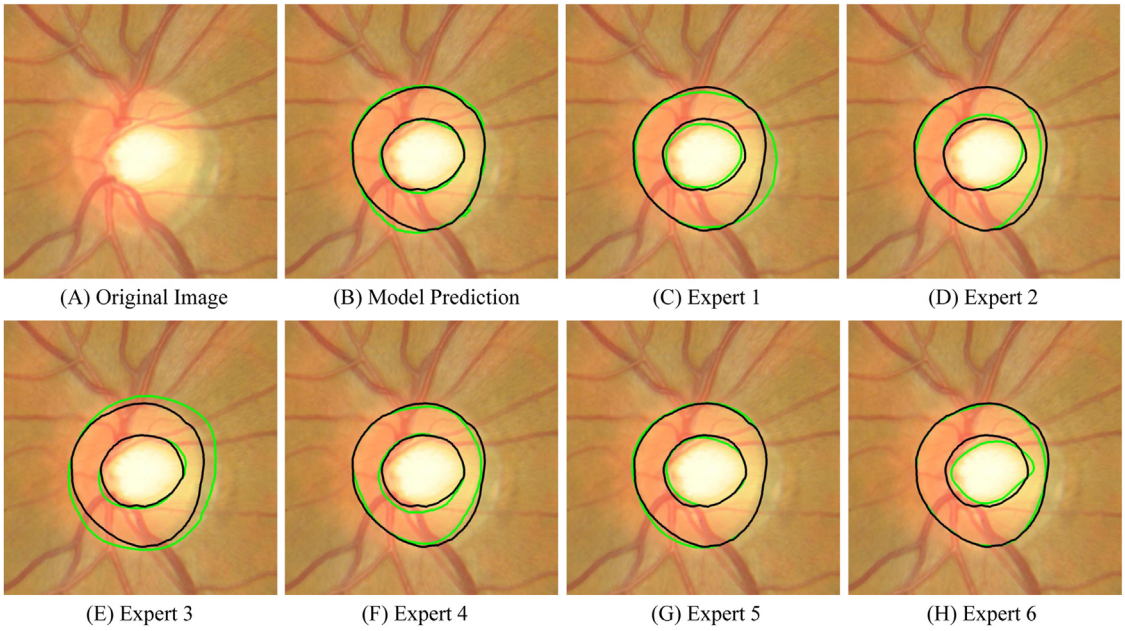

Fig. 4. Comparison of optic disc and cup segmentation boundary between model prediction versus experts. Ground truth label is marked in black color; model prediction and experts' labels are marked in green color. (A), original image; (B), model prediction; (C)–(H), label from expert 1 to expert 6. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

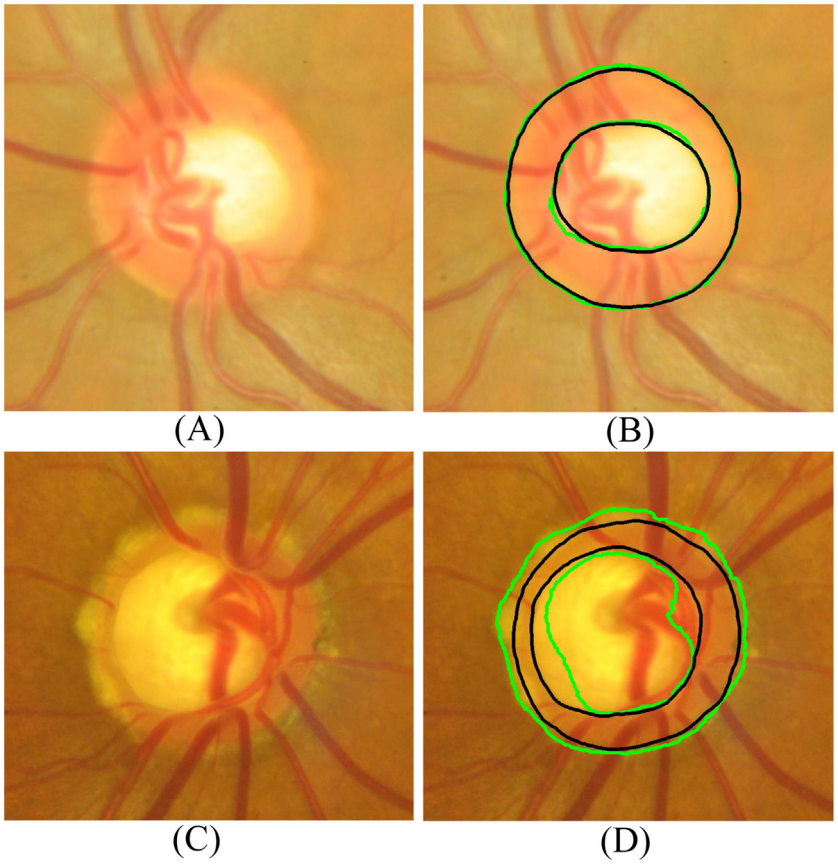

Fig. 5. Representative segmentation performance on the test set. The ground truth segmentation is marked in black color; model segmentation result is marked in green color. (A)(B), Original image and segmentation result for the best performance on test set, with cup dice of 0.97 and disc dice of 0.98; (C)(D), Original image and segmentation result for the worst performance on test set, with cup dice of 0.87 and disc dice of 0.93. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the segmentation performances are evaluated against the average labels from the experts, which are treated as golden-standard in this research.

Fig. 6 shows the performance of CDR measurement. The proposed method achieved a PCC value of 89.69%, mean of absolute error of 0.039 and standard deviation of 0.034. The model's perfor-

Table 1
Optic disc and cup segmentation performance for Magrabia dataset (%).

		Pred	Expt1	Expt2	Expt3	Expt4	Expt5	Expt6	Avg
Disc	Dice	96.99	96.85	96.31	97.28	97.96	97.74	97.72	97.31
	Jacc	94.18	93.93	92.95	94.76	96.01	95.62	95.56	94.80
Cup	Dice	89.13	91.47	90.78	88.42	85.47	87.53	81.97	87.61
	Jacc	81.20	85.15	83.94	80.4	75.91	79.33	71.66	79.40
CDR	PCC	89.69	88.72	88.75	93.38	87.51	94.99	81.90	89.21
	AbsMean	0.039	0.034	0.047	0.035	0.078	0.057	0.059	0.052
	AbsStd	0.034	0.040	0.049	0.036	0.054	0.046	0.065	0.048

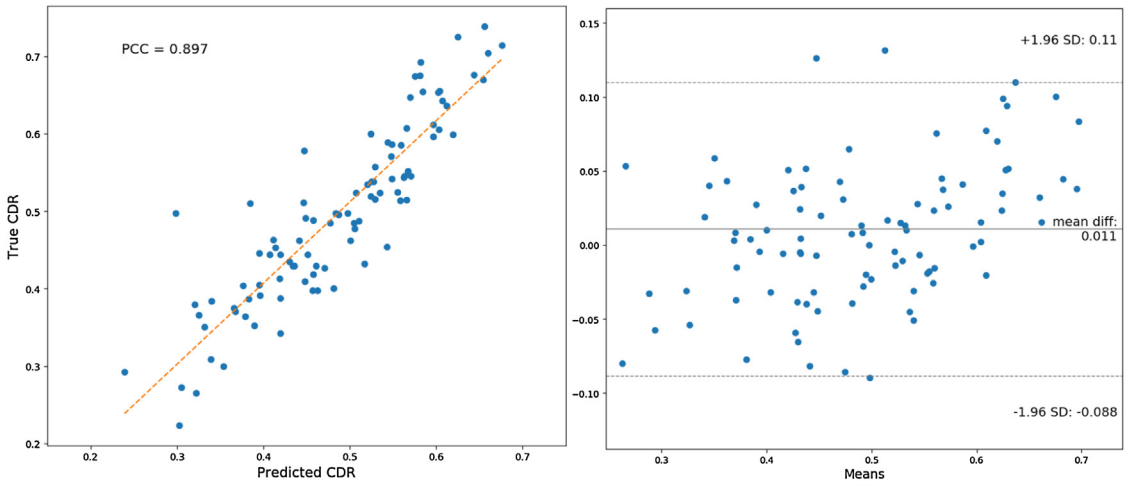

Fig. 6. CDR measurement performance for RIGA dataset (Magrabia folder). (A), Scatter plot of predicted CDR versus that of the ground truth CDR value; (B), Bland-Altman plot of the ground truth CDR value and predicted CDR.

mance is worse than that from the experts 3 and 5, but better than that from the experts 1 2, 4 and 6, in terms of PCC.

Fig. 7 summaries the mutual correlation heatmap among ground truth CDR value, model predicted CDR value and experts' labeled CDR values. The agreement of the model prediction with that of 6 experts falls within range of [79%, 90%], expert 1 with that of the other 5 experts [77%, 85%], expert 2 with the rest experts [75%, 85%], expert 3 with the rest experts [75%, 88%], expert 4 with the rest experts [77%, 82%], expert 5 with the rest experts [79%, 88%] and expert 6 with the rest experts [75%, 81%].

3.3. *DRISHTI-GS database*

Fig. 8 shows two representative segmentation results for the best and worst performance on the DRISHTI-GS database. The best performance achieves a cup dice value of 0.97 and disc dice value of 0.97, while the worst performance arrives at a 0.59 for the cup dice and 0.97 for disc dice.

Table 2 lists the optic disc and cup segmentation performance of the proposed method compared with the related literature on DRISHTI-GS database. When testing the model without fine-tuning, we achieved state-of-the-art performance in terms of Jaccard index for disc segmentation and Dice coefficient for cup segmentation. However, it is worth to point out that the result obtained by Zilly et al. (2017) and Sevastopolsky (2017) were trained or fine-tuned on the DRISHTI-GS database. Meanwhile, the result by our proposed method is originally trained on the RIGA database without fine-tuning. In order to compare the model performance under the same scale, we have fine-tuned the model parameters on the DRISHTI-GS training database, then test the performance on the test database. The fine-tuned model achieved an average disc dice value of 97.38%,

Table 2
Optic disc and cup segmentation performance for DRISHTI-GS Dataset (%).

	Disc		Cup	
	Dice	Jacc	Dice	Jacc
Sevastopolsky (2017)	–	–	85	75
Zilly et al. (2017)	97.3	91.4	87.1	**85.0**
Al-Bander and Zheng (2018)	94.9	90.42	82.82	71.13
Proposed Method †	96.44	93.19	87.39	78.08
Fine-Tuned*	**97.38**	**94.92**	**88.77**	80.42

Bold value indicates the state-of-the-art performance for DRISHTI-GS dataset.
† Performance is evaluated on the whole DRISHTI-DS dataset, including training and test set, in total of 100 images.
* Performance is evaluated on the DRISHTI-DS test set, which contains 50 images.

disc Jacc value of 94.92% and cup dice value of 88.77%, which is the state-of-the-art performance on the DRISHTI-GS test set.

For the calculation of CDR values, the proposed method achieved a PCC value of 84.38%, mean absolute difference of 0.08 and standard deviation of 0.053 against the ground truth labels, which is better than or equivalent to the performance obtained by Zilly et al., which has the mean difference of 0.08 and standard deviation of 0.063 (Zilly et al., 2017). The CDR measurement performance of the proposed method against the ground truth CDR value is displayed in Fig. 9.

3.4. *RIM-ONE database*

Fig. 10 shows the best and worst segmentation performance on the RIM-ONE database, with a cup dice value of 0.96 and disc dice value of 0.98 for the best performance, and cup dice of 0.77 and disc dice of 0.46 for the worst performance.

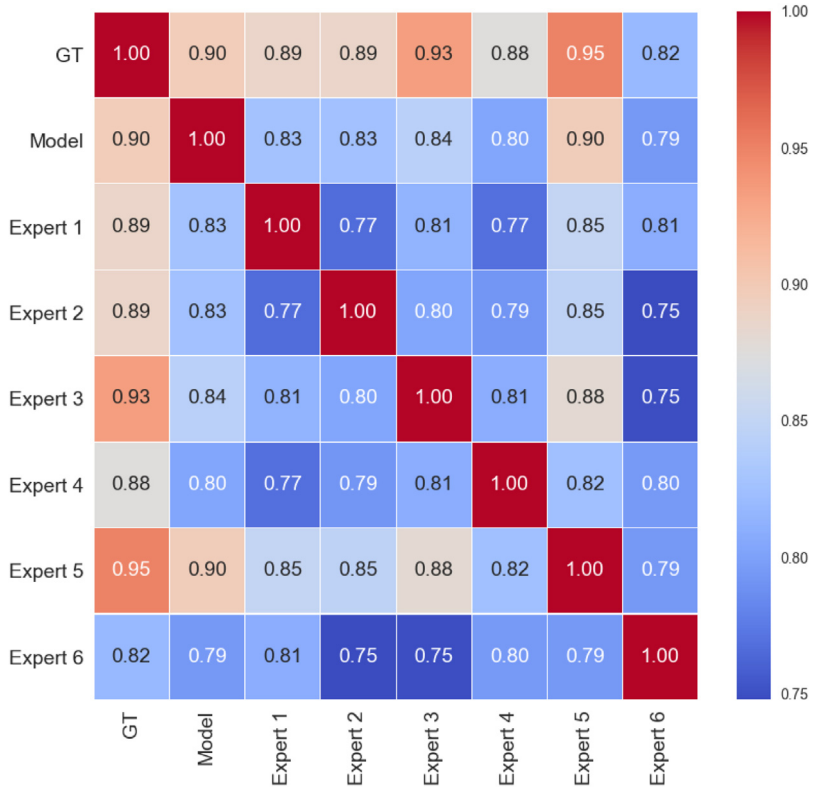

Fig. 7. CDR mutual correlation heatmap among ground truth value, model predicted value and experts value.

Table 3
Optic disc and cup segmentation performance for RIM-ONE dataset (%).

	Disc		Cup	
	Dice	Jacc	Dice	Jacc
Expert 1	97.51	95.20	89.49	81.78
Expert 2	97.60	95.34	89.83	82.27
Sevastopolsky (2017)	95	89	82	69
Zilly et al. (2017)	94.2	89.0	82.4	**80.2**
Al-Bander and Zheng (2018)	90.36	82.89	69.03	55.67
Proposed Method[†]	94.91	90.65	79.32	68.28
Fine-Tuned[*]	**96.10**	**92.56**	**84.45**	74.29

Bold valued indicates the state-of-the-art performance for RIM-ONE dataset.
[†] Performance is evaluated on the whole RIM-ONE dataset, in total of 129 images.
[*] Performance is evaluated on a randomly selected 20% test set with 32 images.

Table 3 lists the optic disc and cup segmentation performance for the proposed model without fine-tuning compared to the experts' labeling and that from the related literature. In order to compare with the method proposed by Zilly et al. (2017) and Sevastopolsky (2017) under the same standard, we have fine-tuned the model on randomly selected 80% images of the RIM-ONE database and then test the performance on the remaining 20%. The fine-tuned model achieves an average disc dice value of 96.10%, Jacc value of 92.56% and cup dice value of 84.45%, which is the state-of-the-art performance on RIM-ONE dataset.

Fig. 11 shows the CDR measurement result of the proposed method versus that of the ground truth CDR value. The proposed method achieved a PCC value of 80.18%, mean of absolute difference of 0.069 and standard deviation of 0.072 against the ground truth value.

3.5. Comparing different training strategies

In order to verify the benefit of model training using pretrained weight versus that of training from scratch, an experiment has been conducted to compare the performance of two training strategies along with the training procedure. The comparison experiment is performed for the disc segmentation model training, with a fixed learning rate of 0.001. As displayed in Fig. 12, training with pre-trained weight converges at value of 0.002 around 26 epochs in terms of validation loss, while training from scratch fluctuates even at the end of the training, around a value of 0.006. The validation dice value of pretrained strategy has basically maintained a better performance than the training from scratch ever since the 3rd epoch. In addition, the validation dice value stabilize around 0.97 for the pretrained weight strategy and 0.95 for the scratch training strategy.

4. Discussions

In this research, we have developed a variant of the classical U-Net architecture, combining the ResNet-34 as encoding down-sampling layers with customized decoding up-sampling layers. The benefits of using a pre-trained weight of ResNet as decoding layers are multi-fold. To begin with, the classical U-Net requires training from scratch, which generally requires more training epochs, thus longer training time. In comparison, the proposed architecture utilizes pre-trained weight, thus requires relatively less training epochs. For example, the original classical U-Net requires around 10 h to train on a Nvidia Titan GPU (Ronneberger et al., 2015). Sevastopolsky's modified U-Nett model was trained for 382 epochs for 2.76 h on Amazon Web Service platform for RIM-ONE dataset, including 159 images and resized to 256×256 (Sevastopolsky, 2017). Our proposed U-Net architecture can be well trained for 30 epochs, with total training time around 2 h, for 655 images resized to 512×512 on a Nvidia Quodro K2100M (compute capability 3.0). The reduced computation of combining pre-trained model with U-Net will generally mitigate the requirement on GPU hardware with high compute capability, as required by most deep learning architectures when training from scratch. Furthermore, training with

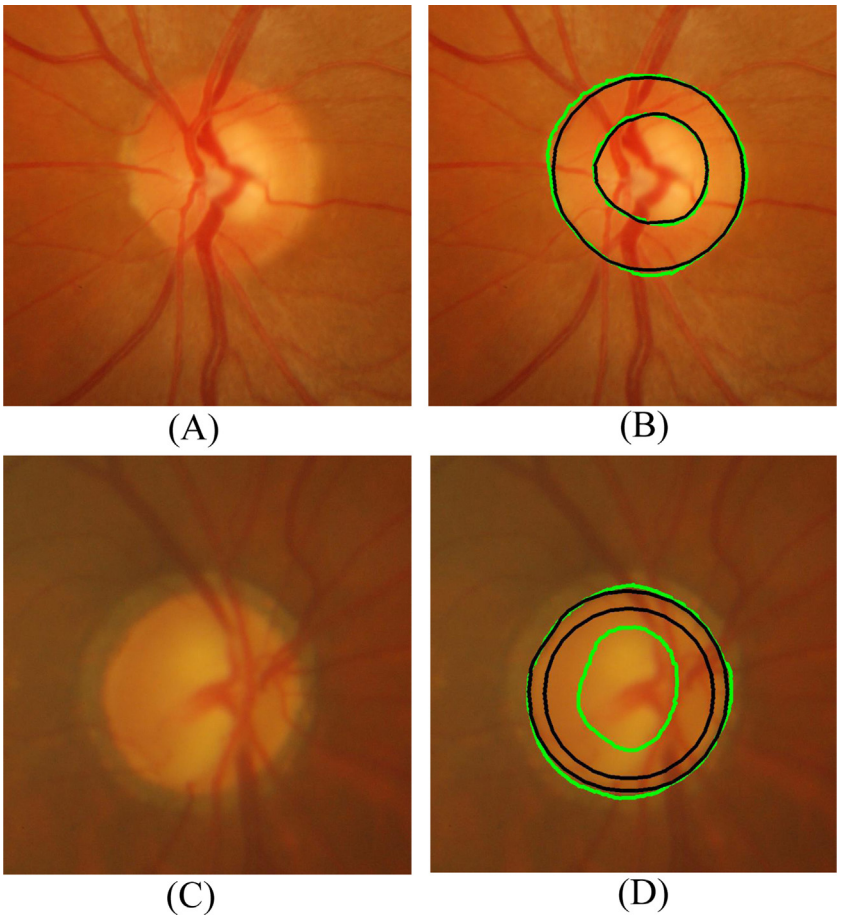

Fig. 8. Representative segmentation performance on the DRISHTI-GS database. The ground truth segmentation is marked in black color; model segmentation result is marked in green color. (A)(B), Original image and segmentation result for the best performance on DRISHTI-GS database, with cup dice of 0.97 and disc dice of 0.97; (C)(D), Original image and segmentation result for the worst performance on DRISHTI-GS database, with cup dice of 0.59 and disc dice of 0.97. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

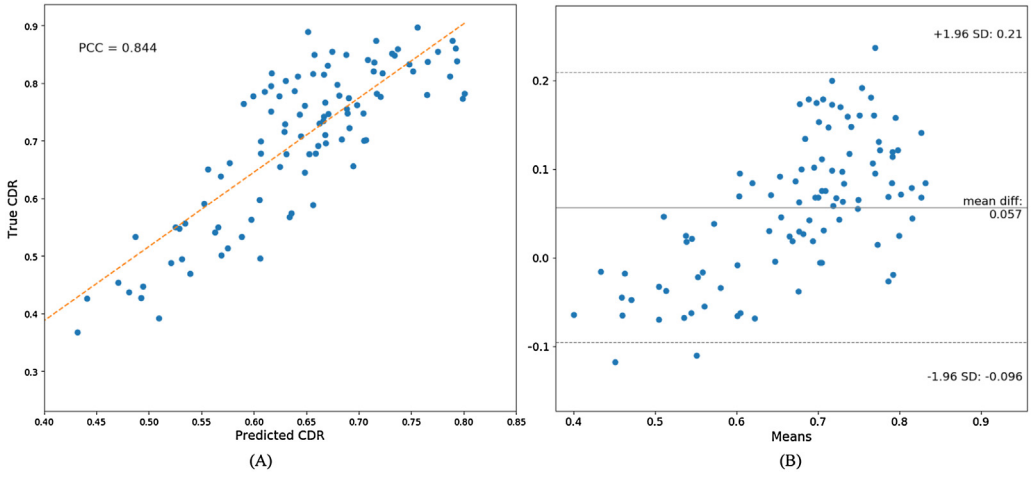

Fig. 9. CDR measurement performance for DRISHTI-GS dataset. (A), Scatter plot of predicted CDR versus that of the ground truth CDR value; (B), Bland-Altman plot of the ground truth CDR value and predicted CDR.

pretrained weight helps to improve the final model performance when the model converges, arriving at a comparatively lower loss and higher validation dice value than training from scratch.

We used the recently available RIGA database for the training of our model. The RIGA database provides labels of optic disc and cup boundary independently graded by 6 experts. We take the majority vote (equal or more than 3 experts) as ground truth labels for each pixels in the image. The experts' label varied in a considerable level from one to another, as can be observed from Table 1 and Fig. 7, which evaluate the experts' performance against ground truth labels and their mutual agreement between each other. The performance of the proposed model is comparable to that of the

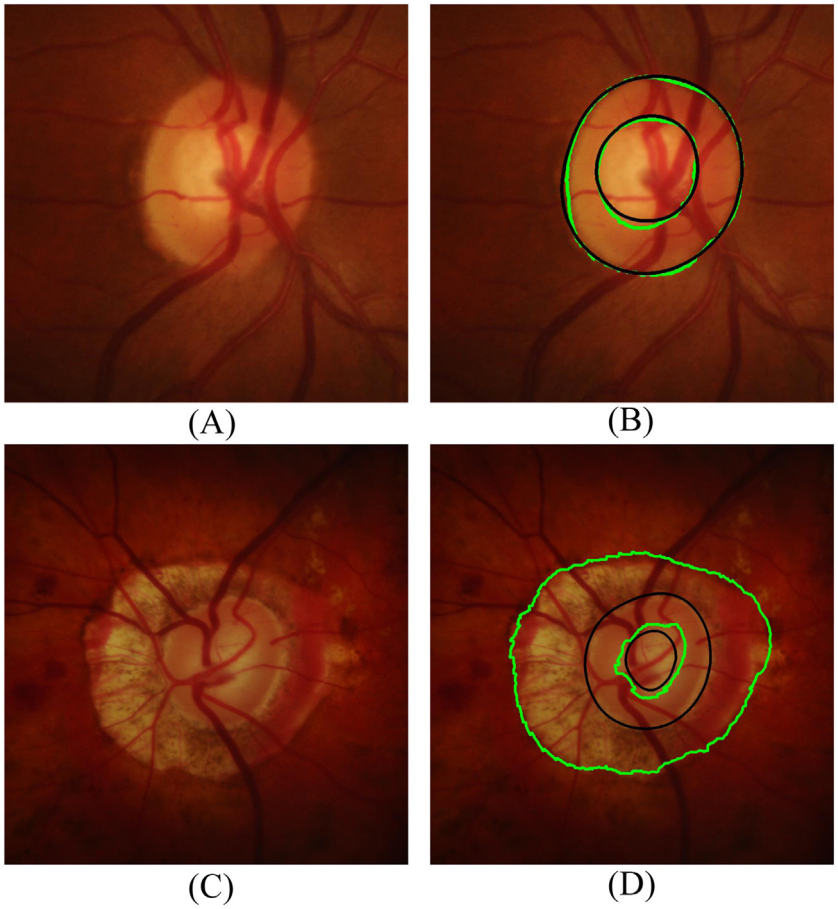

Fig. 10. Representative segmentation performance on the RIM-ONE database. The ground truth segmentation is marked in black color; model segmentation result is marked in green color. (A)(B), Original image and segmentation result for the best performance on RIM-ONE database, with cup dice of 0.96 and disc dice of 0.98; (C)(D), Original image and segmentation result for the worst performance on RIM-ONE database, with cup dice of 0.77 and disc dice of 0.46. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

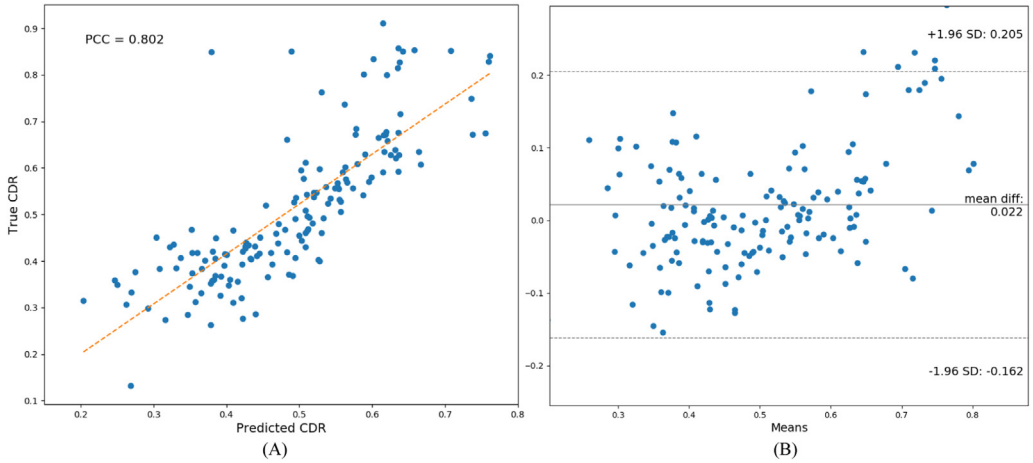

Fig. 11. CDR measurement performance for RIM-ONE Dataset. (A), Scatter plot of predicted CDR versus that of the ground truth CDR value; (B), Bland-Altman plot of the ground truth CDR value and predicted CDR.

experts, with the disc segmentation slightly worse than the average expert performance, but cup segmentation and CDR measurement better than that of the average expert.

To evaluate the accuracy and robustness of the model on unseen data, we tested the model on DRISHTI-GS and RIM-ONE dataset without re-training or fine-tuning. For the optic disc and cup segmentation performance listed in Tables 2 and 3, our model has achieved state-of-the-art performance for disc segmentation on DRISHTI-GS and RIM databases, when evaluated with the Jaccard score. It also achieved state-of-the-art performance for the cup segmentation on DRISHTI-GS database, when evaluated with dice score. It is worse than Sevastopolsky (2017) and Zilly et al. (2017)

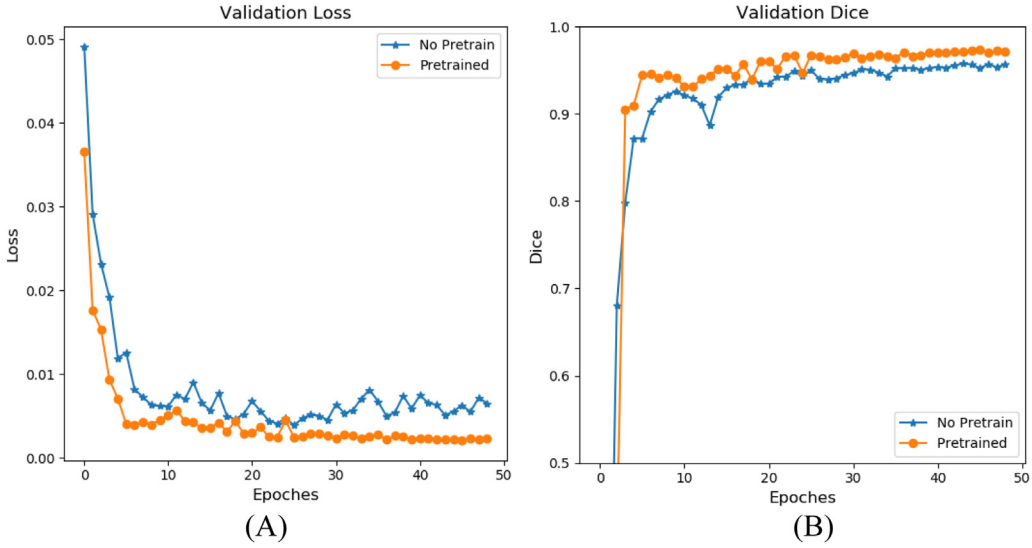

Fig. 12. Comparison of training using pretrained weight versus training from scratch along with training epochs. (A) Comparison of validation loss between two training strategies. (B), Comparison of validation dice between two training strategies.

considering the rest of the evaluation. However, it is worthwhile to point out that the model used by Sevastopolsky and Zilly et al. were trained on the same database where the performances were evaluated. We have also fine-tuned our model on both DRISHTI-DS training dataset and RIM-ONE dataset. The fine tuned model achieves the state-of-the-art performance for cup and disc segmentation in terms of dice value for both DRISHTI-DS and RIM-ONE dataset.

Although the proposed model achieves state-of-the-art performance on the DRISHTI-GS and RIM-ONE datasets, it has several limitations. Firstly, the model performance might be susceptible to images with bad quality. By comparing the model performance without fine-tuning in Tables 2 and 3, there is a decrease in both cup and disc segmentation result, as the image quality of RIM-ONE dataset is worse than that of DRISHTI-GS. Secondly, the model performance might drop for images with severe disc atrophy, especially for patient with pathological myopia. As seen in Fig. 10(D), the disc is mistakenly segmented around the edge of the bright disc atrophy. The main reason of the decreased performance against bad quality and disc atrophy images is because the model was trained on RIGA dataset, which contains mainly good quality images and normal discs. In order to improve the segmentation performance for images like those, the model has to be trained or fine-tuned on more images with less perfect quality and pathological discs with atrophy.

5. Conclusion

In this paper, we have developed and validated a modified U-Net architecture with pre-trained ResNet-34 as encoding down-sampling layers for the segmentation of optic disc and cup. The model can be quickly trained within limited epochs with robust performance comparable to that of the experts. When applied on unseen datasets without re-training or fine-tuning, the model achieves performance comparable to that of the state-of-the-art algorithms which were trained on the datasets. We have also achieved a correlation agreement above 80% and mean absolute error less than 0.08, when compared with ground truth values, for all the databases tested. In the near future, the algorithm will be further tested under the clinical setting to validates its performance and robustness.

Conflict of Interest

There are no conflicts of interest involved for any of the authors.

References

Al-Bander, B., Zheng, Y., 2018. Dense fully convolutional segmentation of the optic disc and cup in colour fundus for glaucoma diagnosis. Symmetry 10 (4), 87.
Almazroa, A., Burman, R., Raahemifar, K., Lakshminarayanan, V., 2015. Optic disc and optic cup segmentation methodologies for glaucoma image detection: a survey. J. Ophthalmol.
Almazroa, A., Alodhayb, S., Osman, E., Ramadan, E., Hummadi, M., Almazroa, A., Alodhayb, S., Osman, E., Ramadan, E., 2018. Retinal fundus images for glaucoma analysis: the RIGA dataset. Med. Imaging 2018: Imaging Inform. Healthc. Res. Appl. 10579 (March), 105790B.
Bourne, R.R.A., 2006. Worldwide glaucoma through the looking glass. Br. J. Ophthalmol. 90 (3), 253–254.
Chaurasia, A., Culurciello, E., 2017. December Linknet: exploiting encoder representations for efficient semantic segmentation. In: 2017 IEEE Visual Communications and Image Processing (VCIP), IEEE, pp. 1–4.
Cheng, J., Liu, J., Wong, D.W.K., Yin, F., Cheung, C., Baskaran, M., Aung, T., Wong, T.Y., 2011. Automatic optic disc segmentation with peripapillary atrophy elimination. Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, 6224–6227.
Cheng, J., Liu, J., Tao, D., Yin, F., Wong, D.W.K., Xu, Y., Wong, T.Y., 2013a. Superpixel classification based optic cup segmentation. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 8151 LNCS (PART 3)., pp. 421–428.
Cheng, J., Liu, J., Tao, D., Yin, F., Wong, D.W.K., Xu, Y., Wong, T.Y., 2013b. Superpixel classification based optic disc and optic cup segmentation for glaucoma screening. IEEE Trans. Med. Imaging 32 (6), 1019–1032.
Damon, W.W.K., Liu, J., Meng, T.N., Fengshou, Y., Yin, W.T., 2012. Automatic detection of the optic cup using vessel kinking in digital retinal fundus images. Proceedings – International Symposium on Biomedical Imaging, 1647–1650.
Decenciere, E., Zhang, X., Cazuguel, G., Lay, B., Cochener, B., Trone, C., Gain, P., Ordonez-Varela, J.R., Massin, P., Erginay, A., Charton, B., Klein, J.C., 2014. Feedback on a publicly distributed image database: the Messidor database. Image Anal. Stereol. 33 (3), 231–234.
Fu, H., Cheng, J., Xu, Y., Wong, D.W.K., Liu, J., Cao, X., 2018. Joint optic disc and cup segmentation based on multi-label deep network and polar transformation. IEEE Trans. Med. Imaging, 1–9.
Fumero, F., Alayon, S., Sanchez, J.L., Sigut, J., Gonzalez-Hernandez, M., 2011. RIM-ONE: an open retinal image database for optic nerve evaluation. In: Proceedings – IEEE Symposium on Computer-Based Medical Systems, no. July, pp. 4–9.
Garway-heath, D.F., Ruben, S.T., Viswanathan, A., Hitchings, R.A., 1998. Vertical cup/disc ratio in relation to optic disc size: its value in the assessment of the glaucoma suspect. Br. J. Ophthalmol. 80 (10), 1118–1124.
Giangiacomo, A., Coleman, A.L., 2009. The epidemiology of glaucoma. In: Glaucoma. Springer Berlin Heidelberg, pp. 13–21.
Iglovikov, V., Shvets, A., 2018. TernausNet: U-Net with VGG11 Encoder Pre-Trained on ImageNet for Image Segmentation. arXiv preprint arXiv:1801.05746.

Issac, A., Partha Sarathi, M., Dutta, M.K., 2015. An adaptive threshold based image processing technique for improved glaucoma detection and classification. Comput. Methods Programs Biomed. 122 (2), 229–244.
Joshi, J.S.K.K., Datt, G., Krishnadas, S.R., 2010. Optic disk and cup boundary detection using regional information. 2010 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 948–951.
Joshi, G.D., Sivaswamy, J., Krishnadas, S.R., 2011. Optic disk and cup segmentation from monocular color retinal images for glaucoma assessment. IEEE Trans. Med. Imaging 30 (6), 1192–1205.
Kumar, S., Giubilato, A., Morgan, W., Jitskaia, L., Barry, C., Bulsara, M., Constable, I.J., Yogesan, K., 2007. Glaucoma screening: analysis of conventional and telemedicine-friendly devices. Clin. Exp. Ophthalmol. 35 (3), 237–243.
Kumar, J.H., Pediredla, A.K., Seelamantula, C.S., 2015. Active discs for automated optic disc segmentation. 2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP), 225–229.
Loshchilov, I., Hutter, F., 2016. Sgdr: Stochastic Gradient Descent With Warm Restarts, . arXiv preprint arXiv:1608.03983.
Mary, M.C.V.S., Rajsingh, E.B., Jacob, J.K.K., Anandhi, D., Amato, U., Selvan, S.E., 2015. An empirical study on optic disc segmentation using an active contour model. Biomed. Signal Process. Control 18, 19–29.
Mary, M.C.V.S., Rajsingh, E.B., Naik, G.R., 2016. Retinal fundus image analysis for diagnosis of glaucoma: a comprehensive survey. IEEE Access 4, 4327–4354.
Mendels, F., Heneghan, C., Thiran, J., 1999. Identification of the optic disk boundary in retinal images using active contours. Proceedings of the Irish Machine Vision and Image Processing Conference, 103–115.
Michelson, G., Warntges, S., Hornegger, J., Lausen, B., 2008. The papilla as screening parameter for early diagnosis of glaucoma. Dtsch. Arztebl. Int. 105 (34–35), 583–589.
Mitchell, P., Smith, W., Attebo, K.H.P.R., 1996. Prevalence of open angle glaucoma in Australia: the Blue mountains eye study. Ophthalmology 103 (10), 1661–1669.
Noor, N.M., Khalid, N.E., Ariff, N.M., 2013. Optic cup and disc color channel multi-thresholding segmentation. Proceedings – 2013 IEEE International Conference on Control System, Computing and Engineering, ICCSCE 2013, 530–534.
Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: convolutional networks for biomedical image segmentation. International Conference on Medical Image Computing and Computer-Assisted Intervention, 234–241.
Sevastopolsky, A., 2017. Optic disc and cup segmentation methods for glaucoma detection with modification of U-Net convolutional neural network. Pattern Recognit. Image Anal. 27 (3), 618–624.
Shankaranarayana, M., Ram, S.M., Mitra, K., Sivaprakasam, K., 2017. Joint optic disc and cup segmentation using fully convolutional and adversarial networks. Fetal, Infant and Ophthalmic Medical Image Analysis, vol. 10554. Springer, Cham, pp. 168–176.
Sivaswamy, J., Chakravarty, A., Datt Joshi, G., Abbas Syed, T., 2015. A comprehensive retinal image dataset for the assessment of glaucoma from the optic nerve head analysis. JSM Biomed. Imaging Data Papers 2 (1), 1004.
Thakur, N., Juneja, M., 2018. Survey on segmentation and classification approaches of optic cup and optic disc for diagnosis of glaucoma. Biomed. Signal Process. Control 42, 162–189.
Zilly, J., Buhmann, J.M., Mahapatra, D., 2017. Glaucoma detection using entropy sampling and ensemble learning for automatic optic cup and disc segmentation. Comput. Med. Imaging Graphics 55, 28–41.