AUTOMATIC IMAGE GRADING FOR DIABETIC RETINOPATHY(DR) AND DIABETIC MACULAR EDEMA(DME)

Junyan Wu, Xiaolong Li, Ting Zhou, Yu Wang

Cleerly, Virginia Tech, University at Buffalo

ABSTRACT

Diabetes mellitus is one of leading cause of vision loss globally. Diabates increase the risk of a range of eye diseases, but the main cause of blindness associated with diabates is diabetic retinopathy(DR). Diabetic Macular Edema(DME) is a complication associated with DR of which retinal thickening or accumulation of fluid can occur at any stage. Early detection for DR and DME is important to save the vision of diabetes patients. In this paper, we propose an ensemble pipline which combines convolutional neural network and feature blending for automatically predicting the DR and DME severity scores. It outperforms the usual end-to-end convolutional neural network method.

Index Terms— Diabetic Retinopathy, Diabetic Macular Edema, DR, DME, CNN

1. INTRODUCTION

The WHO(world health organization) estimates that by the end of 2025, the number of people having diabetes mellitus worldwide will increase to 300 million.

People with long term diabetes is very possible to have an eye disease called diabetic retinopathy, which is triggered by high blood sugar levels. Diabetic retinopathy can cause vision loss in two ways: high blood sugar levels will damage blood vessels in the retina, and in turn cause the blood, fluid leak, as well as the swelling of the retina; sometimes abnormal new blood vessels will grow on the retina, which will increase the pressure within theye. All of these physics changes in the eye can destroy people's vision. Findings of venous beading, retinal neovascularization can be utilized to classify DR severity in one of the two phases known as non-proliferative diabetic retinopathy (NPDR) which is early stage of diabetic eye disease and proliferative diabetic retinopathy (PDR) which is more advanced stage of diabetic eye disease, as shown in Figure 1a and 1b. Having NPDR would make sight blurry, when it reaches to the phase of PDR, then human's central and side vision can be destroyed.

DME occurs when fluid and protein deposit collect on or under the macula of the eye and cause it to thicken and swell. The swelling may distort a persons central vision. The risk of having DME is classified into no risk and two probable risks(Figure 1c,1d respectively) based on the location of hard exudates. In this paper, we will introduce an integrated framework that can automatically grade image for DR and DME.

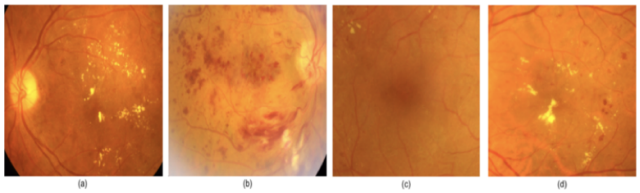

Fig. 1. DR and DME

2. RELATED WORK

Traditional automated DR detection approaches using handcrafted feature dominated the field of DR detection for many years. Given colorful fundus image, these methods extract visual features from the photography on parts of optic disc and blood vessels[1], [2]. Those generic feature extraction methods developed in either computer vision area or statistical community were widely used. With those features at hand, a classification algorithm like KNN or SVM, random forest will be used to identify and localize interested regions like exudates and hemorrhages. Although these methods have been showing progress in automatic diagnosis, the accuracy is still not robust and good enough to be widely accepted and applied. Recently, deep learning is boosting and driving multiple areas, including biomedical imaging research, and some most recent research is highly focused on applying deep ConvNets into DR diagnosis, like [3] in 2016,[4] in 2017. Usually researchers will adopt standard architecture like AlexNet and GoogleNet[5] to build their deep learning model, and some of them are trying to find simplified architecture like [6], [7], or combine both traditional machine learning algorithms and deep learning together to get better performance [8]. However, the fine-tuning or training from scratch on biomedical image data is more difficult compared to those general computer vision tasks, which is partly due to the data variance issues. CNN approaches without sufficient preprocessing of images or large amount of parameters in neural network structures, will sometimes struggle to achieve good performance.

However, when the network is large and deep, it will bring difficulties to training.

In this paper, we propose a unified framework that combines deep feature extractor and statistical feature blending, which could improve the performance to a higher degree in grading DR and DME.

3. METHODS

3.1. Preprocessing

The microaneurysms (MA) appears as dark tiny spots and the hard exudates (EX) appears as bright yellow regions in retinopathic images. Especially for EXs, they have high contrast with the background and usually occur as a continuous space, but the MAs always spark in the images sometimes you even cannot find them. The morphological transformation can be used to highlight or eliminates these lesions. For example, the morphological opening can erase the EXs and highlight the MAs. The closing operation can remove MAs and preserve EXs as shown in the Figures. These operations can be used to denoise specific levels of classifications. For example, the risk of Diabetic Macular Edema (DME) only depends on the location of the hard exudates. Then we can implement morphological closing to mainly preserve the EXs and we call this as a bright preserving operation. We only use dark preserving and bright preserving in image preprocessing, because implementing these methods might bring the risks of eliminating some valuable features of the original images. The dark preserving, bright preserving and original image are shown in 2 by the order of a, b, c.

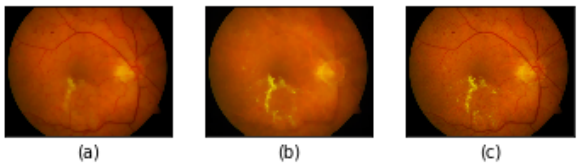

Fig. 2. image preprocessing

3.2. Convolutional Neural Network

Our pipeline adapts the architecture from [9] to solve DR and DMA grading as an image classification problem. Specifically, we modify and fine-tune DenseNet architecture to embed the image data, provide feature layer for further feature blending, while also predict severity score directly as the baseline.

Densenet yields the state of the art in many image classification tasks. It is a network architecture where each layer is directly connected to every other layer in a feed-forward fashion 3. For each layer, the feature maps of all preceding layers are treated as separate inputs whereas its own feature maps are passed on as inputs to all subsequent layers. This design first sloves the vanishing gradient problem when training deep neural work. It also strengthens the feature propagation through the layers, which would be quite helpful when extracting invariant and fine-grained features in DR and DME grading.

Since we use 512*512 size of image as input, which is larger than the standard ImageNet image size, the output size from second final layer of densenet would be very huge, it would be computationally cheaper if we reduces the size of layer followed. We replace the last fully connected layer into 3 fully connected layers and 2 dropout layers. This could let network be fully trained and provide a more condense feature layer for feature blending.

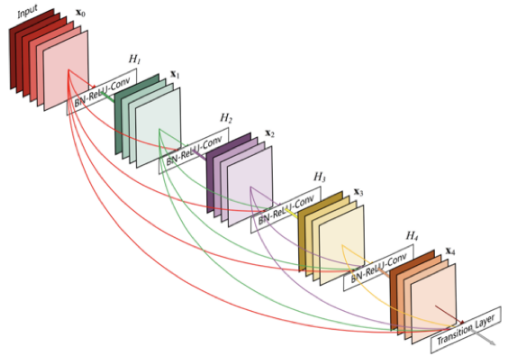

Fig. 3. DenseNet architecture

3.3. Feature Blending

The key to achieve good performance on the disease grading task is robust and invariant features at multiple scale. From our observation, those well-developed networks like VGG16[10] and DenseNet are still struggling with the final accuracy on DR prediction. However, they could serve as very good feature extractors compared to previous methods with hand-crafted features. So we adopt a blending strategy which originally comes from [11]. Feature blending contains two steps. The first step is to pass image data with different augmentation to generate multiple features vectors, and get high level statistical features like mean and standard deviation for every original image. Then we have proposed two main machine learning techniques for the final classification based on the statistical features, which are demonstrated to generate better results for to enhance the results.

One method is gradient boosting tree algorithm[12], which one of the most popular machine learning techniques these days. Boosted trees try to learn a more and more complex decision boundary by successively fitting trees on the error with a low learning rate over hundreds of trees. It benefits from the decision tree algorithm that can learn complex

non-linear decision boundaries. Another technique is a simple fully connected network with light-weight architecture as introduced in section 5. We separately utilize these two methods to build classification models upon the extracted features. The final pipeline is shown in Fig 4.

4. EXPERIMENTS

4.1. Datasets

In our experiment, we use dataset provided by grand-challenge website. The set of color retina images contains 516 images in total and splits into 413 training data and 103 testing data, and labels are provided by clinicians who rated the presence of diabetic retinopathy in each image by a scale of 0,1,2,3,4, which represent No apparent retinopathy, Mild, Moderate, Severe, Proliferative respectively, also diabetic macular edema in each image by a scale of 0,1,2, which represent No apparent hard exudate(s) (HE), outside the radius of disc diameter, inside the radius of disc diameter respectively.

In addition, we use the dataset from Kaggle Diabetic Retinopathy Detection competition. It contains total 35126 high-resolution retina images, including 25810 no DR images, 2443 mild DR images, 5292 Moderate DR images, 873 severe DR images and 708 Proliferative DR images. We use this dataset to enlarge our training dataset and setup a benchmark for our initial training.

4.2. Preprocessing

We crop away the background of all our training and testing images and resize them into the squares of 512. We apply bright preserve and dark preserve filtering method as introduced previously in some of our experiments.

4.3. Augmentation

Common image augmentation techniques have been applied like translation, rotation, stretching, flipping and color augmentation at each image training process.

4.4. Training

We consider the 121 layers DenseNet[13] as our benchmark method to directly predict DME severity score. For the initial training, we only train on Kaggle datasets to predict DR gradings. Since the dataset is highly unbalanced, if trained with general random sampling, the model will see more instances of certain categories instead of a balanced training instances. We implemented the dynamic re-sampling method as used in [11],[14] to let the network train on fully balanced dataset at beginning and then gradually decrease the amount of rare classes. As shown below:

$$\boldsymbol{w_t = r^{t-1} w_0 + (1 - r^{t-1}) w_f, i = 0, 1, 2, ..., k} \quad (1)$$

w_0 stands for the initial resampling weights, w_f is the final resampling weights the model tries to approxmate. w_0 is set to a vector containing the reciprocal of ratio of each class in Kaggle dataset, while the w_f is set to $(1, 2, 2, 2, 2)$. Note the class '0', which means no DR, is the dominant class in the Kaggle dataset. Later we use the pre-trained weights to fine tune on challenge datasets. And we try fine-tuning our network on MSE loss and weighted cross-entropy loss since the re-sampling strategy only brings small performance gain. The bright preserve and dark preserve are also applied separately in our fine tune process to obtain the best performance.

The initial learning rate is 0.0005 and it will decrease 0.1 times every after 30 epochs. The initial training has been done by 200 epochs and fine tuning training will be finished at least after 50 epochs. We use cross-entropy loss and stochastic gradient descent for optimization.

4.5. Feature Extraction

We select several top performance Densenet weights to extract features. We use 50 pseudo random augmentations to get 50 outputs from last second fully connect layer (size of 4096). And we also calculate the mean and standard deviation of 50 feature vectors for each image, and the mean and standard deviation vectors are then concatenated together.

4.6. Feature Blending

We design two experiments training over the input of feature vectors. For the first experiment, we use LightGBM as boosting tree training tool to train our feature vectors. We use dart as our boosting type. And we set learning rate as 0.02. As for the second experiment, we adopt a 5-layer fully connected neural network with Leaky ReLU(0.01) as the activation function, we use both SGD and Adam optimizer to train the network and find Adam is more powerful, the starting learning rate for Adam is set to 0.0001. The architecture of Blendnet is shown in Table 1.

5. RESULTS

The initial training using Kaggle datasets can achieve the overall accuracy of 0.81 for DR grading prediction. However the high accuracy occurs mainly on the majority classes 5. We consider the model we gained from this training as pre-trained network.

Later, we directly apply the pre-trained weights on the challenge's official datasets. The accuracy for DR grading prediction is only 0.73. And the accuracy for DME grading prediction is only 0.77.

We use the output from second last layer of fine-tuning experiments to train a blending model. For DME grading prediction, we got best accuracy of 0.843 by building gradient boosting tree model on combining second last layer from

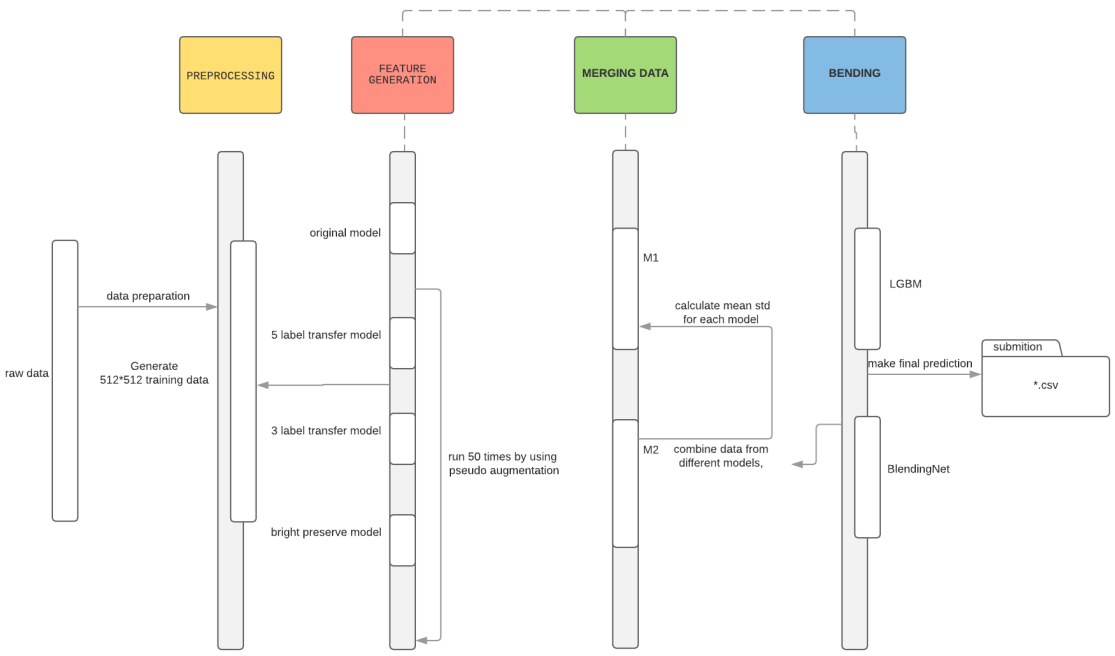

Fig. 4. The overall pipeline of our proposed method.

Table 1. Architecture of blend network

Layers(type)	Output shape
Input	(None, 8192)
dense1(Dense)	(None, 512)
activation1(Relu)	(None, 512)
dense2(Dense)	(None, 256)
activation2(Relu)	(None, 256)
dense3(Dense)	(None, 64)
activation3(Relu)	(None, 64)
dropout1(Dropout)	(None, 64)
dense4(Dense)	(None, 32)
activation1(Relu)	(None, 32)
dense5(Dense)	(None, 5)
activation5(Relu)	(None, 5)
softmax1(Softmax)	(None, 5)

pre-trained network and fine-tuned network in Table 2.

For DR grading prediction, there is no difference between blending on combine networks and pre-trained network. Both can achieve the best accuracy of 0.88 3. However, after we add the feature vector from bright preserved fine-tune training into combining model, the accuracy rises to 0.9. The loss has a significant drop down after we add bright-preserved fine-tuned network feature into combining blending Fig 6.

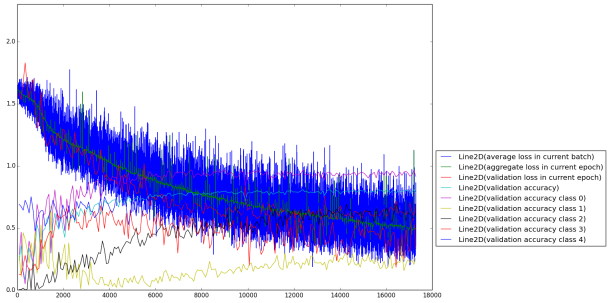

Fig. 5. original training

6. CONCLUSION

In this paper, we have proposed a unified framework for DR and DME grading. It utilizes several state-of-the-art techniques in deep learning to extract robust and invariant features, and feed the features into light-weight classification model to achieve performance gain. Our methods have been validated with the official benchmarks and and are showing perfromance gain on our test set. However, the DR and DME grading still remain a challenging problem as we could see from the low accuracy on the challenge test set. Methods like artificial random occulusion, multi-task learning would be heleful since we already have the localization information of those lesion regions.

Table 2. DME blending performance

Experiment	Overall accuracy
fine tuned pre-trained network	0.77
LGBM + pre-trained network	0.804
LGBM + fine-tuned pre-trained network	0.833
LGBM + pre-trained network + fine-tuned pre-trained network	0.863

Table 3. DR blending performance

Experiment	Overal Accracy
fine tuned pre-trained network	0.73
BlendingNet + pre-trained network	0.80
BlendingNet + fine-tuned pre-trained network	0.83
BlendingNet + pre-trained network + fine-tuned pre-trained network	0.85
LGBM + pre-trained network	0.666
LGBM + fine-tuned pre-trained network	0.634
LGBM + pre-trained network + fine-tuned pre-trained network	0.74
LGBM + fine-tuned pre-trained network + bright-preserved-tuned pre-trained network	0.764

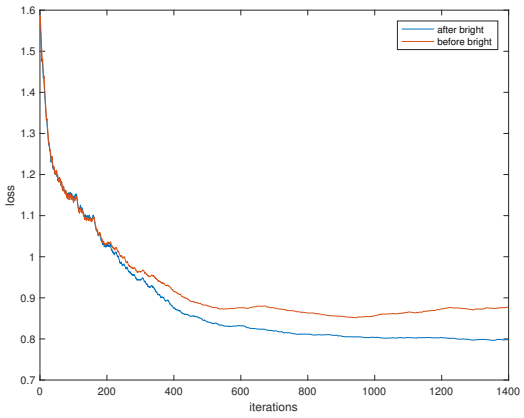

Fig. 6. comparison of adding bright preserved combine blending and original pre-trained network feature blending

7. REFERENCES

[1] M Usman Akram, Shehzad Khalid, Anam Tariq, Shoab A Khan, and Farooque Azam, "Detection and classification of retinal lesions for grading of diabetic retinopathy," *Computers in biology and medicine*, vol. 45, pp. 161–171, 2014.

[2] Karthikeyan Ganesan, Roshan Joy Martis, U Rajendra Acharya, Chua Kuang Chua, Lim Choo Min, EYK Ng, and Augustinus Laude, "Computer-aided diabetic retinopathy detection using trace transforms on digital fundus images," *Medical & biological engineering & computing*, vol. 52, no. 8, pp. 663–672, 2014.

[3] Varun Gulshan, Lily Peng, Marc Coram, Martin C Stumpe, Derek Wu, Arunachalam Narayanaswamy, Subhashini Venugopalan, Kasumi Widner, Tom Madams, Jorge Cuadros, et al., "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *Jama*, vol. 316, no. 22, pp. 2402–2410, 2016.

[4] Ratul Ghosh, Kuntal Ghosh, and Sanjit Maitra, "Automatic detection and classification of diabetic retinopathy stages using cnn," in *Signal Processing and Integrated Networks (SPIN), 2017 4th International Conference on*. IEEE, 2017, pp. 550–554.

[5] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2818–2826.

[6] Zhiguang Wang and Jianbo Yang, "Diabetic retinopathy detection via deep convolutional networks for discriminative localization and visual explanation," *arXiv preprint arXiv:1703.10757*, 2017.

[7] Sebastian Otálora, Oscar Perdomo, Fabio González, and Henning Müller, "Training deep convolutional neural networks with active learning for exudate classification in eye fundus images," in *Intravascular Imaging and Computer Assisted Stenting, and Large-Scale Annotation of Biomedical Data and Expert Label Synthesis*, pp. 146–154. Springer, 2017.

[8] Pallab Roy, Ruwan Tennakoon, Khoa Cao, Suman Sedai, Dwarikanath Mahapatra, Stefan Maetschke, and Rahil Garnavi, "A novel hybrid approach for severity assessment of diabetic retinopathy in colour fundus images," in *Biomedical Imaging (ISBI 2017), 2017 IEEE 14th International Symposium on*. IEEE, 2017, pp. 1078–1082.

[9] Gao Huang, Zhuang Liu, Kilian Q Weinberger, and Laurens van der Maaten, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, vol. 1, p. 3.

[10] Karen Simonyan and Andrew Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[11] M. Antony and S. Brggemann., "Kaggle diabetic retinopathy detection: Team o o solution," 2015.

[12] Jerome H Friedman, "Greedy function approximation: a gradient boosting machine," *Annals of statistics*, pp. 1189–1232, 2001.

[13] A.B. Smith, C.D. Jones, and E.F. Roberts, "Article title," *Journal*, vol. 62, pp. 291–294, January 1920.

[14] Mark JJP van Grinsven, Bram van Ginneken, Carel B Hoyng, Thomas Theelen, and Clara I Sánchez, "Fast convolutional neural network training using selective data sampling: Application to hemorrhage detection in color fundus images," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1273–1284, 2016.