Automatic Classification of Retinal Optical Coherence Tomography Images With Layer Guided Convolutional Neural Network

Laifeng Huang, Xingxin He, Leyuan Fang, *Senior Member, IEEE*, Hossein Rabbani, *Senior Member, IEEE*, and Xiangdong Chen

***Abstract*—Optical coherence tomography (OCT) enables instant and direct imaging of morphological retinal tissue and has become an essential imaging modality for ophthalmology diagnosis. As one of the important morphological retinal characteristics, the structural information of retinal layers provides meaningful diagnostic information and is closely related to several retinal diseases. In this letter, we propose a novel layer guided convolutional neural network (LGCNN) to identify normal retina and three common types of macular pathologies, namely, diabetic macular edema, drusen, and choroidal neovascularization. Specifically, an efficient segmentation network is first employed to generate the retinal layer segmentation maps, which can delineate two lesion-related retinal layers associated with the meaningful retinal lesions. Then, two well-designed subnetworks in LGCNN are utilized to integrate the information of two lesion-related layers. Consequently, LGCNN can efficiently focus on the meaningful lesion-related layer regions to improve OCT classification. The experimental results conducted on two clinically acquired datasets demonstrate the effectiveness of the proposed method.**

***Index Terms*—Optical coherence tomography (OCT), convolutional neural network (CNN), OCT classification.**

I. Introduction

OPTICAL coherence tomography (OCT) is a high-resolution imaging technique which uses the coherent light to capture the micrometer-resolution, 3D tomography images from within biological tissue. For the human eye, with the capability of showing cross-sections of tissue layers, the OCT provides a straightforward method of assessing cellular organization [1] and axonal thickness in macular degeneration [2] and other eye diseases or systemic pathologies which contain ocular signs. Thus, the OCT imaging technique has recently been widely used in the identification and treatment of retinal diseases.

The macula is the central part of the retina and is closely related to humans' vision. Over the past several decades, numerous approaches have been applied for the identification of macular pathology [3]–[8]. In [3], Srinivasan *et al.* utilized the multiscale histograms of oriented gradient (HOG) descriptor as feature vector of a support vector machine (SVM) based classifier for the classification of age-related macular degeneration (AMD), diabetic macular edema (DME) and normal retina. In [4], Albarrak *et al.* proposed a decomposition based approach, coupled with a local feature extraction and Bayesian network classifier, to determine the presence of AMD in retina. Several other feature extractors (e.g., local binary pattern [5] and scale invariant feature transform [6]) and classifiers (e.g., random forest [7] and decision trees [8]) were also developed in retinal OCT images.

Recently, deep learning based method has been demonstrated to be a powerful tool in the field of computer vision and natural language processing [9]–[11]. The convolutional neural network (CNN) as a typical deep learning model can automatically extract a series of deep features hierarchically from the input images, which is very effective for many visual tasks [12], [13]. Also, the CNN model has been applied to retinal OCT image analysis, such as retinal disease classification [14]–[20], lesion detection [21], [22] and retinal layer segmentation [22], [23] of OCT images. In [19], the Wavelet-based CNN is introduced to generate representative features in the spatial-frequency domain for the final diagnosis of abnormal macula. In [22], the U-Net [24] is also introduced to segment the retinal layer and fluid in OCT images. Furthermore, an efficient method, transfer learning, has been applied to OCT field with excellent performance, especially with limited training data [25]. However, most of the aforementioned methods process the whole image and show little attention to the retinal layer regions corresponding to the meaningful lesion regions.

According to clinical experience, retinal diseases are usually caused by the lesions located in the specific retinal layers [26], [27]. For example, the DME appears between inner limiting membrane (ILM) and retinal pigment epithelium (RPE) [26] with fluid accumulation, named ILM-RPE layer, where fluid has different positions and shapes. In addition, the drusen and CNV, belonging to the early and advanced stage of AMD, respectively, are located between RPE and Bruch's membrane (BrM) [27] with distortions in which smooth layer appears wavy, named RPE-BrM layer. Besides, the distortion of CNV is more serious

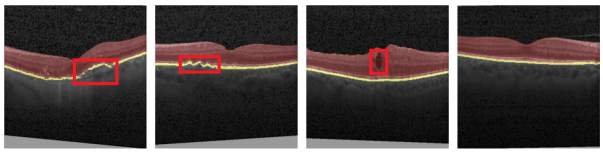
Manuscript received February 6, 2019; revised May 5, 2019; accepted May 8, 2019. Date of publication May 20, 2019; date of current version May 30, 2019. This work was supported in part by the National Natural Science Foundation under Grant 61771192, in part by the National Natural Science Foundation for Young Scientist of China under Grant 61501180, in part by the China Postdoctoral Science Foundation fund under Project 2017T100597, and in part by the National Natural Science Foundation of Hunan Province under Grant 2018JJ3077. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Charles Kervrann. *(Corresponding author: Leyuan Fang.)*

L. Huang, X. He, and L. Fang are with the College of Electrical and Information Engineering, Hunan University, Changsha 410082, China (e-mail: laifeng_huang@hnu.edu.cn; qq244436288@gmail.com; fangleyuan@gmail.com).

H. Rabbani is with the Medical Image & Signal Processing Research Center, Isfahan University of Medical Sciences, Isfahan 81745319, Iran (e-mail: h_rabbani@med.mui.ac.ir).

X. Chen is with the Department of Ophthalmology, First Hospital of Hunan University of Chinese Medicine, Changsha 410082, China (e-mail: 564259166@qq.com).

Digital Object Identifier 10.1109/LSP.2019.2917779



(a) CNV (b) Drusen (c) DME (d) Normal retina

Fig. 1. The red and yellow areas cover the ILM-RPE and RPE-BrM layer, respectively. The red block in OCT B-scan indicates key information about the diagnosis of retinal disease.

than drusen. Detailed illustrations are shown in Fig. 1. Therefore, the layer information should be favorable to identify the location of the lesion region and can be utilized to achieve more effective OCT classification.

In this letter, by considering the information of the lesion-related layers to make medical diagnosis, we propose a novel layer guided convolutional neural network (LGCNN) method for retinal OCT image classification. Specifically, we firstly introduce an efficient segmentation network (ReLayNet [22]) to extract the retinal layer segmentation maps. Then, two lesion-related layers named ILM-RPE layer (layers between ILM and RPE) and RPE-BrM layer (layers between RPE and BrM) are extracted from the layer segmentation maps. To integrate the information of the lesion-related layers for OCT classification, two well-designed subnetworks (i.e., ILM-RPE layer subnetwork and RPE-BrM layer subnetwork) are developed. Specifically, the ILM-RPE layer is incorporated into the ILM-RPE layer subnetwork to enhance the attention, which utilizes the local fluid accumulation information for the identification of DME. On the other hand, the RPE-BrM layer subnetwork is utilized to simply extract the shape distortion information from the RPE-BrM layer for the identification of CNV and drusen. Finally, LGCNN considers the fused information of lesion-related layers to identify the final macular pathology.

The remainder of this letter is organized as follows: The details of our LGCNN method are introduced in Section II. In Section III, experimental results are discussed. Conclusions and future works are suggested in Section IV.

II. Proposed Method

In this section, we will introduce the proposed LGCNN method for retinal OCT image classification in detail. Firstly, ReLayNet [22] is utilized to generate the retinal layer segmentation maps and two lesion-related retinal layers are extracted. Then, two well-designed subnetworks are developed to exploit the lesion-related information from the lesion-related layers. Lastly, a fully connected layer is employed to combine the last connected layers of two subnetworks for the final retinal disease classification.

A. Layer Information Extraction

As shown in Fig. 1, the DME is attributed to fluid accumulation in the ILM-RPE layer [26], while the drusen and CNV present as distortions of the RPE-BrM layer in which smooth layer appears wavy [27]. To obtain discriminative information of macular lesions in the retinal OCT image classification network, layer information needs to be extracted and employed to localize the lesions.

Recently, variety of methods have been developed for retinal layer segmentation, in which ReLayNet [22] is one of the most efficient methods. In [22], ReLayNet is designed as a fully convolutional deep architecture for the end-to-end segmentation of retinal layers and fluid masses in OCT B-scan. We merge multiple retinal layer to get two lesion-related layers (i.e., the ILM-RPE layer and the RPE-BrM layer). More details of ReLayNet can be found in [22].

B. Architecture of LGCNN

The architecture of LGCNN is shown in Fig. 2. Since two lesion-related layers are extracted from the generated segmentation maps, we design two corresponding subnetworks to extract the structural lesion-related information of the segmented layers for identification.

In ILM-RPE layer, fluid accumulation is a very important pathological feature for DME. Due to the unfixed position and various shape of fluid, ophthalmologists usually focus on the fluid region to make the diagnosis decision for DME. Inspired by this diagnostic procedure, we design an ILM-RPE layer subnetwork, which incorporates attention block (layer-integrated-block (LIB)) into the ILM-RPE layer, to focus more on the meaningful lesion-related layer region. Specifically, the ILM-RPE layer subnetwork covers five LIBs and max pooling layers and another two fully connected layers. Given an input feature maps F_i for ith LIB, M_u indicates the resized ILM-RPE layer probability maps and $G(x)$ is defined as the output from convolutional block. Then LIB can be denoted as follows:

$$F_{i+1} = G(M_u \otimes F_i + F_i), \tag{1}$$

which lesion-related feature maps can be emphasized by the ILM-RPE layer probability map. Multiple stacking of LIBs enables the ILM-RPE subnetwork to focus more on the meaningful lesion regions.

In RPE-BrM layer, drusen and CNV are both presented as the distortion of the layer that smooth layer appears wavy. Moreover, the distortion of CNV is more serious than drusen. Thus, ophthalmologists can distinguish them simply by the shape of RPE-BrM layer. Following this diagnostic mechanism, we design a RPE-BrM layer subnetwork, which utilizes the well-known residual blocks, to efficiently identify the retinal diseases. Specifically, the RPE-BrM subnetwork is composed of four residual blocks [28] and max pooling layers and another fully connected layer, to simply extract the distortion information from the PRE-BrM layer for the diagnosis of CNV and drusen.

Then, to mix the structural information of macular pathologies, the fully connection layers from two subnetworks are concatenated and the next fully connection layer outputs the final category. The max pooling layer in LGCNN downsamples the feature maps to accelerate convergence and reduce computational cost. The loss function for optimization is defined as multi-class cross entropy loss:

$$L_{LGCNN} = \sum_{x \in \Omega} q_l(x) log p_l(x), \tag{2}$$

where Ω is the collection of training samples, and $q_l(x)$ and $p_l(x)$ indicates the groundtruth and prediction probability of image x belonging to class l, respectively.

C. Training Strategy of LGCNN

In this letter, a two-step training strategy is adopted for the proposed LGCNN method. Training procedures are as follows.

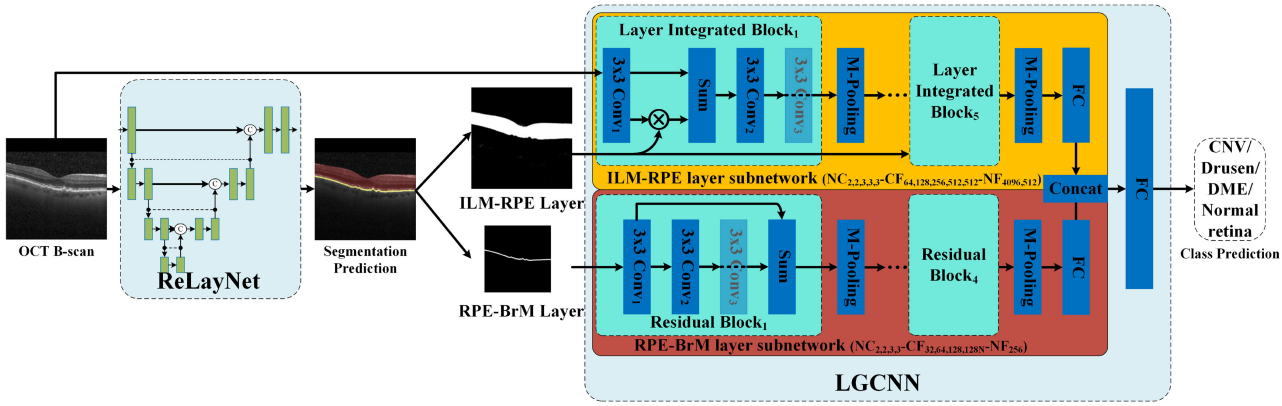


Fig. 2. The architecture of the proposed LGCNN method. The subscript of 'NC', 'CF' and 'NF' in LGCNN indicates the number of convolutional layer in each block, the channel of feature maps in each block and the nodes in fully connected layer, respectively. And the subscript of convolutional layer and block represents the corresponding serial number in each block and subnetwork.

Step I: We adopt a pre-trained model (ReLayNet [22]) without any further training to extract the retinal layer segmentation maps.

Step II: The ILM-RPE and RPE-BrM layer probability maps are extracted from the segmentation map and are fed into the LGCNN with the original OCT B-scan. The parameters of LGCNN are optimized by the back propagation of loss function proposed in (2) between groundtruth and prediction.

III. Experiments

A. Datasets Descriptions

The UCSD dataset is a publicly available dataset and contains 84484 OCT B-scans (Spectralis OCT, Heidelberg Engineering, Germany) captured from 4696 patients at several medical organizations, including the Shiley Eye Institute of the University of California San Diego (UCSD). Such a dataset can be classified into four categories (CNV, DME, drusen and normal retina) with the number of 37455, 11598, 8866 and 26565, respectively. More details about the UCSD dataset are referred to [25].

The HUCM dataset, a total of 8904 496 $\times$ 768 OCT B-scans (Spectralis OCT, Heidelberg Engineering, Germany) with 1581 CNV, 4592 DME, 1563 drusen and 1168 normal retina, is a local dataset, which is acquired and is obtained from the First Hospital of Hunan University of Chinese Medicine (HUCM). This dataset is labeled by five ophthalmologists independently and the final label is determined by the majority voting.

B. Experimental Setups

The pretrained ReLayNet model [22] is downloaded from the website: https://github.com/abhi4ssj/ReLayNet. In addition, the output of ReLayNet is merged to generate the retinal layer segmentation maps which contains the lesion-related layers.

For the training of LGCNN, we employ the Adam optimizer [30] with batch size of 16 B-scans sampled from training set during every iteration. To overcome the problem of class imbalance in the dataset, samples of each class in the training batch are the same. The initial learning rate is 1e-3 and decays by an order of magnitude per 10 epochs. Determined by the convergence of LGCNN model, the number of epoch is set to 20. Validation is done every 100 steps with 1600 B-scans randomly selected from the rest of dataset. No data augmentation is employed in training procedure. We resize the original OCT B-scan to 224 $\times$ 224 as input, while the ILM-RPE layer and the RPE-BrM layer map are resized into 224 $\times$ 224 and 112 $\times$ 112 to input the subnetwork, respectively. And the parameters of model are initialized with xavier algorithm [31].

For the evaluation of LGCNN model generalization and performance in the dataset, we randomly divide the dataset into k equal subsets at patient level. Note that, one of them is used for training and the remaining $k-1$ for validating and testing. The experiment is conducted k times and the final evaluation result is the average of all. Quantization of the classification performance is based on the following evaluation metrics: sensitivity (recall) (SE), precision (PR), specificity (SP), accuracy (ACC), overall accuracy (OA), overall sensitivity (OS) and overall precision (OP) according to the 4-class confusion matrix.

The LGCNN architecture is implemented using Tensorflow toolkit [32] with NVIDIA CUDA (version 9.0) and cuDNN library (version 7.4) by python language. ALL experiments are conducted on a computer running Ubuntu 16.04 LTS with Intel Core i7-9700K CPU, 11G NVIDIA GeForce RTX 2080 Ti GPU and 32G of RAM.

C. Compared Methods

To verify the effectiveness of the proposed LGCNN, four well-known OCT image classifiaiton methods are compared: HOG-SVM [3], transfer learning [25], VGG16 network [29] and IFCNN [17]. The HOG-SVM utilizes the HOG descriptor as a feature extractor and then trains a multiclass SVM classifier for OCT classification. In transfer learning method, a typical CNN network (Inception V3) is firstly trained on the ImageNet dataset [33] and then the last fully connected layer is finetuned on the available OCT dataset. The typical VGG16 architecture can be considered as a baseline network of LGCNN and is trained with the whole image. For comparison, the experimental settings of VGG16 are the same as those of LGCNN as described in Section III-B. IFCNN adopts an iterative fusion strategy to exploit multiple convolutional features within CNN for OCT image classification. In addition, we have also added some experiments about the guided layer. The LGCNN, in which RPE-BrM layer subnetwork is removed, is called LGCNN without RPE-BrM layer subnetwork, which utilizes the whole image and the ILM-RPE layer.

TABLE I
THE AVERAGED SENSITIVITY (SE), PRECISION (PR), SPECIFICITY (SP), ACCURACY (ACC), OVERALL ACCURACY (OA), OVERALL SENSITIVITY (OS) AND OVERALL PRECISION (OP) VALUES WITH STANDARD DEVIATION VALUES ON THE UCSD DATASET ($k = 10$). THE BEST RESULTS FOR EACH ITEM ARE LABELED IN BOLD

Method	Class	SE (%)	PR (%)	SP (%)	ACC (%)	OA/OS/OP (%)
HOG-SVM [3]	CNV	74.6±1.4	89.9±1.4	90.3±1.0	81.8±0.6	70.9±0.3 59.6±1.3 51.0±0.7
	Drusen	38.9±2.7	10.7±4.2	90.5±0.4	89.1±0.2	
	DME	55.1±4.4	15.2±1.5	88.0±0.2	86.8±0.3	
	Normal	70.0±1.4	87.9±1.4	93.7±0.6	84.3±0.6	
Transfer learning [25]	CNV	90.5±1.5	81.5±3.9	86.3±2.3	87.9±0.9	80.3±1.0 73.2±1.3 75.2±1.0
	Drusen	46.1±2.1	56.0±4.8	95.2±0.5	88.6±0.7	
	DME	69.0±4.9	69.4±5.4	95.2±0.7	91.5±0.5	
	Normal	87.2±2.0	90.0±2.6	95.3±1.1	92.6±0.4	
VGG16 [29]	CNV	89.5±2.5	89.5±3.7	91.6±2.4	90.6±1.1	81.2±1.9 73.8±3.1 70.2±2.9
	Drusen	49.0±6.8	37.3±7.3	93.0±0.7	89.5±1.1	
	DME	77.1±3.7	61.6±4.2	94.2±0.6	92.3±0.8	
	Normal	79.6±2.8	92.5±3.3	96.3±1.5	90.1±1.6	
IFCNN [17]	CNV	**94.8±1.9**	87.9±4.3	90.9±2.9	92.4±1.3	87.3±2.2 82.5±3.0 **84.7±2.4**
	Drusen	64.4±8.4	**76.8±7.2**	**97.3±0.8**	93.0±1.7	
	DME	79.2±8.9	**81.9±6.8**	**97.2±1.0**	94.4±1.0	
	Normal	**91.5±3.0**	92.2±4.7	96.4±2.0	**94.8±1.2**	
LGCNN	CNV	93.3±0.9	**91.5±2.4**	**93.3±1.7**	**93.3±0.7**	**88.4±1.3** **84.6±2.1** 82.9±1.8
	Drusen	**71.0±6.6**	65.2±5.1	96.0±0.6	**93.6±1.0**	
	DME	**85.7±2.6**	79.4±3.7	96.8±0.5	**95.4±0.5**	
	Normal	88.5±2.7	**95.5±2.4**	**97.9±1.1**	94.6±0.9	

TABLE II
THE AVERAGED SENSITIVITY (SE), PRECISION (PR), SPECIFICITY (SP), ACCURACY (ACC), OVERALL ACCURACY (OA), OVERALL SENSITIVITY (OS) AND OVERALL PRECISION (OP) VALUES WITH STANDARD DEVIATION VALUES ON THE HUCM DATASET ($k = 6$). THE BEST RESULTS FOR EACH ITEM ARE LABELED IN BOLD

Method	Class	SE (%)	PR (%)	SP (%)	ACC (%)	OA/OS/OP (%)
HOG-SVM [3]	CNV	82.3±3.4	55.1±4.9	91.8±0.9	90.7±0.4	76.0±0.6 74.9±1.3 68.2±0.9
	Drusen	66.6±2.5	53.7±3.8	89.9±0.8	86.4±0.6	
	DME	78.0±0.8	91.6±1.1	88.9±1.2	82.3±0.4	
	Normal	72.9±2.4	72.4±3.3	95.7±0.5	92.6±0.4	
Transfer learning [25]	CNV	74.4±1.4	77.4±1.6	95.1±0.3	91.2±0.4	81.3±0.5 76.7±0.6 79.6±0.5
	Drusen	67.5±2.1	70.8±2.2	93.7±0.4	88.9±0.6	
	DME	92.3±1.2	85.1±1.5	85.4±1.0	88.6±0.4	
	Normal	72.7±1.5	85.7±2.4	97.8±0.4	93.9±0.2	
VGG16 [29]	CNV	85.1±0.8	77.0±1.6	95.2±0.4	93.5±0.2	85.5±0.2 82.9±0.5 81.6±0.7
	Drusen	78.0±2.5	76.5±3.3	95.0±0.7	92.1±0.5	
	DME	89.9±0.9	93.1±1.2	92.3±1.2	91.0±0.2	
	Normal	78.5±3.5	79.4±3.2	96.9±0.5	94.4±0.4	
IFCNN [17]	CNV	83.8±3.5	86.8±3.8	97.1±0.8	94.6±0.4	87.8±0.8 84.3±0.9 86.5±0.9
	Drusen	79.0±4.9	82.4±6.9	96.2±1.4	92.9±0.7	
	DME	**95.6±1.4**	90.3±2.3	90.3±2.0	92.8±0.6	
	Normal	78.8±2.9	**86.5±2.7**	**97.9±0.4**	95.1±0.4	
LGCNN	CNV	**86.7±2.1**	**88.7±2.3**	**97.6±0.5**	**95.6±0.5**	**89.9±0.6** **87.2±0.7** **88.1±1.1**
	Drusen	**81.5±2.0**	**84.1±2.6**	**96.6±0.6**	**93.8±0.5**	
	DME	95.3±0.7	**93.5±1.1**	**93.2±1.1**	**94.3±0.4**	
	Normal	**85.1±1.9**	84.9±3.8	97.7±0.6	**96.1±0.5**	

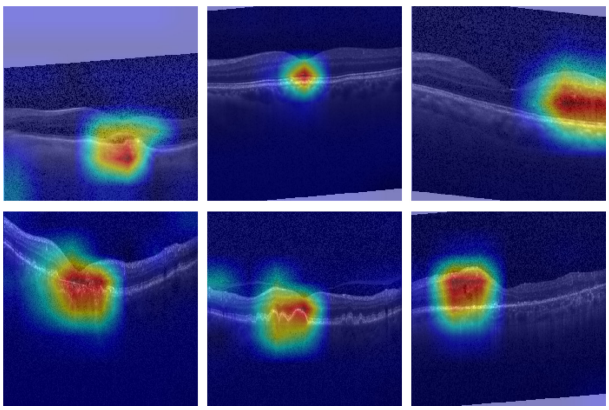


Fig. 3. The highlighted areas of pathology in CNV (left), drusen (middle) and DME (right) from the last convolutional layer. The heat maps above highlight areas that model considers important for diagnosis.

D. Experimental Results

Table I summarizes the SE, PR, SP, ACC, OA, OS and OP of different methods on the UCSD dataset ($k = 10$). As can be observed, deep learning based methods (transfer learning, VGG16, IFCNN and LGCNN) are superior to HOG-SVM method. In addition, the LGCNN achieves the best performance in most of the evaluation metrics. Furthermore, the improvements (in sensitivity) of the proposed LGCNN method over the VGG16 are 4.19%, 44.87%, 11.17%, and 11.14% in four categories (CNV, drusen, DME and normal retina), respectively, which demonstrates the effectiveness of lesion-related layer to guide OCT classification. To provide the ophthalmologists with a more transparent and interpretable diagnostic basis, Fig. 3 shows the heat maps of the proposed method LGCNN generated by Grad-CAM [34]. The heat maps are created by the last convolutional layer and the red highlight shows that the model considers important in identification.

The experiments on the HUCM dataset are conducted with $k = 6$. Table II summarizes seven evaluation metrics of several compared methods on the HUCM dataset. Similarly, LGCNN performs the best among these methods. The generalization capability of the proposed LGCNN can also be seen in comparison to the remaining methods on the two datasets.

TABLE III
THE AVERAGED OVERALL ACCURACY (OA) VALUES WITH STANDARD DEVIATION VALUES ON THE TWO DATASET. THE BEST RESULTS IN EACH DATASET ARE LABELED BOLD

Dataset	UCSD dataset (k=10)			HUCM dataset (k=6)		
Method	VGG16	LGCNN (without RPE-BrM layer subnetwork)	LGCNN	VGG16	LGCNN (without RPE-BrM layer subnetwork)	LGCNN
OA (%)	81.2±1.9	86.8±0.7	**88.4±1.3**	85.5±0.2	88.2±0.4	**89.9±0.6**

As shown in the Table III, it can been seen that introducing information of the guided layers can further improve the results. The ILM-RPE layer subnetwork utilizes the whole image and the ILM-RPE layer and enables the network to more effectively identify the retinal diseases. While the inconspicuous distortion caused by drusen is hardly captured by the ILM-RPE layer subnetwork, RPE-BrM layer subnetwork is adopted to extract the subtle distortion information more effectively. Therefore, the two corresponding subnetworks can capture complementary information and thus achieve higher accuracy.

IV. CONCLUSION

In this letter, we presented a layer guided convolutional neural network named LGCNN for the retinal OCT image classification. Compared to several typical methods, LGCNN utilized the correlation between macular pathologies and the specific retinal layers. Specifically, ReLayNet firstly generated the layer segmentation maps and two lesion-related layer probability maps were extracted. Then, the LGCNN was employed to integrate the lesion-related layer information for classification. Considering the characteristics of different layers, two subnetworks in LGCNN were specifically designed for lesion-related layers, which were very efficient for the final classification. Experiments on the UCSD dataset and the HUCM dataset demonstrated the effectiveness of LGCNN comparing to several typical methods. In the future, we will investigate the expansion of LGCNN for more retinal diseases (e.g., macular hole and macular telangiectasia), even for the detection and segmentation of retinal lesions.

REFERENCES

[1] N. Cuenca, I. Ortuo-Lizarn, and I. Pinilla, "Cellular characterization of OCT and outer retinal bands using specific immunohistochemistry markers and clinical implications," *Ophthalmology*, vol. 125, no. 3, pp. 407–422, 2018.

[2] P. A. Keane, P. J. Patel, S. Liakopoulos, F. M. Heussen, S. R. Sadda, and A. Tufail, "Evaluation of age-related macular degeneration with optical coherence tomography," *Surv. Ophthalmology*, vol. 57, no. 5, pp. 389–414, 2012.

[3] P. P. Srinivasan *et al.*, "Fully automated detection of diabetic macular edema and dry age-related macular degeneration from optical coherence tomography images," *Biomed. Opt. Express*, vol. 5, no. 10, pp. 3568–3577, 2014.

[4] A. Albarrak, F. Coenen, and Y. Zheng, "Age-related macular degeneration identification in volumetric optical coherence tomography using decomposition and local feature extraction," in *Proc. Int. Conf. Med. Image Understand. Anal.*, 2013, pp. 59–64.

[5] G. Lematre *et al.*, "Classification of SD-OCT volumes using local binary patterns: Experimental validation for DME detection," *J. Ophthalmology*, vol. 10, no. 12, pp. 329–346, 2016.

[6] Y. Sun, S. Li, and Z. Sun, "Fully automated macular pathology detection in retina optical coherence tomography images using sparse coding and dictionary learning," *J. Biomed. Opt.*, vol. 22, no. 1, pp. 12–21, 2017.

[7] F. G. Venhuizen *et al.*, "Automated age-related macular degeneration classification in OCT using unsupervised feature learning," *Proc. SPIE*, vol. 9414, 2015, Art. no. 94141I.

[8] R. Koprowski, S. Teper, Z. Wrbel, and E. Wylegala, "Automatic analysis of selected choroidal diseases in OCT images of the eye fundus," *BioMed. Eng. OnLine*, vol. 12, no. 1, 2013, Art. no. 117.

[9] G. E. Hinton, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[10] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[11] C. Xiang, L. Zhang, Y. Tang, W. Zou, and C. Xu, "MS-CapsNet: A novel multi-scale capsule network," *IEEE Signal Process. Lett.*, vol. 25, no. 12, pp. 1850–1854, Dec. 2018.

[12] J. Salamon and J. P. Bello, "Deep convolutional neural networks and data augmentation for environmental sound classification," *IEEE Signal Process. Lett.*, vol. 24, no. 3, pp. 279–283, Mar. 2017.

[13] S. Zhang, H. Yang, and Z.-P. Yin, "Transferred deep convolutional neural network features for extensive facial landmark localization," *IEEE Signal Process. Lett.*, vol. 23, no. 4, pp. 478–482, Apr. 2016.

[14] C. S. Lee, D. M. Baughman, and A. Y. Lee, "Deep learning is effective for classifying normal versus age-related macular degeneration OCT images," *Ophthalmology Retina*, vol. 1, no. 4, pp. 322–327, 2017.

[15] R. Rasti, H. Rabbani, A. Mehridehnavi, and F. Hajizadeh, "Macular OCT classification using a multi-scale convolutional neural network ensemble," *IEEE Trans. Med. Imag.*, vol. 37, no. 4, pp. 1024–1034, Apr. 2018.

[16] J. De Fauw *et al.*, "Clinically applicable deep learning for diagnosis and referral in retinal disease," *Nature Med.*, vol. 24, no. 9, pp. 1342–1350, 2018.

[17] L. Fang, Y. Jin, L. Huang, S. Guo, G. Zhao, and X. Chen, "Iterative fusion convolutional neural networks for classification of optical coherence tomography images," *J. Vis. Commun. Image Represent.*, vol. 59, pp. 327–333, 2019.

[18] L. Fang, C. Wang, S. Li, H. Rabbani, X. Chen, and Z. Liu, "Attention to lesion: Lesion-aware convolutional neural network for retinal optical coherence tomography image classification," *IEEE Trans. Med. Imag.*, doi: 10.1109/TMI.2019.2898414.

[19] Y. Rong *et al.*, "Surrogate-assisted retinal OCT image classification based on convolutional neural networks," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 1, pp. 253–263, Jan. 2019.

[20] R. Rasti, A. Mehridehnavi, H. Rabbani, and F. Hajizadeh, "Automatic diagnosis of abnormal macula in retinal optical coherence tomography images using wavelet-based convolutional neural network features and random forests classifier," *J. Biomed. Opt.*, vol. 23, no. 3, pp. 35–45, 2018.

[21] L. Fang *et al.*, "Automatic detection and recognition of multiple macular lesions in retinal optical coherence tomography images with multi-instance multilabel learning," *J. Biomed. Opt.*, vol. 22, no. 6, 2017, Art. no. 66014.

[22] A. G. Roy *et al.*, "ReLayNet: Retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional networks," *Biomed. Opt. Express*, vol. 8, no. 8, pp. 3627–3642, 2017.

[23] L. Fang, D. Cunefare, C. Wang, R. H. Guymer, S. Li, and S. Farsiu, "Automatic segmentation of nine retinal layer boundaries in OCT images of non-exudative AMD patients using deep learning and graph search," *Biomed. Opt. Express*, vol. 8, no. 5, pp. 2732–2744, 2017.

[24] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Med. Image Comput. Comput.-Assisted Intervention*, 2015, pp. 234–241.

[25] D. S. Kermany *et al.*, "Identifying medical diagnoses and treatable diseases by image-based deep learning," *Cell*, vol. 172, no. 5, pp. 1122–1131, 2018.

[26] F. E. Hirai, M. D. Knudtson, B. E. Klein, and R. Klein, "Clinically significant macular edema and survival in type 1 and type 2 diabetes," *Amer. J. Ophthalmology*, vol. 145, no. 4, pp. 700–706, 2008.

[27] A. Abdelsalam, L. Del Priore, and M. A. Zarbin, "Drusen in age-related macular degeneration: Pathogenesis, natural course, and laser photocoagulation-induced regression," *Surv. Ophthalmology*, vol. 44, no. 1, pp. 1–29, 1999.

[28] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 630–645.

[29] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2015. [Online]. Available: http://arxiv.org/abs/1409.1556

[30] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014. [Online]. Available: https://arxiv.org/abs/1412.6980

[31] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. Conf. Artif. Intell. Statist.*, 2010, pp. 249–256.

[32] M. Abadi *et al.*, "Tensorflow: A system for large-scale machine learning," in *Proc. Symp. Operating Syst. Des. Implementation*, 2016, pp. 265–283.

[33] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[34] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 618–626.