Retinal Abnormalities Recognition Using Regional Multitask Learning

Xin Wang[1], Lie Ju[1], Xin Zhao[1], and Zongyuan Ge[1,2](✉)

[1] Airdoc LLC, Beijing, China
{wangxin,julie,zhaoxin,gezongyuan}@airdoc.com
[2] Monash eResearch Center, Monash University, Clayton, Australia

Abstract. The number of people suffering from retinal diseases increases with population aging and the popularity of electronic screens. Previous studies on deep learning based automatic screening generally focused on specific types of retinal diseases, such as diabetic retinopathy and glaucoma. Since patients may suffer from various types of retinal diseases simultaneously, these solutions are not clinically practical. To address this issue, we propose a novel deep learning based method that can recognise 36 different retinal diseases with a single model. More specifically, the proposed method uses a region-specific multi-task recognition model by learning diseases affecting different regions of the retina with three sub-networks. The three sub-networks are semantically trained to recognise diseases affecting optic-disc, macula and entire retina. Our contribution is two-fold. First, we use multitask learning for retinal disease classification and achieve significant improvements for recognising three main groups of retinal diseases in general, macular and optic-disc regions. Second, we collect a multi-label retinal dataset to the community as standard benchmark and release it for further research opportunities.

1 Introduction

Many retinal diseases, such as glaucoma, age-related macular degeneration (AMD) and diabetic retinopathy (DR), lead to irreversible vision loss or even blindness [1]. Fundus photograph, where the microcirculation can be observed directly, is widely used to examine and screen eye diseases for early intervention in the asymptomatic stage. Deep learning has been used for computer-aided diagnosis of retinal diseases on fundus images [2–5]. For DR screening, [2,3] used convolutional neural networks (CNNs) to perform lesion detection and DR grading. Although the performance of those methods is impressive in controlled experimental settings, most of them are designed and tested for one specific retinal disease only. Real clinical scenarios where some patients suffer from various

Electronic supplementary material The online version of this chapter (https://doi.org/10.1007/978-3-030-32239-7_4) contains supplementary material, which is available to authorized users.

D. Shen et al. (Eds.): MICCAI 2019, LNCS 11764, pp. 30–38, 2019.
https://doi.org/10.1007/978-3-030-32239-7_4

and multiple retinal diseases are not considered in those models. For instance, people who suffer from DR may suffer from macular edema and optic-disc edema as well. Diseases such as hypertensive retinopathy, retinal detachment, epiretinal membrane, macular hole and macular edema may be co-occurrent with one another which leads to a challenging multi-label categorisation problem.

Moreover, global and local region information fusion has been used to improve the model performance. A novel Disc-aware ensemble network is proposed for glaucoma screening in [5], which integrates the deep hierarchical context of the entire retina and the local optic region. However, most existing methods use global and local network to target for diseases with the same distribution, region-based disease information is not explicitly used. Many researches show that the symptom from glaucoma is often associated with elevated intraocular pressure in optic-disc area. Various medical signs linked to some specific eye diseases including hemorrhage, blood vessel abnormalities (tortuosity, pulsation and new vessels) and pigmentation may tend to appear frequently in one region (e.g. macular or optic-disc region).

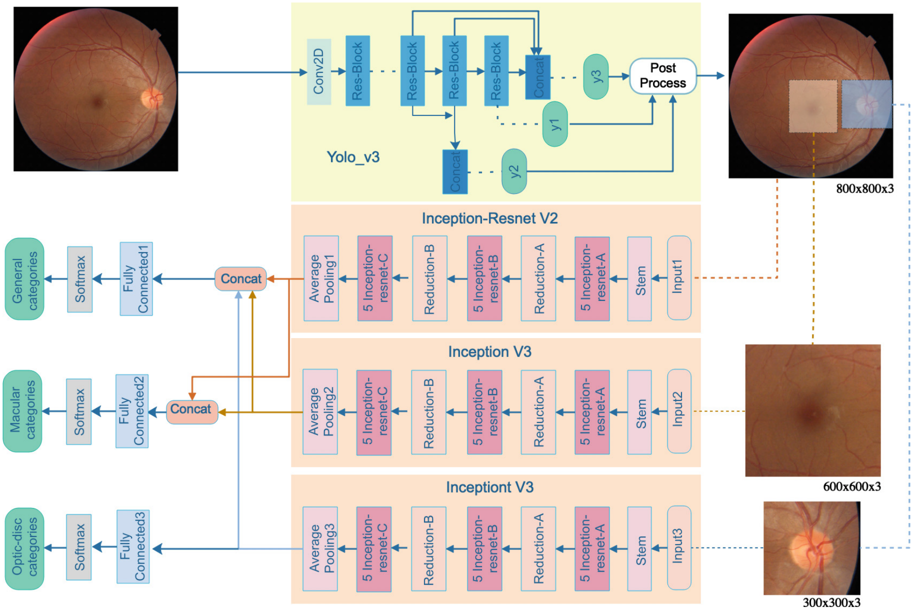

Fig. 1. The multi-label classification network has been split into three sub-networks and trained for three mutual exclusive tasks: a general task to detect diseases affect the whole retina (DR, CRVO/BRVO etc.), a **macular** sub-network to identify macular diseases (drusen, macular edema etc.) and a **optic-disc** network component to detect optic-disc related diseases (glaucoma, optic atrophy etc.). Because the features representing each of these region tasks are relevant, we design a hierarchical fusion strategy to combine late semantic representations.

In this paper, we propose a multi-task deep learning framework for identifying multi-label retinal diseases using fundus images. Our method is inspired by how ophthalmologists observe fundus images to make diagnoses and determine therapies. They firstly scan the whole fundus image to examine if there is any abnormality in color, texture or any dispersed lesion in general and then switch their attention to macular and optic-disc regions which are more vital to center vision. In order to simulate ophthalmologists' behaviour, we adopt the idea of multi-stream network [6] by using three separate sub-networks to extract features from **macula**, **optic-disc** and whole **fundus** regions separately. Additionally, we design the network in a multi-task manner by employing the prior clinical knowledge about label co-occurrence dependencies in each task (as shown in Fig. 1 from the Appendix). Our framework consists of two stages. The first stage contains a joint detector for detection of **optic-disc** and **macula** regions. The second stage is composed of a semantic multi-task network where each task is trained with exclusive region-related disease labels to output disease categories of whole fundus, optic-disc and macula concurrently, as shown in Fig. 1. To evaluate the performance of our proposed method, we collect and relabel a number of 200,817 images with 36 categories in the dataset and 17,385 images from this dataset contain more than one labels.

2 Datasets

Data Annotation and Preparation. To the best of our knowledge, there does not exist a large fundus dataset contains multi-label retinal diseases. To evaluate our proposed method, we present a multi-label dataset that contains 200,817 fundus images, including 18,614 re-labelled images from Kaggle contest dataset [1] and 182,203 de-identified samples collected from several private hospitals. Apart from that, 2,000 fundus images are annotated with optic-disc and macular location bounding boxes for training the region detection network. All the images from either Kaggle or private hospitals are re-labelled by three ophthalmologists. The labels of one fundus image preserve only if at least two ophthalmologists are in agreement. We choose 36 retinal diseases that are commonly examined during the screening including diseases affected entire retina (diabetic retinopathy etc.), optic-disc (glaucoma etc.) and macula (drusen, edema, membranes etc.). The dataset is divided for training (80%), validating (10%) and testing (10%) in this work.

Multi-label. Among all of these images, 183,432 images have single-class labels, 16,849 images have dual-class labels and 536 images have triple-class labels. The distribution of the categories and their co-occurrences are shown in Fig. 1 from the Appendix. To summarise, the most frequently co-occurrent labels are central or branch retinal vein occlusion (CRVO/BRVO) with macular edema (4,147 samples), and pathologic myopia with choroidal neovascularization (CNV) (2,467 samples). We will make the test dataset publicly available to the community for benchmark comparison and extenable future study.

[1] https://www.kaggle.com/c/diabetic-retinopathy-detection.

3 Methods

Overview. Our proposed method contains two parts: (1) macular and optic-disc region detection; (2) semantic multitask learning for retinal disease classification. We first train a joint CNN detector to localise optic disc and macula regions. The architecture extends the Yolov3 [7] with geometric constraints. The detected optic-disc and macular region bounding boxes along with whole fundus image are resized to 300 × 300, 600 × 600, 800 × 800 respectively and then fed into a three-stream multi-label disease classification network. The classification network uses the idea of semantic feature fusion to categorise regionally based diseases.

3.1 Macular and Optic-Disc Region Detection

Optic-disc and the macula are critical regions in the diagnosis of many diseases. To identify these regions, recent works such as [8] trained a Yolo detector and demonstrated accurate detection results. We extend Yolov3 [7]'s structure to infer and calibrate optic disc and macula regions. Yolov3 predicts candidate bounding boxes with 3 different scales (i.e. $y1$, $y2$ and $y3$ in Fig. 1) and perform postprocessing techniques such as non-maxima suppression [9] to select the most appropriate bounding boxes for each object. With prior knowledge of the centre distance between optic-disc and macula is approximately equivalent to two and half times as the diameter of the optic disc [10], it becomes reasonable to model and infer target regions from one another. Therefore, based on the bounding box annotations of macula and optic-disc, we add an auxiliary bounding box including those two regions as an extra cue for training and inference.

Geometric Constraints. Let (x_0^{OD}, y_0^{OD}, x_1^{OD}, y_1^{OD}) denote the locations (upper left and bottom right coordinates of optic-disc for the bounding box) and (x_0^{MA}, y_0^{MA}, x_1^{MA}, y_1^{MA}) as macular bounding box. The auxiliary bounding box is defined as,

$$\begin{aligned} x_0^{AUX} &= \min(x_0^{OD}, x_0^{MA}), x_1^{AUX} = \max(x_1^{OD}, x_1^{MA}) \\ y_0^{AUX} &= \min(y_0^{OD}, y_0^{MA}), y_1^{AUX} = \max(y_1^{OD}, y_1^{MA}) \end{aligned}$$

as shown in Fig. 2(a). During training, the auxiliary bounding box is regressed to assists the model learning the implicit relations between the macula and optic-disc regions. While in inference process, the auxiliary bounding box could be used to deduce the optic disc or macula location in some circumstances where one of the two regions is failed to be detected. Because the individual region detectors are less than perfect, the candidate region window is not always correct or even missing occasionally, especially when macula or optic-disc region is at low quality or with occlusion. In such circumstances, when the proposed auxiliary bounding box can be used to deduce the optic disc or macula location explicitly, as shown in Fig. 2(b) and (c), the missing macula bounding box location is calculated as:

$$
\begin{aligned}
&s = sgn(|x_1^{AUX} - x_1^{OD}| - |x_0^{AUX} - x_0^{OD}|), w^{AUX} = x_1^{AUX} - x_0^{AUX} \\
&x_0^{MA} = \max\left(s\left(x_1^{AUX} - \frac{1}{3}w^{AUX}\right), x_0^{AUX}\right), x_1^{MA} = x_0^{MA} + \frac{1}{3}w^{AUX} \\
&y_0^{MA} = y_0^{AUX}, y_1^{MA} = y_1^{AUX}
\end{aligned}
$$

where, sgn indicates signum function and s indicates whether the detected optic-disc box is at the left of the auxiliary box; w^{AUX} denotes the width of the auxiliary box. Based on the prior domain knowledge, we approximate the location of the macular box to be at either left or right (depends on which eye being scanned) one-third part of the auxiliary box, in which of in the opposite direction to the optic-disc box. To be complete, in the rare case where both macula and optic-disc boxes are missing even with the assistance of auxiliary box, the whole fundus image is adjusted and used as input to the macula and optic-disc stream.

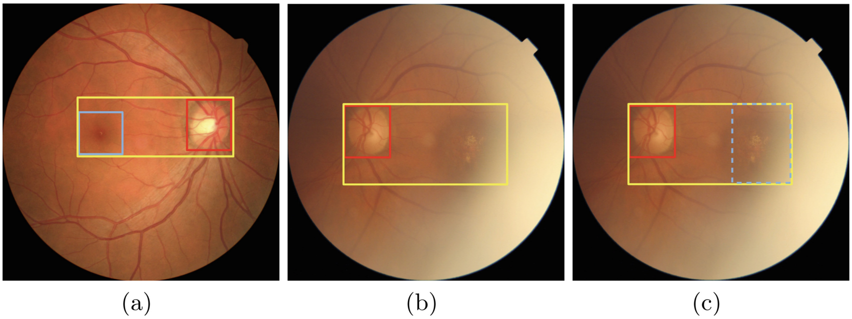

Fig. 2. Illustration of auxiliary bounding box AUX. (a) annotations of optic-disc bounding box OD in red, macular box MA in blue and the auxiliary bounding box AUX of joint region in yellow; (b) the detection result with missing macula MA because of low quality/blurry region; (c) localisation of macula MA in dash blue from (b) through AUX post-processing described in Sect. 3.1 (Color figure online)

3.2 Semantic Multitask Learning for Retinal Disease Classification

There are three main groups of retinal diseases, general retina diseases indicate diseases that influence the entire retina, such as DR, hypertensive retinopathy and CRVO/BRVO etc. Macular diseases include age-related macular degeneration (AMD), macular edema and macular hole, while optic-disc diseases contain glaucoma, optic-disc edema, optic atrophy etc. By analyzing medical signs of these diseases, many studies found that some diseases belong to different tasks share some common pathological features or have some medical signs correlations with each other. From those observations, we propose a collaborative multi-task learning framework with three streams. Each stream of the network represents learning process for one group of retinal diseases. One important concept of the proposed framework is that part of the model layer is shared across independent tasks. In this work, the layer sharing strategy is semantically designed with respect to the knowledge of regionally disease correlation.

General task stream is designed to be a task that determines the general retinal disease affecting the entire retina based on the features from optic-disc and macular regions as well as the whole fundus image. The motivation to design this branch is inspired by the fact that most general retinal diseases are diagnosed in consideration of lesions on the macula and optic disc. For instance, one of the vital medical signs to diagnose proliferative diabetic retinopathy (PDR) is whether there is neovascularisation on the optic-disc. As for diagnosing pathologic myopic, atrophy on macula can serve as a strong evidence. Therefore, it is reasonable to add region-specific features of macula and optic-disc as supporting information to determine the general retinal disease category. **Macular task stream** takes advantage of features from macular region and entire retina because some sub-type macular edema diseases are closely correlated to general retinal disease such as DR. To further improve the intra-class performance of this branch, we made macular edema into three sub-categories, macular edema caused by DR or hypertensive retinopathy, CRVO/BRVO and other diseases. The same operation is applied to pathologic myopia w/o CNV. **Optic-disc task stream** is a relatively independent task because its categories are self-contained and the regionally independent compared to other regions.

Inference: During inference process, each task stream outputs either healthy or diseased class which has the highest confidence score. A patient is diagnosed as healthy when all task streams (macular, optic-disc and whole fundus regions) indicate non-disease outcome.

Implementation: Inception-V3 is used as backbone for optic-disc and macular task streams, and Inception-Resnet-V2 is used for general retinal task[2]. Each task stream is initialised with ImageNet pre-trained parameters and trained separately before jointly trained together.

4 Experiments

4.1 Results on Optic-Disc and Macula Joint Detection

The joint detection YoloV3 network with joint constraint is trained and validated on 2,000 fundus images with annotated macular and optic-disc bounding boxes. All the 2,000 images are divided into training, validating and testing in the

[2] We experimented with all Inception-V3/Inception-Resnet-V2 settings and figured out a mixture of them gave the best performance of all. Network training used Adam with a learning rate of 1e-5 which decayed every three epochs with ratio of 0.9 for total 14 epochs. We used Keras distributed machine learning system with 8 replicas running each on a NVidia 1080Ti GPU. The input size of general stream is 800, while the input size of macular and optic-task stream is 600 and 300 respectively.

proportion of 8:1:1. Experiment on the testing dataset obtains 83.45% mAP with minimal IoU 0.5, with macula(90.03% AP) and optic disc (96.73% AP). We then run the trained detector on the whole fundus dataset with 200,817 images. After carefully checked by three ophthalmologists, only 4,824 (about 2.4% out of the whole dataset) macular or optic-disc regions are not correctly detected. Most of these cases are either blurry or with very serious lesions (see Appendix Fig. 2).

4.2 Quantitative Evaluation on Multitask Learning

Table 1 shows average precision and recall for task based classification. We evaluate the effectiveness of applying multi-stream on macular task and general task. For macular task, we first obtain the results of one-stream network which is trained on macular images (detected by YoloV3 detector) with 14 macular (MA) relevant retinal diseases (**MA-One-Stream (with single task)**), and compare it to **MA-Two-Stream (with single task)** which takes both the whole fundus image and macular image as inputs. We then conduct the same experiment on the general disease task (GC). In both experiments of MA and GC, our multi-stream method works better than single stream network baseline. We then extend the multi-stream experiments with extra multitask label learning, marginal performance improvements can be observed for **GC-Three-Stream (with multitask)**. More detailed confusion matrix of these frameworks for various tasks are shown in Figs. 3 and 4 in the Appendix.

Table 1. Results on task based classification

Methods	Average recall	Average precision
MA-One-Stream (single task)	68.8%	61.3%
MA-Two-Stream (single task)	**73.6%**	67.6%
MA-Two-Stream (multitask)	73.1%	**68.6%**
GC-One-Stream (single task)	65.2%	60.8%
GC-Three-Stream (single task)	67.5%	61.5%
GC-Three-Stream (multitask)	**67.8%**	**62.2%**

The second set of results in Table 2 focuses on fully 36 diseases trained multitask deep neural network performance for each disease group (macula, optic-disc and whole fundus). As Table 2 shows, in this setting our proposed multi-task method works much better than the baseline single stream model. The performance of all diseases has improvements in both precision and recall, while diseases relevant to macula and disc acquire greater improvement than diseases not relevant to these two regions compared with single stream.

Table 2. Results on 36 category classification on different disease regions

Methods	Regions	Average recall	Average precision
One-Stream	Macula	60.9%	61.1%
Three-Stream	Macula	**62.6%**	**63.0%**
One-Stream	Disc	57.5%	69.7%
Three-Stream	Disc	**62.4%**	**70.0%**
One-Stream	General	69.2%	60.5%
Three-Stream	General	**70.6%**	**61.4%**

4.3 Visualization

In order to gain a better understanding of the proposed model, we draw the class activation maps (CAM) [11] for each stream with respect to each task. As shown in Fig. 3, the case in Fig. 3 is a fundus image with PDR, macular edema and other optic-disc disease[3]. Our model gives three correct labels via the multitask architecture. From the CAM generated by the general task of individual regions, we can observe that when PDR, proliferative membranes and hemorrhage appear over the whole fundus image, neovascularisation grows around the optic-disc, and more activation of exudates from macular region are taken into consideration. Moreover, lesions from the entire fundus are also used to help distinguish diabetic macular edema (DME) from other macular diseases. The membrane and neovascularisation on the optic-disc are highlighted by its task stream to correctly give the label of other optic-disc disease.

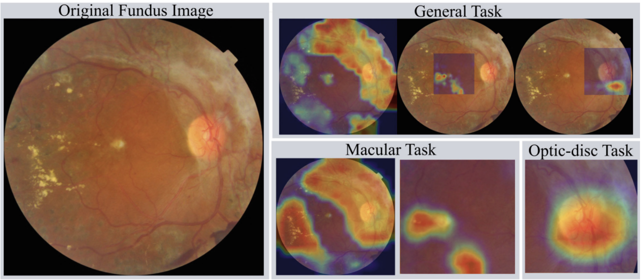

Fig. 3. Class activation maps (CAM) of a challenging multi-label sample, CAM generated for each task from a fundus image with PDR, macular edema and other optic-disc disease.

[3] We set a label named as other general, macular or optic-disc disease to indicate a gathering of rare diseases in each task. This label for optic-disc task represents disease such as morning glory syndrome, melanocytoma of optic disc, membrane tissue on the optic disc and etc.

From visualization of this case, we can see that the proposed model can mine the region-specific information with respective to various disease types and improve the performance of disease recognition through collaborative multi-label learning.

5 Conclusion

In this work, we demonstrate the effectiveness of multitask learning approach for recognising the general, macular and optic-disc diseases, as opposed to single task classification. The presented method and new benchmark open the possibility for large scale multi-label retinal abnormalities recognition.

References

1. Tham, Y.-C., Li, X., Wong, T.Y., Quigley, H.A., Aung, T., Cheng, C.-Y.: Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. Ophthalmology **121**(11), 2081–2090 (2014)
2. Wang, P., et al.: Development and validation of a deep-learning algorithm for the detection of polyps during colonoscopy. Nat. Biomed. Eng. **2**(10), 741 (2018)
3. Playout, C., Duval, R., Cheriet, F.: A multitask learning architecture for simultaneous segmentation of bright and red lesions in fundus images. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11071, pp. 101–108. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00934-2_12
4. Grassmann, F., et al.: A deep learning algorithm for prediction of age-related eye disease study severity scale for age-related macular degeneration from color fundus photography. Ophthalmology **125**(9), 1410–1420 (2018)
5. Fu, H., et al.: Disc-aware ensemble network for glaucoma screening from fundus image. IEEE Trans. Med. Imaging **37**(11), 2493–2501 (2018)
6. Simonyan, K., Zisserman, A.: Two-stream convolutional networks for action recognition in videos. In: Advances in Neural Information Processing Systems, pp. 568–576 (2014)
7. Redmon, J., Farhadi, A.: Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767 (2018)
8. Araújo, T., Aresta, G., Galdran, A., Costa, P., Mendonça, A.M., Campilho, A.: UOLO - automatic object detection and segmentation in biomedical images. In: Stoyanov, D., et al. (eds.) DLMIA/ML-CDS -2018. LNCS, vol. 11045, pp. 165–173. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00889-5_19
9. Bodla, N., Singh, B., Chellappa, R., Davis, L.S.: Soft-NMS-improving object detection with one line of code. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 5561–5569 (2017)
10. Sinthanayothin, C., Boyce, J.F., Cook, H.L., Williamson, T.H.: Automated localisation of the optic disc, fovea, and retinal blood vessels from digital colour fundus images. Br. J. Ophthalmol. **83**(8), 902–910 (1999)
11. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., Torralba, A.: Learning deep features for discriminative localization. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2921–2929 (2016)