DISCOVERING ABNORMAL PATCHES AND TRANSFORMATIONS OF DIABETICS RETINOPATHY IN BIG FUNDUS COLLECTIONS

Yuqian ZHOU[1] and Shuhao LU[2]

[1]Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Hong Kong, China
`yzhouas@ust.hk`
[2]Department of Mechanical and Aerospace Engineering, The Hong Kong University of Science and Technology, Hong Kong, China
`slu@ust.hk`

ABSTRACT

Diabetic retinopathy (DR) is one of the retinal diseases due to long-term effect of diabetes. Early detection for diabetic retinopathy is crucial since timely treatment can prevent progressive loss of vision. The most common diagnosis technique of diabetic retinopathy is to screen abnormalities through retinal fundus images by clinicians. However, limited number of well-trained clinicians increase the possibilities of misdiagnosing. In this work, we propose a big-data-driven automatic computer-aided diagnosing (CAD) system for diabetic retinopathy severity regression based on transfer learning, which starts from a deep convolutional neural network pre-trained on generic images, and adapts it to large-scale DR datasets. From images in the training set, we also automatically segment the abnormal patches with an occlusion test, and model the transformations and deterioration process of DR. Our results can be widely used for fast diagnosis of DR, medical education and public-level healthcare propagation.

KEYWORDS

Transfer Learning, Diabetic Retinopathy, Retinal Fundus Images, Computer-aided Diagnose, Big data analytics

1. INTRODUCTION

Diabetes is one of the most common diseases in the US affecting 86 million people, which is more than 1 out of 3 adults in America [1]. Diabetic retinopathy (DR) is a retinal damage caused by long-term effect of diabetes. It has been statistically shown that around 80% of diabetes patients who have had diabetes more than 20 years suffered from DR [2]. However, timely treatment is extremely beneficial and research indicated that early diagnosis resulted in at least 90% reduction of new patients [3]. Currently, the diagnostic screening is time-consuming and laborious. Diagnosis of DR mostly depends on manual inspection of retinal fundus images, hence

several weeks are often required for the test results. The need of expensive medical equipment and expertise also lead to a considerably high cost for population screening while regular screening of diabetic patients was considered as necessary. Consequentially, the development of automated DR detection is of vital importance to assist the clinicians for better diagnosis, and reduce the time and cost in diagnostic process.

Many studies have been reported in automated diagnosis of DR using different techniques. The diagnostic scheme of different classifications of DR depends on the weighting of various patterns and their locations [11]. Oktoeberza et al. reported 94.5% accuracy using image processing based on red channel of the retinal fundus images and optic disc segmentation detection [4] while Pratt et al. reported 75% correct detection using Convolutional Neural Networks (CNN) [5]. In terms of glaucoma and DR detection based on optic disc images of healthy and unhealthy subjects, Abdel-Ghafar et al. achieved 65% of accuracy [6]. Mookiah et al. proposed a DR detection with two class of DR and non-DR using SVM and kNN algorithms and 85% of accuracy was achieved [7]. Some other studies proposed various methodologies to process the fundus images including morphological features, color histogram and deep learning with accuracy ranging from 60% to 90% for different eye-related diseases [8, 9].

However, it is shown that automated detection using algorithms was no better than manual inspection since they are all trained on hand-labeled images. The utilization of automated detection can only simplify and cut down the time cost in diagnostic process. Besides this, special situations related to abnormal patterns still require the diagnosis from professional doctors. A kind of human-involved semi-supervised learning are also recommended to improve the accuracy [10]. Therefore, in addition to improve the detection accuracy, discovering and analyzing the characteristics and mechanisms, as well as the diagnosis techniques widely utilized by clinicians is another meaningful topic.

In this paper, we treat the DR detection problem as a regression problem instead of widely used classification, by regressing the severity degree using discrete labels from a web-based large-scale eye fundus dataset. We proposed a big-data-driven automatic computer-aided diagnosing (CAD) system based on transfer learning, starting from AlexNet [15] pre-trained on generic images, and retraining the custom regressor. In addition to achieving a satisfactory diagnosing result, we are also interested in the following important questions.

First, what is the region of one specific eye image the clinicians always focus on when they diagnosed? Can we reveal the fact by big data analytics? We show those regions by an occlusion test automatically, and provide the reference and advices for young doctors, as the experience from professional clinicians.

Second, are there any characteristics of abnormal patterns in different stages of DR? Segmentation of the abnormal patches exists in large-scale database based on occlusion test. This will help the clinicians focus more on distinctive DR cases, which also improves their diagnosis skills.

Finally, we model the transformations and deterioration process of DR in a continuous way by ranking the regressive scores. In this case, the development process of DR is revealed and help doctors and the public understand DR.

2. METHODOLOGY

A large-scale dataset was used in this study. Abnormal pattern segmentations were carefully detected using transferred regressive convolutional neural network. At the end, a reasonable transformation model of DR was achieved. The details are covered as follows.

2.1. Dataset

The database we use in this paper is provided by the kaggle competition, where large-scale and high-resolution retinal images are given by an online diagnosing website called EyePACS [13]. This organization aims at matching patients potentially suffering from vision impairment from diabetic retinopathy with professional clinical doctors, by allowing the patient to upload their own eye fundus images. Some Samples in this database are shown in Figure 1. The images are labeled in five stages: Normal (0), Mild (1), Moderate (2), Severe (3) and Proliferative DR (4) according to the severity of illness. All the labels are given by the professional doctors. In total, there are 35,126 images of left and right eyes in the training set provided. However, those images are from different devices of distinct clinical organizations, and only one rater exists for one image. Therefore, noisy data requires more preprocessing and filtering before staring the training process.
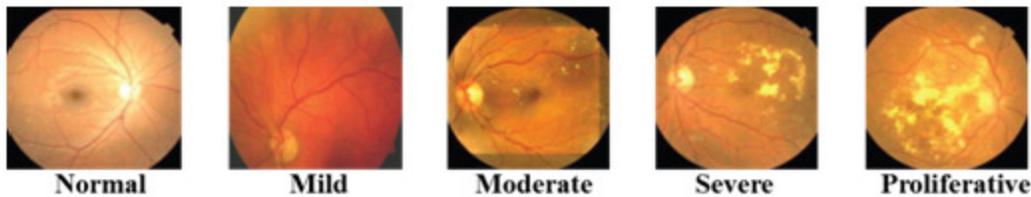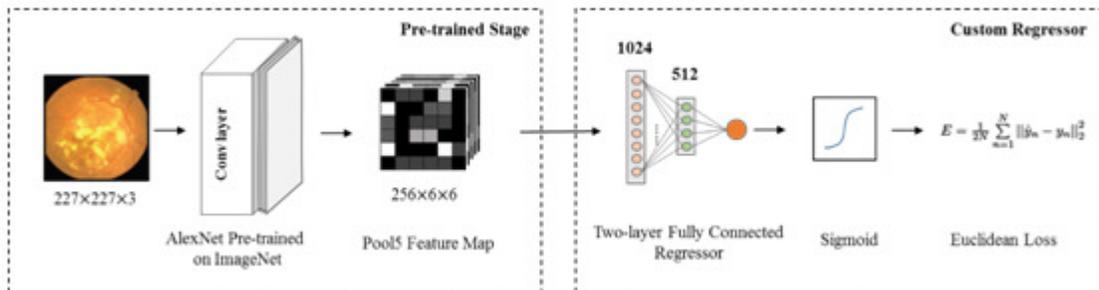

Figure 1. Some samples in the EyePACS dataset

Figure 2. System architecture of our transferred regressive CNN

2.2. Transferred Regressive Convolutional Neural Network

In the previously reported result, the whole network is designed and trained from scratch. This may yield over-fit if the training set are not large enough. However, it is shown that reusing the bottom layers of network trained on generic images will facilitate the development process of vision system, and benefit from the rich mid-level features for images in the new dataset.

The method we model the order of severity of Diabetics Retinopathy is using a regressive convolutional neural network (RegCNN). In this paper, we propose a transfer learning model, which utilizes the five bottom layers of a revised version of AlexNet [15], and connect a custom two-layer regressor with a nonlinear sigmoid function, regressing the severity according to the diagnostic labels assigned by the professional doctors. The sigmoid function is used to restrict the regression value inside a specific interval from 0 to 1. The custom regressor is fine-tuned in a fully supervised way using discrete labels, but it regresses to continuous labels. To fully utilize the high-level feature maps in pooling layer five, we do not fine-tune the bottom layers. We then utilize the Euclidean loss to train the whole network, which is defined as,

$$E = \frac{1}{2N} \sum_{n=1}^{N} \|\widehat{y_n} - y_n\|_2^2 \tag{1}$$

where N is the number of image samples, and $\widehat{y_n}$ is the estimated label, y_n is the ground truth label.

The overall structure of the network is shown in Figure 2. The regressed labels are then further used to validate the reliability of the system and model the illness severity transformations using all the training images.
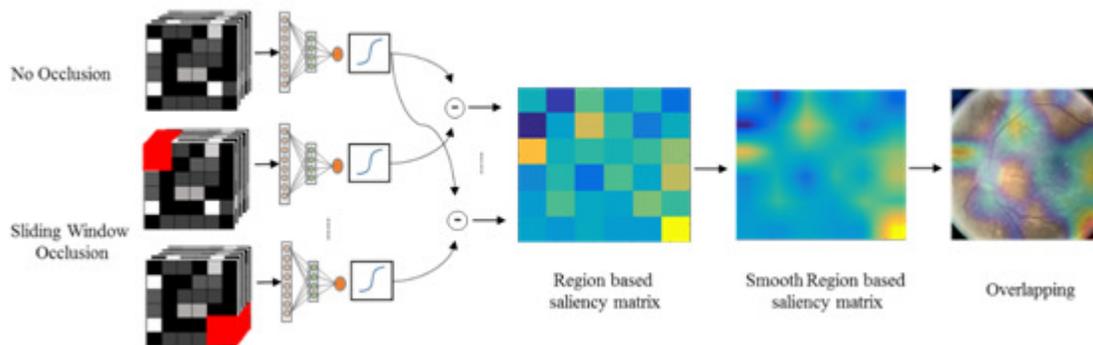
2.3. Abnormal pattern segmentation

Figure 3. The procedure of occlusion test.

Clinicians diagnose DR by checking and finding the abnormal patterns, like swelling vessels, proliferative tissues etc. We want to figure out which regions inside the eye fundus images actually help clinicians label the illness stages. Those abnormal patches contain important pattern closely related to the diagnoses. Segmenting them out automatically will highlight the critical patterns of diagnosing, assist the training process and benefit the medical education. Also, feedback given by the computer will help the clinicians prevent misdiagnosis, and mutually improve the diagnosing accuracy through human-computer interaction.

We demonstrate an occlusion test for segmentation tasks. In pooling layer five of AlexNet, the size of the feature maps is six by six, segmenting the whole pixel space into 36 small patches. We slide a window and zero out the enclosed patch across all the feature maps in pooling layer five, and monitor the regression value dropping. If the regression value of this image drops dramatically, we will regard the patch as an abnormal one, and segment it out. The process is similar in spirit to calculate the jacobian matrix with respect to neurons in each position. High

values inside the matrix mean that severity regression value is more sensitive to the patterns enclosed in this patch. Without the patterns in this patch, the patient will not be that severe since the severity degree regressed will drop. Finally, we conduct a k-means clustering on these patches and organize the typical pattern for clinical education.

We further compute the salience abnormal region for each input sample by using the above method with window size one. Patches with larger regression value dropping will be more salient and important for the model to regress the severity. The occlusion process is illustrated in Figure 3. The salient region helps the clinicians focus on the abnormal pattern localized and searched automatically by the CAD system, and verity the reliability of the diagnosis given by the computer. It will make the diagnosis process faster and more accurate.

2.4. Transformation Modelling

With the continuous regression value corresponding to the severity of DR, we want to discover the transformations and deteriorating process in the large-scale eye fundus images database. The discovery will benefit in two aspects: First, it helps the clinicians discover the intermediate states between each discrete stage label. Secondly, it helps people understand the big picture of disease transformations and will benefits the public-level healthcare propagation. In this paper, we rank and distribute the DR disease cases in the Computer-given space according to the regression value.
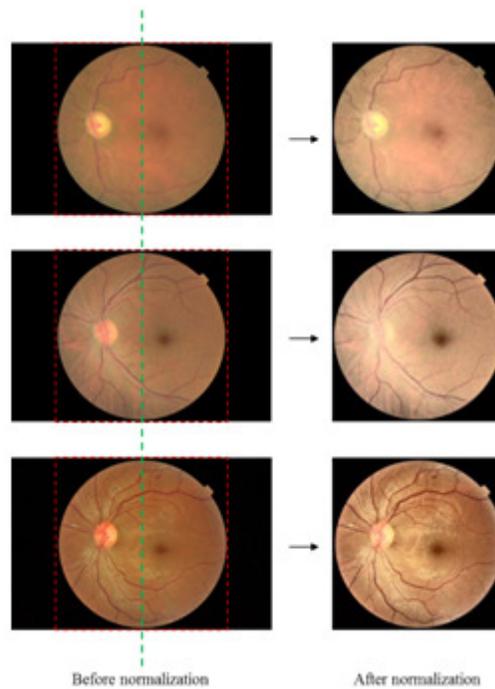
3. EXPERIMENT

Figure 4. Some examples of preprocessing result

All fundus images need to be pre-processed to eliminate the noises from both data and labels. A supervised model then was trained for regression. Besides this, occlusion test was conducted and abnormal segmentation can be found. Finally, transformations of DR were discovered. The details are illustrated as follows.

3.1. Data pre-processing

The problem of web-based database comes from plenty of noisy data and label. In the EyePACS dataset, images are collected in a crowdsourcing way, thus there is no standard for devices and illumination before collection. Normalization processing should be conducted. There are mainly three problems of the original eye fundus images, which are color variance, illumination variance, and scaling variance. Flipping and rotation do not influence the diagnosis result, thus we do not consider normalizing then in image directions.

In this case, we preprocessed the images in center cropping, color normalization, and contrast normalization. The whole normalization process is illustrated in Figure 4. For cropping, we first make the height of original images identical to each other, and crop the square regions around the center line (the blue lines). After processing, the eye fundus images seem uniform and normalized.

3.2. Regression Experiment

We want to train the model in a supervised way using the discrete labels assigned. However, there are some problems when using those labels. First, the labels are ranged from 0 to 4, but sigmoid function will only generate the values between 0 and 1. Thus, we need rescaling all the labels before training. Secondly, the samples are unbalanced across all the labels. Normal samples are far more than proliferative samples. Although the distribution somehow shows the probability of occurrence of those stages in real life, it will make the system biased to lower value. The number of samples of each label is shown in Figure 5. Finally, we redistribute the samples number by random data selection and discarding many samples. The number of samples for each stage in the whole dataset becomes normal (3000), mild (2443), moderate (2000), severe (873) and proliferative (708). Then the set is split into training set, validation set and testing set, in 80%, 10% and 10%. We stop training when the model obtains the minimum validation loss. The training process is conducted with two slaves in Amazon web platform to speed up the stochastic gradient descent. An alternative way may be data augmentation through rotation and flipping to balance the sample number.

Intuitively, regression model is more meaningful since the loss is based upon the severity, so that the loss between a proliferative sample and a normal one is larger than the loss between two closed illness stages. However, for classification model, the loss will be the same between each pair of classes. Thus, for testing result, regression model will show a clustering pattern around each value, which is more meaningful than classification, i.e. classifying a normal sample to proliferative one with only tiny probability difference. To validate the intuition, we also train a classification model using the same features and bottom layers as the regression model, where the loss function is replaced by cross entropy of softmax function. We compare the regression and classification result distribution for each class. For the classification mode, we directly got the discrete distribution through the confusion matrix. For the regression model, we use a kernel

smoothing function to estimate the distribution of regression value after scaling back for each labeled class. The result is shown in Figure 6.
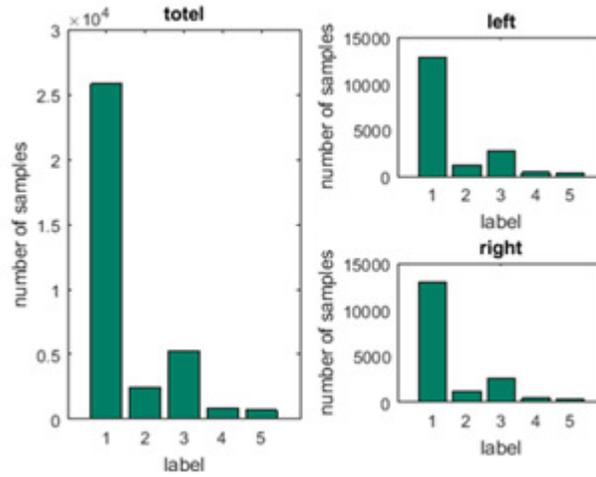

Figure 5. The label distribution in the original dataset.

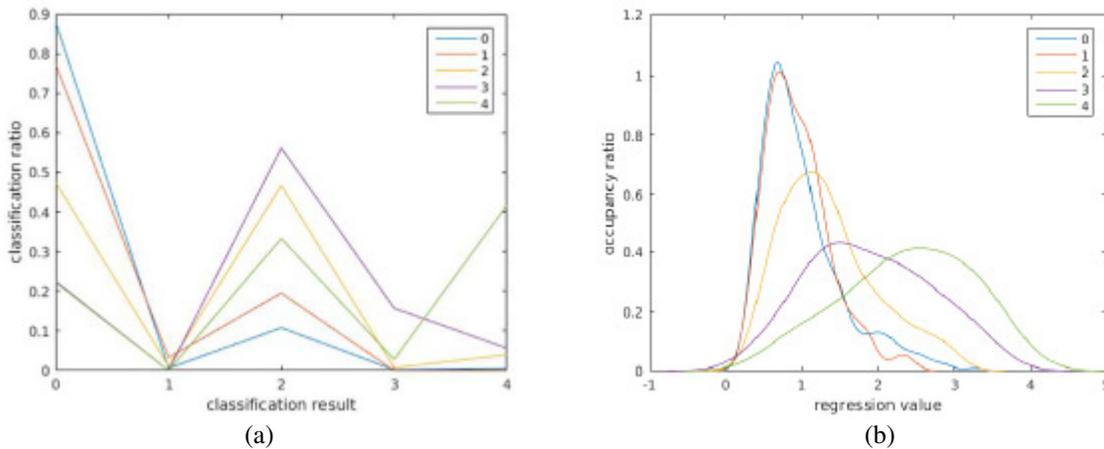(a)   (b)

Figure 6. The result distribution on the testing set of two models.

As shown, in the classification mode, it achieves a 61% accuracy and normal cases contribute a lot to the accuracy. And the detection may skip stages, i.e. classifying proliferative one to normal one, which is unacceptable. However, for the estimated distribution of regression value for the regression model, it demonstrates a more meaningful distribution. Most of the cases can be differentiated according to the severity. And the regression value for each labeled class shows a normal-distribution-like pattern both in the training and testing set.

Because of the characteristics of sigmoid function, values cannot be regressed to exact zero or one, or exceed this range, thus, the mean of the regression values of proliferative samples may not be able to fall on value 4, instead, shift to the left around 3. To compare the result with other classification models and build a detector, we consider the similar distribution in the training set, and split the range of 0 to 4 into five stages, avoiding overlapping. The method to achieve the

non-overlapped distribution is starting from the estimated peak value, and extending to both sides according to the standard deviation (SD). We then use this separation to classify the testing set. The result of range separation is illustrated in Figure 7. The testing result and one-vs-all ROC curve is shown in Figure 8.
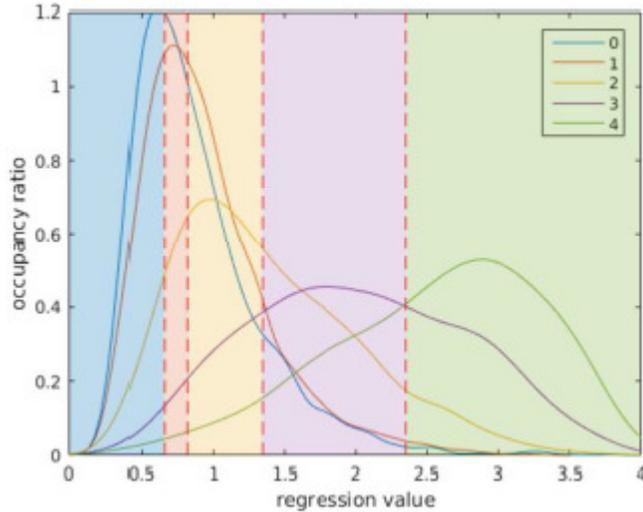

Figure 7. The testing value range of regression.
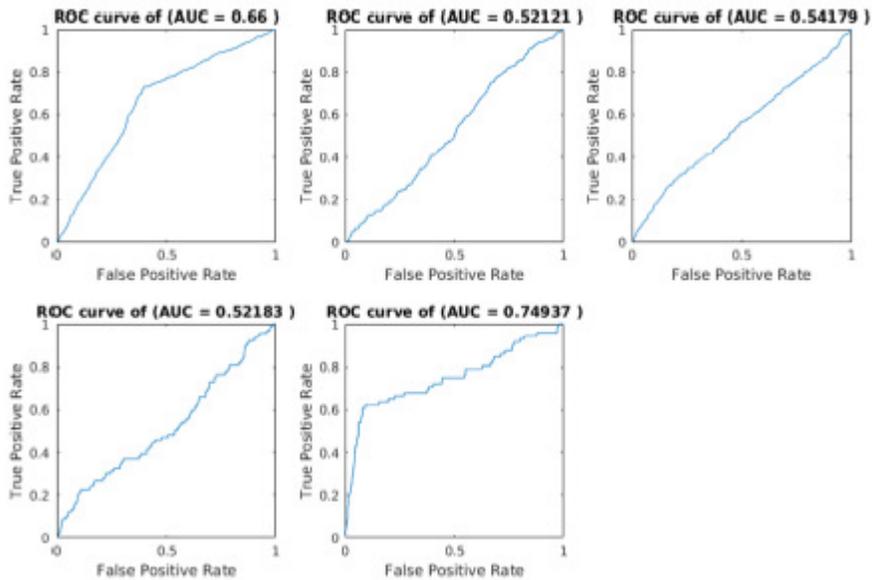

Figure 8. The ROC curve of the one-vs-all classification result.
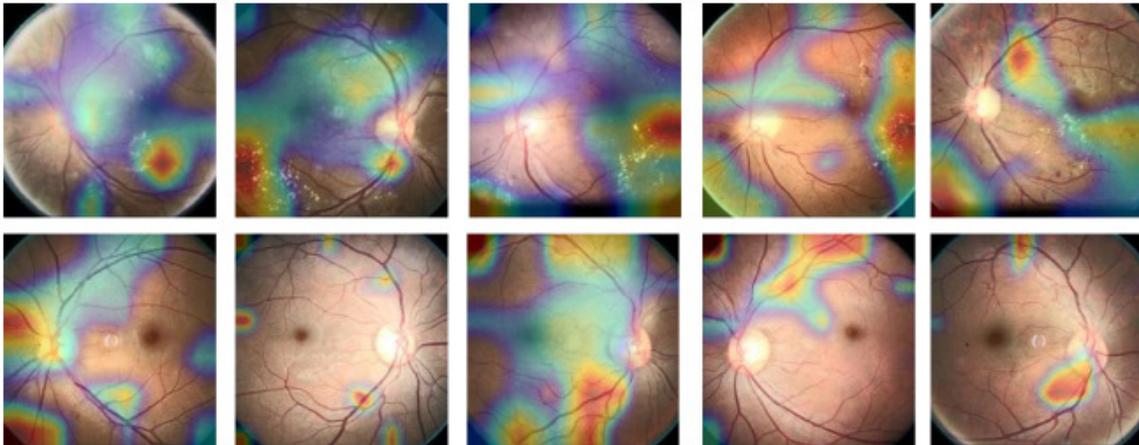
Figure 9. Some DR focus examples detected by occlusion test.

Figure 10. The clustering result and some samples of abnormal patch segmented by occlusion test.

However, the testing result is not satisfactory enough for an accurate detector. It mostly results from noises in original images and labels, and large overlapping region among classes after regression. As shown in Figure 7, the overlapping region between two classes are too large, making it hard for this model to differentiate two distinct stages. But it is still better than classification model as a binary detector (one-vs-all). Better regressive model still should be explored and developed in the future work.

3.3. Occlusion test and abnormal segmentation

We ran the occlusion test with sliding window size setting to one to generate the saliency map. Some random selected examples of the saliency map generated for each class are shown in Figure 9. As shown, images of late period of DR (the first row) will show patterns like yellow spot or proliferative tissue shape. However, images from early period of DR (the second row) only show the swelling or blocked vessels. The occlusion test is able to detect and focus on those abnormal patterns, and judge the severity. We can conclude that our regression model is using the same mechanism as the professional clinicians.

Furthermore, we segmented the abnormal patches by running the occlusion test with sliding window size setting to two. And the dropping difference threshold of segmentation is 0.2. After the test, we get totally 2772 patches. We extract the deep features using AlexNet again, and use k-means to cluster them with k=20. Finally, we visualize the segmentation using t-SNE [13] method according to the pairwise distance. The visualization is shown in Figure 10. In the first two columns, the 20 clusters are illustrated. From the graph, we found that similar pattern will be clustered together. In the third column, some segmented abnormal patches for each stage are shown. From our result, the segmented patches for one specific stage are from some fixed clusters. For the late period, the patches will be more related to proliferative tissue and yellow spot. For the early stages, the patches are related to vessels. Such segmented patches can be further used for medical education.
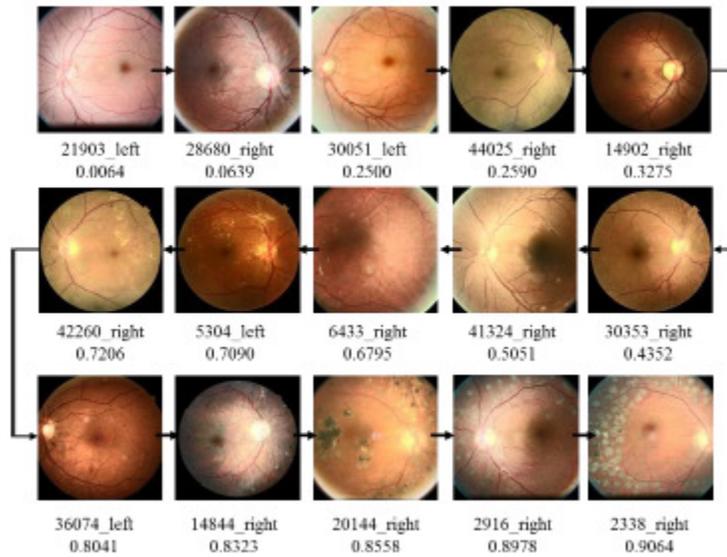

Figure 11. The transformations we model according to the regression scores.

3.4. Transformations

Finally, we discovery the transformations of DR by the ranking of the regression scores in the training set. From the ranking, we randomly select the fundus images and plot them in order of increasing regression scores. The transformation is shown in Figure 11. Through big-data-driven analysis, the system is able to regress and rank the stages of DR automatically, and we can find the deteriorating process from normal eyes in a more detailed and continuous way. The organized data will help the clinicians clearly understand the development stages of DR, and diagnose more accurately.

4. CONCLUSIONS

In this paper, we proposed a big-data-driven automatic diagnosis system based on transfer learning of deep CNN. Instead of doing classification directly, our system focuses on regressing the severity of diabetic retinopathy. Based on the trained model, we conducted an occlusion test to search for the abnormal patches in the eye fundus images, and segmented them out for medical

education. We revealed the mechanism of diagnosing DR utilized by professional clinicians and computer-aided system by showing the saliency map obtained from occlusion test. Furthermore, according to the regression value, we modeled the transformations and deteriorating process of DR. Our system will benefit efficient and accurate self-inspection for patients, and contribute to medical education and public-level healthcare propagation by assisting the learning process of young clinicians.

ACKNOWLEDGEMENTS

REFERENCES

[1]   C. f. D. Control and Prevention, "National diabetes statistics report: estimates of diabetes and its burden in the United States, 2014," Atlanta, GA: US Department of Health and Human Services, vol. 2014, 2014.

[2]   P. J. Kertes and T. M. Johnson, Evidence-based eye care: Lippincott Williams & Wilkins, 2007.

[3]   R. J. Tapp, J. E. Shaw, C. A. Harper, M. P. De Courten, B. Balkau, D. J. McCarty, et al., "The prevalence of and factors associated with diabetic retinopathy in the Australian population," Diabetes care, vol. 26, pp. 1731-1737, 2003.

[4]   K. W. Oktoeberza, H. A. Nugroho, and T. B. Adji, "Optic Disc Segmentation Based on Red Channel Retinal Fundus Images," in International Conference on Soft Computing, Intelligence Systems, and Information Technology, 2015, pp. 348-359.

[5]   H. Pratt, F. Coenen, D. M. Broadbent, S. P. Harding, and Y. Zheng, "Convolutional Neural Networks for Diabetic Retinopathy," Procedia Computer Science, vol. 90, pp. 200-205, 2016.

[6]   R. Abdel-Ghafar and T. Morris, "Progress towards automated detection and characterization of the optic disc in glaucoma and diabetic retinopathy," Medical informatics and the Internet in medicine, vol. 32, pp. 19-25, 2007.

[7]   M. R. K. Mookiah, U. R. Acharya, C. K. Chua, C. M. Lim, E. Ng, and A. Laude, "Computer-aided diagnosis of diabetic retinopathy: A review," Computers in biology and medicine, vol. 43, pp. 2136-2155, 2013.

[8]   J. Nayak, R. Acharya, P. S. Bhat, N. Shetty, and T.-C. Lim, "Automated diagnosis of glaucoma using digital fundus images," Journal of medical systems, vol. 33, pp. 337-346, 2009.

[9]   U. R. Acharya, S. Dua, X. Du, and C. K. Chua, "Automated diagnosis of glaucoma using texture and higher order spectra features," IEEE Transactions on information technology in biomedicine, vol. 15, pp. 449-455, 2011.

[10] M. J. Greaney, D. C. Hoffman, D. F. Garway-Heath, M. Nakla, A. L. Coleman, and J. Caprioli, "Comparison of optic nerve imaging methods to distinguish normal eyes from those with glaucoma," Investigative ophthalmology & visual science, vol. 43, pp. 140-145, 2002.

[11] E. T. D. R. S. R. Group, "Grading diabetic retinopathy from stereoscopic color fundus photographs—an extension of the modified Airlie House classification: ETDRS report number 10," Ophthalmology, vol. 98, pp. 786-806, 1991.

[12] R. Raina, A. Battle, H. Lee, B. Packer, and A. Y. Ng, "Self-taught learning: transfer learning from unlabeled data," in Proceedings of the 24th international conference on Machine learning, 2007, pp. 759-766.

[13] EyePACS Dataset. Available: https://www.kaggle.com/c/diabetic-retinopathy-detection/data.

[14] L. v. d. Maaten and G. Hinton, "Visualizing data using t-SNE," Journal of Machine Learning Research, vol. 9, pp. 2579-2605, 2008.

[15] A. Krizhevsky, I. Sutskever and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in Advances in Neural Information Processing Systems, 2012, pp. 1097-1105.

AUTHORS

Yuqian ZHOU, M.phil candidate of HKUST. His research focuses on affective computing based on deep vision, pattern recognition, machine learning and medical image processing.

Shuhao LU, Ph.D candidate of HKUST. His research focuses on Biomechanics, Ophthalmology, and medical sensor.