Assisting Retina image Diagnosis with DNN-based Object Detection

Yung-An Hsieh, Sheng-Chun Kao, Hanwen Cao

School of Electrical and Computer Engineering
Georgia Institute of Technology
Atlanta, United States

ABSTRACT

Our aim is to develop a system to detect diabetic retinopathy in color digital retinal images automatically and to evaluate its potential in diabetic retinopathy assistance diagnosis. A system was used involving pre-processing to normalize color and enhance edge and labeled according to ground truth. The system was trained using 100 images from two datasets and evaluated by accuracy and precision. The system can achieve 40.5% average area accuracy, 79.1% average area precision, 73.8% case accuracy and 99.5% case precision. This system could be used to help doctor identify diabetic retinopathy with high precision and accuracy, even when the image has no perfect contrast and lighting condition.

Index Terms**—** diabetic retinopathy, assistance diagnosis, objection detection, YOLO, R-CNN

1. INTRODUCTION

Diabetic retinopathy is a typical retinal disease that affects millions of people all over the world, several works have been done on automatic diagnosis using machine learning based methods. [1] However, after discussed with the cooperating physician, we found that "currently a more helpful task of AI for doctors is not diagnosing the disease, but discovering symptoms that assist them to diagnose". Therefore, we aimed to find a disease detection method that can detect symptoms on retina images to serve as the first-stage diagnosis of Diabetic Retinopathy. In the traditional diagnose framework, limited numbers of doctors identify symptoms of thousands of patients, which is a time consuming process. Our proposed framework is to use machines to identify retinal disease symptoms automatically and quickly to help doctors diagnose. Because of the automatic symptoms detection, doctors can save time finding symptoms on retina images, and concentrated more on diagnosis.

There are several ways to detect and localize objects in given images. In this paper we describe our progress in the development of an automated system for identifying and localization of diabetic retinopathy. Two object detection networks were compared for the suitability of diabetic retinopathy detection. The system were trained on one hundred digital fundal images from DIARETDB1 [2] and IDRiD [3]. Since our work is different from normal object detection and retina image diagnosis, we proposed two new performance metric, which are case accuracy and case precision. We also used different pre-processing methods and discussed about their effect on training and detection. Our system can detect the retinopathy that is hard to find by traditional algorithm like Pattern matching and Background cancelation with high precision.

2. RELATED WORKS

Sinthanayothin [4] et al. applied recursive region growing segmentation (RRGS) technique. This system treat haemorrhages and microaneurysms (HMA) as one group, but it can only distinguish these group with exudates. Niemeijer [5] et al. proposed a method to detect candidate red lesions (microaneurysms and haemorrhages) based on pixel classification. Usher [6] et al. used an RRGS, adaptive intensity thresholding to extract the candidate red lesions. Candidate red lesions were classified using a neural network. These two methods are used for national annual screening for diabetic retinopathy, but not designed to localize disease in retina images. None of them concentrate on detect and localize each disease in pixel precision. Girshick [8] et al. and Cai [9] et al. achieved high precision for detecting normal object. But their model cannot be applied to retina images and diabetic retinopathy directly. achieved high precision for detecting normal object. But their model cannot be applied to retina images and diabetic retinopathy directly. Our work aims to identify diabetic retinopathy with high precision and accuracy to help doctors diagnose by identifying and labeling the possible disease regions precisely in given funds images. In this way, this system can speed up the diagnosis process and relief the burden of doctors.

3. IMAGE LABELING

One hundred color digital fundal images from DIARETDB1 - Standard Diabetic Retinopathy Database and IDRiD - Indian Diabetic Retinopathy Image Dataset were used from training and validation. Each image contains ground truth of diabetic retinopathy and optical disc. First we generate labels of four kinds of diabetic retinopathy, hard exudates, soft exudates, microaneurysms and haeorrhages, and optic disc based on the ground truth.

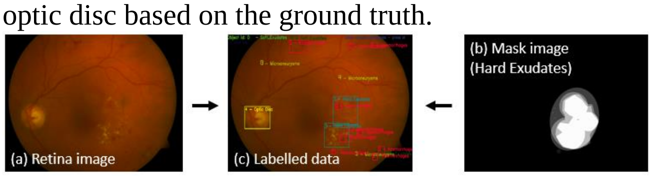

Fig. 1. Original fundus image (a), mask image (b), and generated labelled image (c) for object detection training

Each label was a rectangular bounding box covering the mask made by doctors as shown in Fig. 1. These funds images with labels were then pre-processed by color normalization, contrast enhancement and edge enhancement.

4. IMAGE PRE-PROCESSING

Different fund images from different patients taken under different lightening conditions may have different color and brightness. Therefore, we normalize the color and brightness of all the input images. After that, the contrast of images was enhanced to make the details stand out of the backgrounds. Finally, we used edge detection to enhance the edge of diabetic retinopathy. We further compared different pre-processing methods. The final input images of DNN-based network were images weighted combining these different methods. Fig. 2. shows the idea of weighted summation. By changing the weights, we observe the detection results.

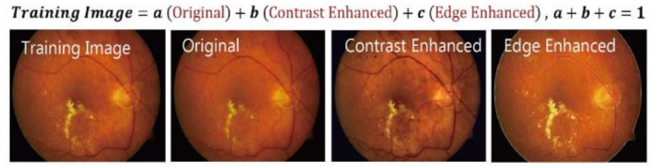

Fig. 2. Combination of original image and different enhanced images

5. FASTER R-CNN AND YOLO

The first DNN-based network we used was Faster R-CNN. Faster R-CNN is a state-of-the-art object detection networks depend on region proposal algorithms to hypothesize object locations [10]. In Faster R-CNN, the region proposal network and detection network are trained separately based on the bounding box region only. The second one is YOLO (You Only Look Once) [11]. It only has a Single network for both bounding box and classification. And it is train by whole image, preserving whole image information. A comparison of Faster R-CNN and YOLO is listed in the table I.

TABLE I. Quick Comparison between two algorithms

	Faster R-CNN	**YOLO**
Resource	Large	Small
latency	High	Low
Computation	High	Low
Image info.	No	Yes
Summary	Able to perform well, even better than YOLO in some performance criteria. But demand very high computation power, very high hardware resources.	A balanced trade-off between performance and Complexity. Potential of widely usage in resource-constrained device. **We choose YOLO to develop more significant approach for Retina image**.

We finally choose YOLO to develop our system and detect diabetic retinopathy.

6. EXPERIMENTS SETUP

In this chapter, we would discuss about how our experiments were carried out. Specifically, the dataset, performance metric and validation methods would be talked about.

6.1. Datasets

We used two public datasets. DIARETDB1 [2] contains 89 images. The second dataset we used is

IDRiD [3], which consists of 54 color fundus images. From these images we picked 100 images into 5 sets to training and validate. The validation process is described below.

6.2. Performance Metric

There are several validation methods for object detection. The conventional metrics for disease diagnosis are accuracy and precision, where accuracy differentiate the patient and healthy cases correctly, and precision determines the patient cases correctly. We used accuracy to measure the ability of our system to identify symptoms and precision to evaluate how accuracy the symptoms are localized. Accuracy can be calculated by $\frac{tp+tn}{tp+tn+fp+fn}$, and precision can be calculated using $\frac{tp}{tp+fp}$, where tp is true positive, tn is true negative, fp is false positive and fn is false negative.

The traditional definitions of accuracy and precision would not work well. Therefore, we modified metrics that to make them suitable for our object detections task. First, we used area-overlapped based accuracy and precision, which are similar to the definitions above. They basically concentrate on the overlapped area between ground truth and the detected region, then they could tell us how accurate the system is in terms of detecting symptoms and locating them. Besides, we came up with case-based accuracy and precision, which only consider true or false values. In case-based accuracy and precision, creating a smaller or shifted detected region is still adequate good for symptoms detection. What they concern about is whether the system can identify the possible symptoms. In this way, not only do we consider how precision our detection is, but also the ability to identify symptoms. The difference between area-based and case-based metrics is shown in Fig. 3.

6.3. Validation

We used cross validation to validate the results of our experiments. The set of input images were divided into subsets, and separate experiments and validation were performed on each subset.

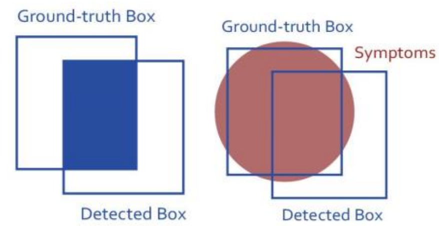

Fig. 3. Area-based metrics (left) considers the overlap (blue) region of ground truth and detected box. In case-based metrics (right), the symptoms are considered detected even if the detected box is not perfectly overlap with ground truth box

7. EXPERIMENT RESULTS

In our experiments, we aim to discuss the performance of our models and the effect of weighting among different image preprocessing method. We implement two image preprocessing method: first, color contrast, which enhance the color contrast of the image; second, edge enhancement, which enhance the image by its edge. We experiment through different combination of weighting, following formula (1). The experiments results are summarized in table II. It shows the seven different combination of weighting that we experiment on. For example, for the first case (a, b, c) = (1, 0, 0), it means we use the original image to train our model. For the last case (a, b, c) = (0.33, 0.33, 0.33), it tells that we equally average three processed image.

Also, due to the fact that accuracy is much more important than precision in the case of disease diagnosis, our criteria of comparing their performances is that we select the top accuracy with adequate precisions. Thus, the case (a, b, c) = (0.25, 0.25, 0.5) is the best performance in these different settings. It shows that edge information is more important in our retina image testbed. Also, our model can achieve about 80% accuracy with 99% percipiency.

We show one of the testing result in Fig. 4(a). The bounding box and its tags are the predicted results by our models. After combing the prediction and the original image, our system will generate the symptoms suggestion image for doctor, as shown in Fig. 4. Also, we can observe that most of the mutated region are correctly diagnosis. Our model, can provide accurate prediction. Fig. 4(b) shows another test case, which has very few and not obvious symptoms. Therefore, it can have more chances to be missed by doctors with their raw-eyes reading. However, our model can distinct these very minor symptoms accurately. This can help doctors to improve their experiences of diagnosing.

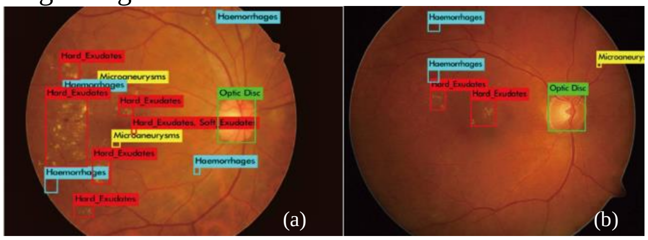

Fig. 4. The generated suggestion image of one test case.

In our experiment, we implement five-fold cross validation. To be more specific, we have 100 images in our dataset. We divide them into 80 images for training and 20 images for validation. By the 80-20 percent rules, we will have five different sets for training and validation.

TABLE II. Experiment Results of Different Weighting across Different Pre-processing Methods

a, b, c	**Area Accuracy**	**Area Precision**	**Case Accuracy**	**Case Precision**
1,0,0	40.3%	71.8%	76%	97.5%
0,1,0	39.7%	76.7%	78%	100%
0,0,1	39.3%	82.1%	69%	97.5%
1/2,1/4,1/4	40%	78.6%	78%	100%
1/4,1/2,1/4	40%	72.1%	78%	97.5%
1/4,1/4,1/2	40%	72.1%	80%	100%
1/3,1/3,1/3	40.4%	81.8%	75%	100%

In table III, we can observe that we have rather consistent performance across different sets. It tells that we are not in the over-fitting situation. Our average case accuracy is 73.8% and case precision is 99.5%. It means we have 73.8% rate that we make the right call. In addition, approximately, we will only falsely miss the symptoms one time out of 200 test cases.

TABLE III. Results of Five-fold Cross Validation

	Area Accuracy	**Area Precision**	**Case Accuracy**	**Case Precision**
Set 1	37.6%	80.3%	66%	100%
Set 2	36.7%	77.2%	75%	100%
Set 3	43.7%	81.7%	78%	100%
Set 4	40.4%	72.5%	80%	100%
Set 5	43.9%	83.9%	70%	97.5%
Average	40.5%	79.1%	73.8%	99.5%

8. DISCUSSION

Our machine-learning based object detection model on the retinal image symptoms diagnosis shows good results. However, we want to discuss about how machine learning really helps compare to the conventional methods? Intuitively, we tried pattern matching, and one of the result is shown in Fig. 6(a). It tells that no symptoms are correctly detected by the conventional pattern matching method. It is because the retina images are full of variability, and the symptoms patterns are not standardized. Therefore, without higher dimension description, we are hard to find the matched pattern directly.

In addition, we also implement, background cancelation method. The background is shown in Fig. 6(b) and the predicted image is in Fig. 6(c). We also come to no correct perdition. The conventional background cancelation method can't be very helpful, because the position, the angle, the size, the rotation are all different across the dataset. Therefore, it is hard to find a valid background for background cancellation. However, with our Machine learning based model, we have learned higher dimension feature, and our model is more robust to different variability.

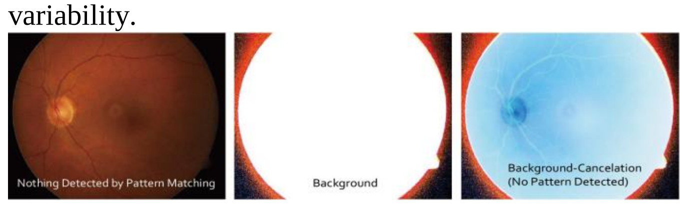

Fig. 5. (a) The result of pattern matching. (b) The background in background cancelation method. (c) The result after enhanced by background cancelation.

9. CONCLUSION

We proposed a framework that can increase the efficiency of retinal disease diagnosis. The framework has a ML-based symptoms detections stricter as a first-stage diagnosis for the doctors. For the ML-based symptoms detection structure, we proposed a modified YOLO model. The experiment results show that our model can achieve about 78% accuracy, with 99% percipiency. It shows that our model can successfully detect most of the symptoms correctly, and the high potential for it to be applied in the real diagnosis system.

10. ACKNOWLEDGEMENT

We would like to express the deepest appreciation to Huck Yang and Doctor Meng Tian, who have provided the essential assistance and invaluable advice.

11. REFERENCES

[1] R. Santhakumar, M. Tandur, E. R. Rajkumar, K. S. Geetha, G. Haritz, and K. T. Rajamani, “Machine learning algorithm for retinal image analysis,” In Proc. IEEE Region 10 Conf. (TENCON), Nov. 2016, pp. 1236–1240.

[2] DIARETDB1 dataset: http://www.it.lut.fi/project/imageret/diaretdb1/

[3] IDRiD dataset: https://idrid.grand-challenge.org/Home/

[4] Sinthanayothin C, Boyce JF, Williamson TH, Cook HL, Mensah E, LalS. Automated detection of diabetic retinopathy on digital fundus image.

[5] Niemeijer M, van Ginneken B, Staal J, Suttorp-Schulten MS, Abramoff MD. Automatic detection of red lesions in digital color fundus photographs.

[6] Usher D, Dumskyj M, Himaga M, Williamson TH, Nussey S, Boyce J. Automated detection of diabetic retinopathy in digital retinal images: a tool for diabetic retinopathy screening. Diabet Med 2004;21:84–90.

[7] Girshick, Ross, et al. "Region-based convolutional networks for accurate object detection and segmentation." IEEE transactions on pattern analysis and machine intelligence 38.1 (2016): 142-158.

[8] Cai, Zhaowei, et al. "A unified multi-scale deep convolutional neural network for fast object detection." European Conference on Computer Vision. Springer, Cham, 2016.

[9] Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. In Advances in neural information processing systems (pp. 91-99).

[10] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. IEEE conference on computer vision and pattern recognition