

# Lesion Detection and Grading of Diabetic Retinopathy via Two-Stages Deep Convolutional Neural Networks

Yehui Yang<sup>1,2(✉)</sup>, Tao Li<sup>2</sup>, Wensi Li<sup>3</sup>, Haishan Wu<sup>4</sup>, Wei Fan<sup>1</sup>,  
and Wensheng Zhang<sup>2</sup>

<sup>1</sup> Big Data Lab, Baidu Research, Beijing, China  
yangyehuisw@126.com

<sup>2</sup> Institute of Automation, Chinese Academy of Sciences, Beijing, China

<sup>3</sup> The Beijing Moslem's Hospital, Beijing, China

<sup>4</sup> Heyi Ventures, Beijing, China

**Abstract.** We propose an automatic diabetic retinopathy (DR) analysis algorithm based on two-stages deep convolutional neural networks (DCNN). Compared to existing DCNN-based DR detection methods, the proposed algorithm has the following advantages: (1) Our algorithm can not only point out the lesions in fundus color images, but also give the severity grades of DR. (2) By introducing an imbalanced weighting scheme, more attentions will be payed on lesion patches for DR grading, which significantly improves the performance of DR grading under the same implementation setup. In this study, we label 12, 206 lesion patches and re-annotate the DR grades of 23, 595 fundus images from Kaggle competition dataset. Under the guidance of clinical ophthalmologists, the experimental results show that our lesion detection net achieves comparable performance with trained human observers, and the proposed imbalanced weighted scheme also be proved to significantly enhance the capability of our DCNN-based DR grading algorithm.

**Keywords:** Diabetic retinopathy · Deep convolutional neural networks · Fundus images · Retinopathy lesions

## 1 Introduction

Diabetes is an universal chronic disease around some developed countries and developing countries including China and India [1–4]. The individuals with diabetic have high probability for having diabetic retinopathy (DR) which is one of the most major causes of irreversible blindness [5, 6]. However, according to the report from Deepmind Health<sup>1</sup>, 98% severe vision loss led by DR can be prevented by early detection and treatment. Therefore, the quickly and automatically detecting of DR is critical and urgent to reduce burdens of ophthalmologists, as well as providing timely morbidity analysis for the massive patients.

<sup>1</sup> <https://deepmind.com/applied/deepmind-health/research/>.

According to the *International Clinical Diabetic Retinopathy Disease Severity Scale* [5,9], the severity of DR can be graded into five stages: normal, mild, moderate, severe and proliferative. The first four stages can also be classified as non-proliferative DR (NPDR) or pre-proliferative DR, and NPDR may turn to the proliferative DR (PDR, the fifth stage) with high risk if without effective treatment. The early signs of DR are some lesions such as microaneurysm (MA), hemorrhages, exudate etc. Therefore, lesion detection is a less trivial step for the analysis of DR. There are plenty of literatures focus on detecting lesions in retina. Haloi *et al.* [10] achieve promising performance in exudates and cotton wool spots detection. Later, Haloi [4] try to find MAs in color fundus images via deep neural networks. van Grinsven *et al.* [7] propose a selective sampling method for fast hemorrhage detection. Additionally, Srivastava *et al.* [11] achieve robust results in finding MA and hemorrhages based on multiple kernel learning.

However, the aforementioned algorithms do not attach the DR severity grades of the input fundus images, which is vital for the treatment of DR patients. Recently, Seoud *et al.* [8] propose an automatic DR grading algorithm based on random forests [19]. By leveraging deep learning techniques [13–16], Gulshan *et al.* [5] take efforts to classify the fundus images into normal and referable DR (defined as moderate and worse DR) with the annotation of 54 United States licensed ophthalmologists on over 128 thousands fundus images. Similarly, Sankar *et al.* [12] using DCNN to grade DR into normal, mild DR and several DR. Pratt *et al.* [2] predict the severity of DR according to the five-stage standard of *International DR Scale standard* [5,9]. Even though these DR grading algorithms seem to have achieved promising performance, they still exist the following problems:

- (1) The aforementioned DCNN-based DR grading methods can only give the DR grade but can not indicate the location and type of the existing lesions in the fundus images. However, the detailed information about the lesions may be more significant than a black box for clinicians in treatment.
- (2) The above end-to-end DCNN<sup>2</sup> may not suitable to learn features for DR grading. Compared to the size of the input image, some tiny lesions (eg., MAs and some small hemorrhages) are such un conspicuous that they are prone to be overwhelmed by the other parts of input image via end-to-end DCNN. However, such lesions are critical for DR grading according to the standard [9].

To address the above issues, we propose two-stages DCNN for both lesion detection and DR grading. Accordingly, our method composed of two parts: local network to extract local features for lesion detection and global network to exploit image features in holistic level for DR grading.

Instead of end-to-end DR grading, we construct a weighted lesion map to differentiate the contribution of different parts in image. The proposed weighted lesion map gives imbalanced attentions on different locations of the fundus image

---

<sup>2</sup> End-to-end DCNN grading means that directly feed the input images into DCNN, then output the DR grades of the images.

in terms of the lesion information, i.e., the patches with more severe lesions will attract more attention to train the global grading net. Such imbalanced weighted scheme significantly improve the capability of the DR grading algorithm.

Compared to the existing DCNN-based DR analysis algorithms, the proposed algorithm has the following advantages and contributions:

- (1) We propose a two-stages DCNN-based algorithm which can not only detect the lesions in fundus images but also grade the severity of DR. The two-stages DCNNs learn more complete deep features of fundus images for DR analysis in both global and local scale.
- (2) We introduce imbalanced attention on input images by weighted lesion map to improve the performance of DR grading network. To the best of our knowledge, this is the first DNN-based work resorting imbalanced attention to learn underlying features in fundus images for DR grading.

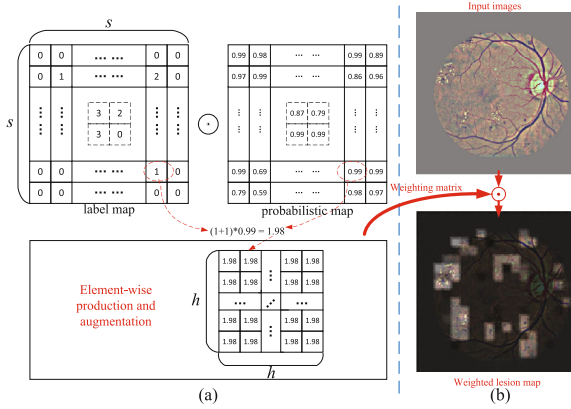
## 2 Methods

In this section, we present the details of the proposed two-stages DCNN for lesion detection and DR grading. First, the input fundus images will be preprocessed and divided into patches, and the patches are classified by the local net into different lesion types. Then, the weighted lesion map is generated based on the input image and output of local net. Third, the global network is introduced for grading the DR of input image.

- (1) *Local Network.* To detect the lesions in the fundus image, the input images are divided into  $h \times h$  patches using sliding windows with stride  $h - ov$ , where  $ov$  is the overlapped size between two adjacent patches. The local network is trained to classify the patches into 0 (normal), 1 (microaneurysm), 2 (hemorrhage), 3 (exudate), which are the main indicators to NPDR.
- (2) *Weighted Lesion Map.* Two maps are generated when all the patches in fundus image  $\mathbf{I} \in \mathbb{R}^{d \times d}$  are classified by the local network. One is label map  $\mathbf{L} \in \mathbb{R}^{s \times s}$ , which records the predicted labels of the patches. Wherein  $s = \lfloor (d - h) / (h - ov) \rfloor$ , and  $\lfloor \cdot \rfloor$  is the floor operator. The other map is probabilistic map  $\mathbf{P} \in \mathbb{R}^{s \times s}$ , which retains the the biggest output probability of the softmax layer (the last layer of the local net) [17] for each patch label.

As illustrated in Fig. 1(a), we construct a weighting matrix for each input image as: (1) Integrating the label map and probabilistic map as  $\mathbf{LP} = (\mathbf{L} + \mathbf{1}) \odot \mathbf{P}$ , where  $\odot$  is the element-wise product and  $\mathbf{1} \in \mathbb{R}^{s \times s}$  is an all one matrix<sup>3</sup>. (2) Each entry in  $\mathbf{LP}$  is augmented to a  $h \times h$  matrix (corresponding to the patch size). (3) Jointing the augmented matrixes into the weighting matrix  $\mathbf{M}_{\mathbf{I}} \in \mathbb{R}^{d \times d}$  according to the relative locations in the input image  $\mathbf{I}$ . The weighting matrix is constructed into the same size of input image, and the values in the intersection areas are set as the average values between adjacent expanded matrixes.

<sup>3</sup> The motivation of the addition of the all one matrix is to avoid totally removing the information in the patches with label 0.



**Fig. 1.** The illustration of the construction of weighted lesion map. (a) Jointing label map and probabilistic map into weighting matrix. (b) Illustration of imbalanced attention on weighted lesion map.

The weighted lesion map of input image  $\mathbf{I}$  is defined as  $\mathbf{I}^* = \mathbf{M}_{\mathbf{I}} \odot \mathbf{I}$ . The entries in the weighting matrix  $\mathbf{M}_{\mathbf{I}}$  implicit the severity and probability of lesions in local patches. Therefore, the image patches have more severe lesion patches with higher probability will get higher weights in the weighted lesion map. As seen in Fig. 1(b), imbalanced attentions are payed on the weighted lesion map by highlighting the lesion patches.

- (3) *Global Network.* The global network is designed to grade the severity of DR according to the *International Clinical Diabetic Retinopathy scale* [9]. In this study, we focus on the advanced detection on NPDR, and the NPDR can be classified into four grades: 0 (normal), 1 (mild), 2 (moderate), 3 (severe). The global network is trained with the weighted lesion maps, and the output is the severity grade of the testing fundus images.

### 3 Experimental Evaluation

#### 3.1 Data Preparation

- (1) *Database:* The Kaggle database contains 35, 126 training fundus images and 53, 576 testing photographs. All the images are assigned into five DR stages according to the international standard [9]. The images in the dataset come from different models and types of cameras under various illumination.

According to our cooperant ophthalmologists, although the amount of images in this dataset is relatively big, there exist a large portion of biased labels. Additionally, the dataset do not indicates the locations of the lesions which are meaningful to clinicians. Therefore, we select subset from Kaggle database for re-annotation. The subset consists of 23, 595 randomly selected images in terms

of the four grades of NPDR, where 22,795 for training and 800 for testing (each NPDR grade contains 200 testing images).

The training and testing patches for lesion detection are cropped from the training and testing images respectively, which totally contains 12,206 lesion patches and over 140 thousands randomly cropped normal patches. Licensed ophthalmologists and trained graduate students are invited or payed to annotate the lesions in the images and re-annotate DR grades of the fundus images.

- (2) *Data Preprocessing and Augmentation.* In this study, contrast improvement and circular region of interesting extraction are conducted on the color fundus images as [7]. For lesion detection, all the images are resized to  $800 \times 800$ , and the relative ratio between the sample height and length is kept by padding before resizing the raw images. The input sample size of global network are turned into  $256 \times 256$  to reduce the computational complexity.

Data augmentation are implemented to enlarge the training samples for deep learning, as well as to balance the samples across different classes. Inspired by [14], the augmentation methods include randomly rotation, cropping and scaling. The samples after augmentation is split into training and validating set for tuning the deep models, and the testing samples are not put into augmentation. The testing patches for local lesion detection are generated from testing images.

- (3) *Reference Standard and Annotation.* For training the local network, the patches are first annotated by two over three months trained observers, and both of them have experience in medical image processing. Then all the samples are checked by a clinical ophthalmologist. For training the global network, the labels of the fundus images are selected from Kaggle annotations by the trained observers, then the label biases are corrected by the clinical ophthalmologist.

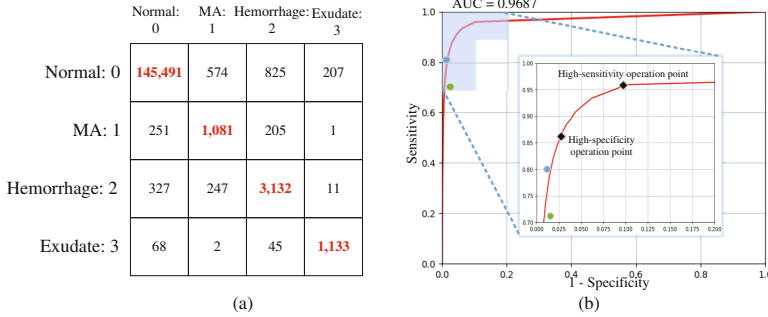
For the testing sets, firstly, all the testing lesion patches and DR grades of fundus images are annotated independently among all the trained observers and ophthalmologist, then the discrepancy patches are selected for further analyzing. Finally, the references of the samples are determined only by achieving the agreement of all annotators.

### 3.2 The Identification of Lesions

To evaluate the performance of local network for lesion recognition, we record the *recall* and *precision* for each class of lesion in testing fundus images on Table 1. The second line of the table present the number of different types of lesions in testing set. The left and right values in the table denote *recall* and *precision* respectively. Two baseline algorithms are take into comparison: random forests [19] and support vector machine (SVM) [18]. We use the default setting with a *Python* toolkit named *Scikit-learn* (<http://scikit-learn.org/stable/>) except that the number of RF trees is turned from 10 to 500. As seen in the table, the proposed local network significantly outperforms the random forests and SVM

**Table 1.** Recall (the left values) and precision (the right values) of lesion recognition

| Lesion number | MA                   | Hemorrhage           | Exudate              |
|---------------|----------------------|----------------------|----------------------|
|               | 1538                 | 3717                 | 1248                 |
| Local network | <b>0.7029/0.5678</b> | <b>0.8426/0.7445</b> | <b>0.9079/0.8380</b> |
| Random forest | 0.0078/0.06704       | 0.2754/0.1011        | 0.6859/0.1941        |
| SVM           | 0.4153/0.0251        | 0.0108/0.0548        | 0.0787/0.0318        |



**Fig. 2.** (a) Lesion confusion matrix. The value of  $(i, j)$ -th entry of the matrix denotes the number of class  $i$  patches with prediction as class  $j$ . Wherein,  $i, j \in \{0, 1, 2, 3\}$  according to the first and second axes respectively. (b) ROC curve (shown in red) of the proposed algorithm over lesion detection. The black diamonds on the curve indicate the sensitivity and specificity of our lesion detection algorithm on high-sensitivity and high-specificity operating points. The green and blue dots present the performance of two trained human observers on binary lesion detection on the same testing dataset.

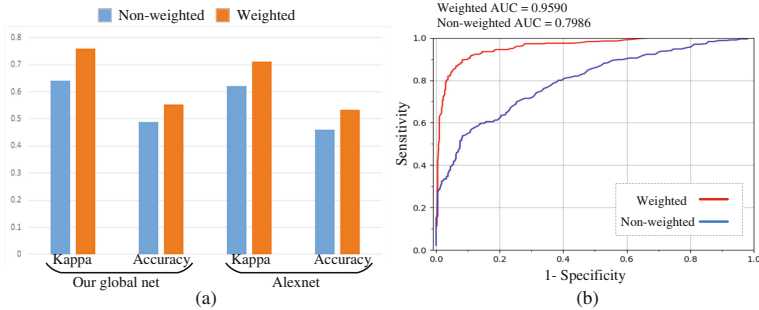
under same training images, which indicate the powerful ability of DCNN in learning task-driven features. In addition, we also shown the confusion matrix for lesion recognition in Fig. 2(a).

To show the importance of local net in finding retina lesions, we also train a binary classifier to distinguish the lesion patches from normal ones in the testing set. Receiver operating characteristics (ROC) curve is drawn with *sensitivity* and *specificity* in Fig. 2(b), and the value of area under curve (AUC) is 0.9687. The black diamonds on the red curve highlight the performance of the proposed algorithm at high-specificity (*sensitivity* : 0.863, *specificity* : 0.973) and high-sensitivity points (*sensitivity* : 0.959, *specificity* : 0.898). The green and blue dots correspond to the performance of two trained observers on binary lesion detection. As shown in the figure, the proposed algorithm can achieve superior performance than the trained observers by setting proper operating points.

### 3.3 Grading the DR Severity of Fundus Images

In this paper, we focus on the grading on NPDR, which can be classified into 0 to 3 stages: normal, mild, moderate and severe respectively. To prove the

importance of the proposed weighting scheme, we compare the *Kappa score* and *Accuracy* of grading networks with and without weighting (*non-weighted* for simplification) scheme under the same implementation setup. The results are shown in Fig. 3(a), and we also list the performance of the popular AlexNet under the condition of weighted and non-weighted scheme. As seen in Fig. 3(a), the proposed global grading net with weighting scheme achieve 0.767 *Kappa score*, and both the grading DCNNs achieve superior results with weighted lesion map, which prove the effectiveness of the proposed weighting scheme.



**Fig. 3.** (a) Illustration of importance of the proposed imbalanced weighting scheme on NPDR grading. (b) ROC curves of our global grading network in referable DR detection under weighted (red) and non-weighted (blue) conditions.

Since the symptom of some milder DR are too un conspicuous to be spotted, the judgements of milder DR are not easy to be unified even among licensed ophthalmologists. Therefore, similar with [5], we also train our global net to distinguish referable DR from normal images. The ROC curves on *sensitivity* and *1-specificity* are illustrated in Fig. 3(b). The performance of referable DR detection with weighted scheme is shown in red, and the AUC of the proposed algorithm is 0.9590. On the other side, the performance of the same network under non-weighted scheme is shown in blue, and the corresponding AUC is 0.7986. The results further prove the superior of the proposed imbalanced weighting scheme against our end-to-end grading net.

## 4 Conclusion

In this paper, we proposed two-stages DCNN to detect abnormal lesions and severity grades of DR in fundus images. The experimental results have shown the effectiveness of the proposed algorithm, and this study can provide valuable information for clinical ophthalmologists in DR examination. However, there still exist limitations need to be solved in our future work, such as collecting more high quality annotated fundus data, and paying attention to more types of lesions. Moreover, diabetic macular edema is also an import open issue needed to be addressed.

## References

1. Shaw, J.E., Sicree, R.A., Zimmet, P.Z.: Global estimates of the prevalence of diabetes for 2010 and 2030. *Diabetes Res. Clin. Pract.* **87**(1), 4–14 (2010)
2. Pratt, H., Coenen, F., Broadbent, D.M.: Convolutional neural networks for diabetic retinopathy. *Procedia Comput. Sci.* **90**, 200–205 (2016)
3. Bhaskaranand, M., Cuadros, J., Ramachandra, C., et al.: EyeArt + EyePACS: automated retinal image analysis for diabetic retinopathy screening in a telemedicine system. In: OMIA (2015)
4. Haloi, M.: Improved microaneurysm detection using deep neural networks. arXiv preprint (2015). [arXiv:1505.04424v2](https://arxiv.org/abs/1505.04424v2)
5. Gulshan, V., Peng, L., Coram, M., et al.: Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *J. Am. Med. Assoc.* **316**(22), 2402–2410 (2016)
6. Kocur, I., Resnikoff, S.: Visual impairment and blindness in Europe and their prevention. *Br. J. Ophthalmol.* **86**(7), 716C722 (2002)
7. van Grinsven, M.J., van Ginneken, B., Hoyng, C.B., et al.: Fast convolutional neural network training using selective data sampling: application to hemorrhage detection in color fundus images. *IEEE Trans. Med. Imaging* **35**(5), 1273–1284 (2016)
8. Seoud, L., Chelbi, J., Cheriet, F.: Automatic grading of diabetic retinopathy on a public database. In: OMIA (2015)
9. American Academy of Ophthalmology. International Clinical Diabetic Retinopathy Disease Severity Scale (2012). <http://www.icoph.org/dynamic/attachments/resources/diabetic-retinopathy-detail.pdf>
10. Haloi, M., Dandapat, S., Sinha, R.: A Gaussian scale space approach for exudates detection classification and severity prediction. arXiv preprint (2015). [arXiv:1505.00737](https://arxiv.org/abs/1505.00737)
11. Srivastava, R., Duan, L., Wong, D.W.K., et al.: Detecting retinal microaneurysms and hemorrhages with robustness to the presence of blood vessels. *Comput. Methods Programs Biomed.* **138**, 83–91 (2017)
12. Sankar, M., Batri, K., Parvathi, R.: Earliest diabetic retinopathy classification using deep convolution neural networks. *Int. J. Adv. Eng. Technol.* **7**, 466–470 (2016)
13. Gu, J., Wang, Z., Kuen, J., et al.: Recent advances in convolutional neural networks. arXiv preprint (2016). [arXiv:1512.07108v2](https://arxiv.org/abs/1512.07108v2)
14. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: NIPS (2012)
15. Szegedy, C., Liu, W., Jia, Y., et al.: Going deeper with convolutions. In: CVPR (2015)
16. He, K., Zhang, X., Ren, S., et al.: Deep residual learning for image recognition. In: CVPR (2016)
17. Bishop, C.M.: *Pattern Recognition and Machine Learning*. Information Science and Statistics. Springer, New York (2006)
18. Cortes, C., Vapnik, V.: Support-vector networks. *Mach. Learn.* **20**(3), 273–297 (1995)
19. Kam, H.T.: The random subspace method for constructing decision forests. *IEEE Trans. Pattern Anal. Mach. Intell.* **20**(8), 832–844 (1998)