2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Automatic Recognition of Ocular Surface Diseases on Smartphone Images Using Densely Connected Convolutional Networks

Rong Chen[1], Wankang Zeng[1], Wenkang Fan[1], Fang Lai[1], Yinran Chen[1], Xiang Lin[2], Liying Tang[2], Weijie Ouyang[2], Zuguo Liu[2], and Xiongbiao Luo[1], Senior Member, IEEE

Abstract— **Ocular surface disorder is one of common and prevalence eye diseases and complex to be recognized accurately. This work presents automatic classification of ocular surface disorders in accordance with densely connected convolutional networks and smartphone imaging. We use various smartphone cameras to collect clinical images that contain normal and abnormal, and modify end-to-end densely connected convolutional networks that use a hybrid unit to learn more diverse features, significantly reducing the network depth, the total number of parameters and the float calculation. The validation results demonstrate that our proposed method provides a promising and effective strategy to accurately screen ocular surface disorders. In particular, our deeply learned smartphone photographs based classification method achieved an average automatic recognition accuracy of 90.6%, while it is conveniently used by patients and integrated into smartphone applications for automatic patient-self screening ocular surface diseases without seeing a doctor in person in a hospital.**

I. INTRODUCTION

Ocular surface disease (OSD) is a common eye disease related to eye surface structures including the cornea, conjunctiva and eyelids [1]. The prevalence of OSDs diagnosed on symptoms ranges from 7% to 52% worldwide [2]. Most of OSDs such as abnormal conjunctiva and cornea are commonly undiagnosed due to lack of standard description of symptoms and too many subtypes of OSD in practice. More unfortunately, OSD patients of all ages can develop photophobia, intermittent blurred vision, pain, limited ability to perform daily activities, and even depression in some cases [1], [3]. Therefore, it is important to accurately recognize OSDs at the early stage and treat them in an appropriate way without OSD exacerbation.

Ophthalmologists usually use typical medical instruments to examine ocular surface disease at the hospital. Corneal topography is the earliest non-invasive imaging technique to detect the cornea [4]. Clinical slit-lamp is most commonly performed to check OSD [5]. Optical coherence tomography is a powerful imaging method to examine various ocular conditions [6]. However, these clinical medical instruments are expensive and only available at the hospital. Ophthalmic outpatients usually require higher costs and longer visit durations [7]. On the other hand, a tremendous amount of ophthalmic outpatients (most of them are nonemergency eye concerns) bring huge labor and mental burden to ophthalmologists, resulting in screening thoroughly insufficient to spare patients these diseases. Therefore, ophthalmologists require an efficient tool or strategy to assist them to accurately and thoroughly screen OSDs beyond ophthalmology clinics.

In recent years, deep learning has been widely applied in medical image analysis, especially in automatic diagnosis diseases. Many common automatic classification network frameworks such as VggNet [8], ResNet [9], and DenseNet [10] have achieved great success. In addition, to better embed mobile devices, some advanced lightweight network like ShuffleNet [11], MobileNets [12], [13], and GhostNet [14] have also been developed. Hence, it is a promising strategy to use deep learning for recognizing ocular surface diseases automatically.

Based on these clinical problems and facts discussed above, this work aims to establish an automatic recognition framework to efficiently and effectively screen OSDs. Several highlights of this work are clarified as follows. Technically, we modify densely connected convolutional network for reducing the amount of parameters and calculation. We reduce the depth of networks and use the hybrid unit instead of standard convolution to learn more diverse features of ocular surface image that have enough diversity and flexibility to meet the needs of recognizing ocular surface diseases. On the other hand, we introduce smartphone cameras that are easy for OSD patients to take images, enabling patients themselves to conveniently and timely monitor the development of OSDs. Additionally, our efficient and effective strategy provides ophthalmologists with a promising way to reduce their labor and mental load, as well as cost and visit duration to improve patient efficiency in ophthalmology clinics.

II. APPROACHES

Fig. 1 details our deeply learned smartphone image ocular surface disease classification method on the basis of modifying the DenseNet architecture. We basically follow the architecture of DenseNet that consists of an initial convolution layer, three dense blocks, two transition blocks and a classification layer [10].

A. Dense Block

Different from the original DenseNet extracting single features, each dense block contains some hybrid units rather than a standard convolution (Conv). The hybrid unit consists of a 3×3 standard convolution and a 3×3 depthwise

[1]Rong Chen, Wankang Zeng, Wenkang Fan, Fang Lai, Yinran Chen, and Xiongbiao Luo are with the School of Informatics, Xiamen University, Xiamen 361005, China. (Correspondence: Xiongbiao Luo xbluo@gxmu.edu.cn)

[2]Xiang Lin, Liying Tang, Weijie Ouyang, and Zuguo Liu are with the School of Medicine, Xiamen University, Xiamen 361102, China.

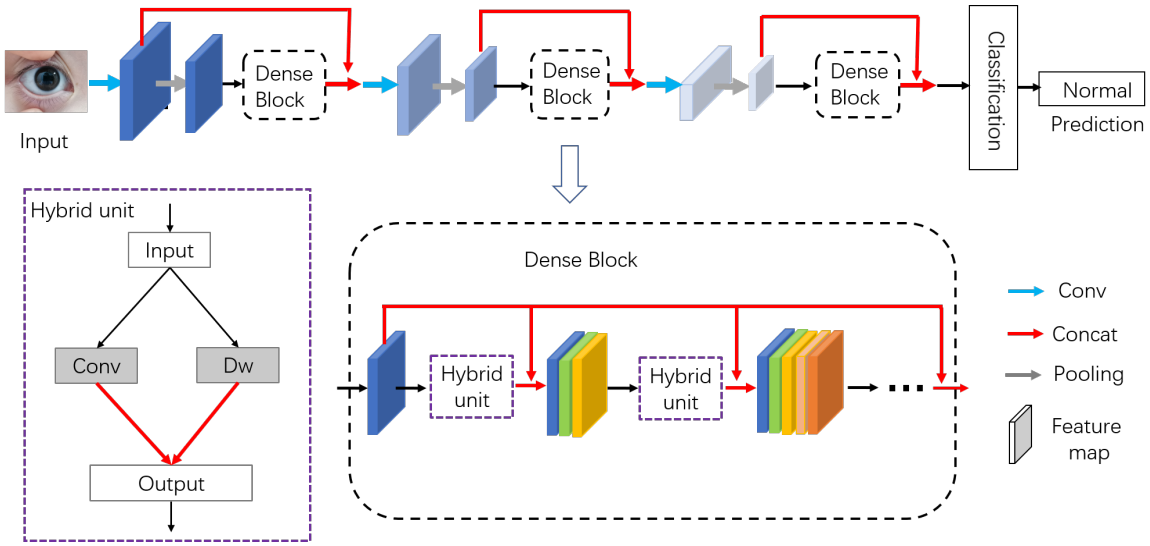


Fig. 1: Architecture of our proposed method based on DenseNet. The hybrid unit consists of a standard convolution (Conv) and a depthwise separable convolution (Dw).

separable convolutions (Dw) [12]. Each output is fixed to k (k denotes the growth rate) feature map that indicates the amount of new information contributed by this operation. These ouputs are concatenated. Therefore, each hybrid unit can learn the $2k$ feature map than the k feature map before. After extracting features, there will be a set of batch normalization (BN), rectified linear unit (ReLU), and 1×1 Conv. The BN layer can make overall sample distribution stable, limit the fluctuation of gradient to a certain level and improve the speed and performance of networks[15]. ReLU is a nonlinear activation function. The 1×1 kernel is used to reduce the number of channels in order to improve computational efficiency.

B. Transition Block

Each transition block includes four parts: BN, ReLU, a convolution layer, and an average pooling layer. Transition block is to connect two adjacent dense blocks and reduce the size and number of feature maps to improve efficiency. A convolution layer implemented by 1×1 kernel reduces the number of channels, while a pooling layer is implemented by 2×2 kernel to average pool.

In our experiments, we classify OSDs based on the DenseNet structure. The initial convolution layer consists of convolutions of 7×7 kernel, BN, ReLU, and max-pooling of 3×3 kernel. Three dense blocks connected by two transition blocks are used to learn features. Three dense blocks contain 2, 5 and 8 hybrid units respectively. The classification layer is to compress 359-channel feature maps to 2-channel for 2 classes. At the end, a softmax function estimates the probability of a photo classified into either **1** or **0**.

III. EXPERIMENTS

A. Data and Training

In data acquisition, three ophthalmologists used different smartphones with various camera sizes (e.g., 3648×2736, 4032×3024, 5120×3840) to acquire ocular surface images from over 1000 patients (with their consents). All the ocular surface images were collected under the following conditions: A typical distance (less than 20.0 cm) between eyes and smartphones, almost consistent lighting, and different viewing angles. Each ocular surface smartphone image was discussed and labeled together by three ophthalmologists. Eventually, we obtained 953 annotated ocular surface smartphone images including 467 normal and 486 abnormal images showing 12 types of ocular surface diseases.

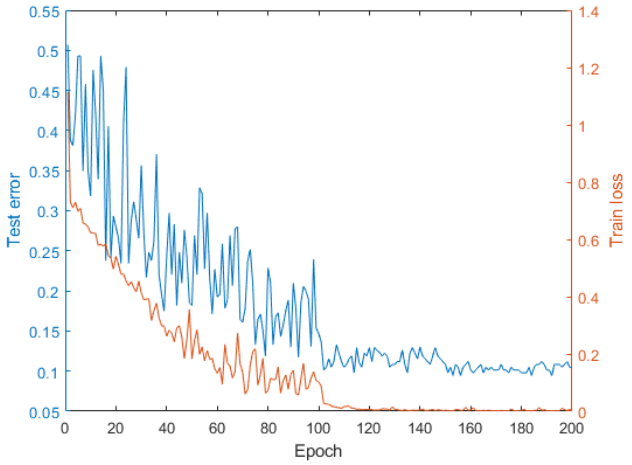


Fig. 2: Train loss and test error of OSDs recognition.

We randomly divided all the annotated data into training and testing by a ratio of $7:3$. We resized all images to 682×512 in our experiments. We did data augmentation by random flipping and cropping to alleviate the problem of a small dataset in training. While our dense connection can save each layer information at the maximum level, which is helpful

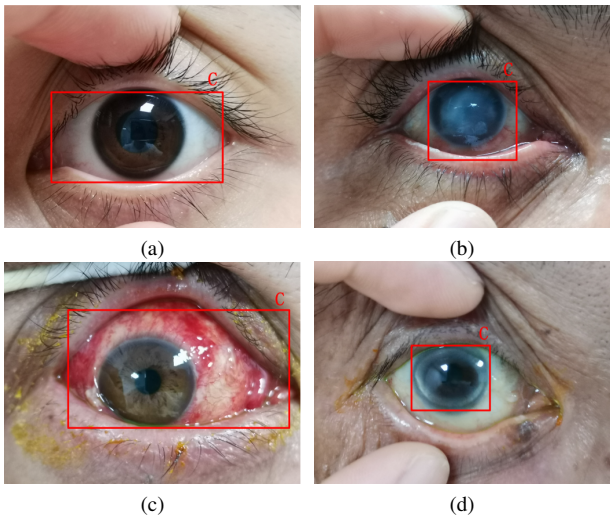


Fig. 3: Correctly recognized ocular surface diseases: (a) normal and (b) abnormal cornea-conjunctiva and (c) abnormal conjunctiva and (d) abnormal cornea.

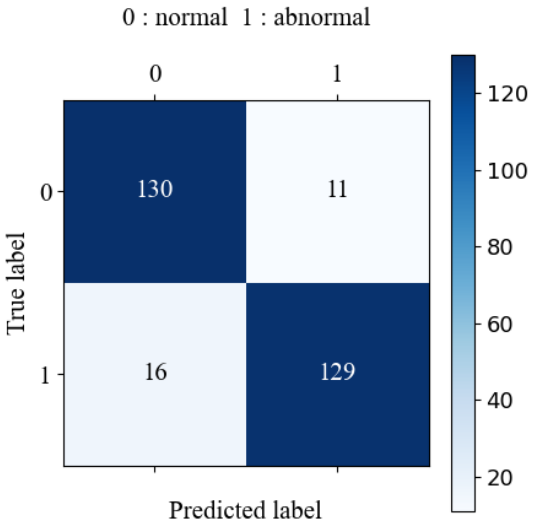


Fig. 4: The confusion matrix of our classification method.

for training [16]. The hyper-parameter k was set to 12 (our experiments showed that higher accuracy can be obtained than other smaller or larger ones). We used a logistic loss function to train model which was optimized by stochastic gradient descent. We trained our network for 200 epochs with a batch size of 16 and the initial learning rate was set to 0.15 and lower by 10 times at epoch 100 and 150. The training was performed by a momentum of 0.9. Fig. 2 shows the train loss and test error of recognizing OSDs. The speed of the training decline is quite fast at the early stage and the training procedure was gradually converged after 100 epochs.

B. Performance

Fig. 3 displays the correctly recognized cases of using our automatic OSD classification method. The confusion matrix (Fig. 4) demonstrates that 11 normal cases were misjudged as abnormal, while 16 photos with ocular suface diseases were incorrectly classified as normal. The average area under the receiver operating characteristic curve (AUC) of OSDs' recognition was about 0.96. We also calculated the specificity 92.20%, precision 92.14%, recall 88.97%, F1 score 90.53% and accuracy 90.56%.

C. Comparison

We compared the results of automatic recognition of ocular surface diseases with other state-of-the-art deep learning methods, including deep convolutional neural network like VggNet [8], ResNet [9] and DenseNet-121 [10], and the lightweight networks like ShuffleNet [11], MobileNet series [12], [13], and GhostNet [14]. The results are summarized in Table I. For a fair comparison, all parameters are identical to our method. Our method only has 0.31M parameters and 1.94B FLOPs with a recognition accuracy of 90.56%. Our method reduces lots of parameters and FLOPs and consistently outperforms than VggNet, ResNet and DenseNet. Compared with the lightweight networks, our network achieves best performance after using the hybrid units to learn features and dense connection. Our approach also shows quite competitive results on number of parameters and computation. In addition, our hybrid units can improve the recognition accuracy to 90.56% that is much better than only standard convolutions (87.76%).

TABLE I: Comparison of other state-of-the-art deep learning methods for ocular surface diseases classification.

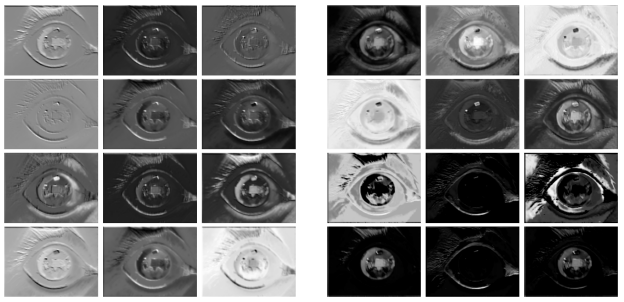
Method	Params(M)	FLOPs(B)	Acc(%)
VggNet-19 [8]	139.58	135.53	N/A
ResNet-50 [9]	23.51	28.93	74.83
ResNet-101 [9]	42.50	55.06	87.41
DenseNet-121 [10]	0.60	2.68	89.86
ShuffleNet v2 [11]	1.26	1.04	83.22
MobileNet v1 [12]	3.21	4.06	83.92
MobileNet v2 [13]	2.23	2.20	87.06
GhostNet [14]	3.62	0.82	86.71
Ours (only Conv)	0.16	1.50	87.76
Ours (hybrid unit)	0.31	1.94	**90.56**

IV. Discussion and Conclusion

This work proposes automatic recognition of ocular surface diseases in accordance with smartphone images and densely connected neural networks. The experimental results demonstrate that our approach completely outperforms currently available network architectures and achieves a high accuracy of 90.6%, providing a promising way to precisely screen and monitor ocular surface diseases without going to hospital.

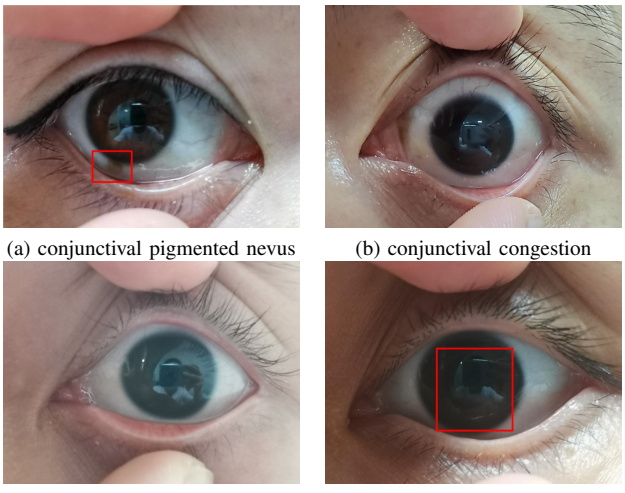
While it is generally hard to establish deeper convolutional neural networks with many operations on mobile phones due to their limited computation power, DenseNet can reduce

parameters by dense connections and feature reuses [10]. We follow the basic architecture of DenseNet and reduce the depth of the network, greatly reducing the amount of parameters and FLOPs. Meanwhile, we use a light-weight depthwise separable convolution, which can learn more diverse image features without increasing too much computation. We visualize the feature maps learned by the hybrid units and only standard convolution in the first layer of the first dense block (Fig. 5). Although the generated feature maps are extracted from the primary layer, there are significant differences between them, which implies that the generated features by hybrid units have enough diversity and flexibility to meet the needs of recognizing ocular surface diseases.



(a) The standard convolution (b) The hybrid unit

Fig. 5: Comparison of the feature maps generated from the first layer of the first dense block in the only standard convolution and the hybrid unit, respectively.



(a) conjunctival pigmented nevus (b) conjunctival congestion

(c) normal (d) normal

Fig. 6: Incorrectly classified ocular surface diseases.

Although our method works well, it still has several potential limitations. First, training samples are uneven. We did not have much training samples. Ocular surface disease is generally a complex and prevalent eye disorder with more than 150 subtypes in clinical practice. We will collect massive smartphone images to deal with this problem. Next, there are some unavoidable image problems in Fig. 6. Fig. 6(a) is conjunctival pigmented nevus whose diseased region is small. Fig. 6(b) is conjunctival congestion at Level 1 that is similar to the normal image. Fig. 6(c) is out of focus blurring. Fig. 6(d) is the obvious reflection in the corneal region. These problematic images undoubtedly deteriorate the performance of the model training useful information, resulting in a classification failure. We can consider taking a short video at different angles to supplement information for enhancing the recognition performance of networks. In future work, we will further validate our method in smartphone applications.

ACKNOWLEDGMENTS

This work was supported partly by the National NaturalScience Foundation of China under Grant 61971367, the Natural Science Foundation of Fujian Province of China under Grant 2020J01004, the Fujian Provincial Technology Innovation Joint Funds under Grant 2019Y9091, and the NEC Laboratories China.

REFERENCES

[1] Esen K Akpek and et al, "Dry Eye Syndrome Preferred Practice Pattern," *Ophthalmology*, vol. 126, no. 1, pp. 286, 2019.

[2] Heiko Pult and et al, "The development and evaluation of the new ocular surface disease index-6," *The Ocular Surface*, vol. 17, pp. 817–821, 2019.

[3] Jennifer P Craig and et al, "TFOS DEWS II Report Executive Summary," *Ocular Surface*, vol. 15, no. 4, pp. 802–812, 2017.

[4] Colin M Kirkness, "Corneal Topography, Principles and Applications," *British Journal of Ophthalmology*, vol. 83, no. 12, pp. 1409–1409, 1999.

[5] Marcus-Matthias Gellrich, *The Slit Lamp: Applications for Biomicroscopy and Videography*, Springer-Verlag Berlin Heidelberg, 2014.

[6] Johannes F. de Boer and et al, "Twenty-five years of optical coherence tomography: the paradigm shift in sensitivity and speed provided by fourier domain oct," *Biomedical Optics Express*, vol. 8, no. 7, pp. 3248–3280, 2017.

[7] Michelle R Hribar and et al, "Data-Driven Scheduling for Improving Patient Efficiency in Ophthalmology Clinics," *Ophthalmology*, vol. 126, no. 3, pp. 347–354, 2019.

[8] Karen Simonyan and et al, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *International Conference on Learning Representations*, 2015.

[9] K. He and et al, "Deep Residual Learning for Image Recognition," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[10] Gao Huang and et al, "Densely Connected Convolutional Networks," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[11] Xiangyu Zhang and et al, "Shufflenet: An extremely efficient convolutional neural network for mobile devices," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[12] Andrew Howard and et al, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[13] Mark Sandler and et al, "Mobilenetv2: Inverted residuals and linear bottlenecks," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[14] Kai Han and et al, "GhostNet: More Features from Cheap Operations.," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

[15] Sergey Ioffe and et al, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," *International Conference on Machine Learning*, pp. 448–456, 2015.

[16] Seonhoon Kim and et al, "Semantic sentence matching with densely-connected recurrent and co-attentive information," in *Proceedings of the AAAI conference on artificial intelligence*, 2019, vol. 33, pp. 6586–6593.