IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 14, 2018, accepted January 1, 2019, date of publication January 15, 2019, date of current version March 29, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2891975

Integrating Handcrafted and Deep Features for Optical Coherence Tomography Based Retinal Disease Classification

XUECHEN LI[1], LINLIN SHEN[1,2], MEIXIAO SHEN[3], AND CONNOR S. QIU[4,5]

[1]College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518000, China
[2]Guangdong Key Laboratory of Intelligent Information Processing, Shenzhen University, Shenzhen 518000, China
[3]School of Ophthalmology and Optometry, Wenzhou Medical College, Wenzhou 325000, China
[4]St Mary's Hospital, Isle of Wight NHS Trust, Newport PO30 5TG, U.K.
[5]Faculty of Medicine, Imperial College London, London SW7 2AZ, U.K.

Corresponding author: Linlin Shen (llshen@szu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61702337, Grant 61672357, and Grant U1713214, and in part by the Science and Technology Project of Guangdong Province under Grant 2018A050501014.

ABSTRACT Deep neural networks (DNNs) have been widely applied to the automatic analysis of medical images for disease diagnosis and to help human experts by efficiently processing immense amounts of images. While the handcrafted feature has been used for eye disease detection or classification since the 1990s, DNN was recently adopted in this area and showed a very promising performance. Since handcrafted and deep feature can extract complementary information, we propose, in this paper, three different integration frameworks to combine handcrafted and deep feature for optical coherence tomography image-based eye disease classification. In addition, to integrate the handcrafted feature at the input and fully connected layers using existing networks, such as VGG, DenseNet, and Xception, a novel ribcage network (RC Net) is also proposed for feature integration at middle layers. For RC Net, two ''rib'' channels are designed to independently process deep and handcrafted features, and another so-called ''spine'' channel is designed for the integration. While dense blocks are the main components of the three channels, sum operation is proposed for the feature map integration. Our experimental results showed that the deep networks achieved better classification accuracy after the integration of the handcrafted features, e.g., scale-invariant feature transform and Gabor. The RC Net showed the best performance among all the proposed feature integration methods.

INDEX TERMS Artificial intelligence, deep learning, optical coherence tomography, feature integration.

I. INTRODUCTION

Most of today's ophthalmologists identify eye diseases by visual observation and interpretation. Optical coherence tomography (OCT) has become a powerful imaging modality for non-invasive diagnosis of various retinal abnormalities, such as choroidal neovascularization (CNV) [1]–[3], diabetic macular edema (DME) [4], [5] and drusen [6]. Since accurate diagnosis has great impact for timely treatment, the process can be improved by employing disease-specific computer-aided diagnostic (CAD) systems based on OCT images [7].

Ophthalmologists have found some morphological changes such as variations in thickness of retinal layers and the choroid, the presence of DME, and drusen caused by eye diseases before any noticeable vision deterioration [8], [9]. The traditional OCT image based retinal disease detection methods relied on segmentation and feature analysis of retinal layers [10]–[12]. Eltanboly *et al.* [10] employed a Markov–Gibbs random field (MGRF) model for retinal layer segmentation and an auto-encoder was used for diabetic retinopathy (DR) detection. The method showed promising performance on a small OCT dataset (52 OCT scans). The performance of the method in [10] relied highly on the segmentation of retinal layers whilst its robustness has not been proven on large datasets. Duan *et al.* [11] proposed a geodesic distance method (GDM) for retinal layer segmentation. The curvature of the layer was employed for drusen detection. The segmentation method in [11] relied on the gradient of the layer borders and was thus sensitive to noise. Lemaître *et al.* [12] proposed a local binary pattern (LBP)-based method for DME detection. The method achieved 80%

accuracy on a small dataset (32 OCT volumes, 16 DME and 16 normal).

In recent years, deep learning showed great performance in many medical image analysis tasks. Many deep learning-based methods have been developed for retina segmentation and eye disease detection using OCT images [7], [13], [14]. Roy *et al.* [13] proposed a U-net-based fully connected network (FCN) for retina segmentation. The network showed over 0.9 Dice on the public Duke SD-OCT segmentation dataset. Lee *et al.* [14] proposed a convolutional neural network (CNN)-based method for age-related macular degeneration (AMD) detection. About one hundred thousand OCT scans were selected for training and testing the network. The network showed 0.87, 0.88 and 0.93 accuracy on image, macular and patient level, respectively. However, the dataset is not publicly available. Kermany *et al.* [7] implemented an Inception network based approach for multiple eye diseases detection. The network was pre-trained on the imageNet database and applied to OCT images using transfer learning technology. The classification accuracy was 0.97 and 0.93 on the ''full model'' and ''limited model'', respectively. The full model was trained using about eighty thousand OCT images while the limited model was trained using one thousand OCT images. Both models were tested using the same test set with one thousand images.

In recent years, fusion of handcrafted features and deep features has been used in medical image classification and face recognition. Antropova *et al.* [15] employed size, shape, texture, and morphology of segmented lesion areas as handcrafted features and fused them with deep features for diagnosis of breast cancer. The fusion was done at the feature level. Wang *et al.* [16] employed about 25 handcrafted features extracted using the first and second order filters and combined them with DNN to reduce the false positives of lung nodule detection. The fusion was also at the feature level. Xie *et al.* [17] introduced a handcrafted feature integration method for lung nodule classification. The gray-level co-occurrence matrix (GLCM) and Fourier descriptor were employed and fused with deep features vectors at the end of the network. Xie *et al.* [18] improved their work by employing DNN to the handcrafted features extracted on the ROI image. Zhang *et al.* [19] proposed a RGB-D DNN for face recognition. Depth image was employed as handcrafted feature image. A three channel network was proposed for feature integration. The handcrafted and deep features were fully integrated through the network. In summary, both handcrafted and deep features have been applied for OCT image based eye disease diagnosis. However, to the best of our knowledge, no work combining the two features has been done in this area.

In many cases of medical image recognition applications, large numbers of annotated samples are difficult to obtain. When lacking of training samples, the deep learning-based method often cannot achieve satisfactory performances, e.g. low accuracy or overfitting. Employing handcrafted features can help alleviate the problem. In this work, we presented a deep learning-based framework to combine handcrafted and deep features for eye disease classification. The handcrafted features, i.e. Scale-invariant feature transform (SIFT) and Gabor were employed in this work since they can extract edge and texture features, which are significant for OCT image diagnosis. In addition to integrate handcrafted features at input and output of FC layer using existing networks like VGG [20], Xception [21] and DenseNet [22] we also proposed a dense block based Ribcage network (RC Net) to integrate handcrafted features at different layers of the network. All proposed networks were tested using the dataset collected in [7], which is the only large OCT dataset available in the literature. The main contribution is that a novel integration network, e.g. RC Net, was proposed to integrate the handcrafted and deep features and significant improvement of the performance has been achieved.

The paper was organized as below. Section 2 introduces the databases used to train and evaluate the networks, SIFT and Gabor features and the integration networks employed for eye disease classification. Section 3 presents the experimental results of the proposed methods. Finally, Section 4 draws a conclusion.

II. MATERIALS AND METHODS

A. DATASET

We employed the publicly available dataset proposed in [7] to train and test our method. The OCT images were selected from retrospective cohorts of adult patients from the Shiley Eye Institute of the University of California San Diego, the California Retinal Research Foundation, Medical Center Ophthalmology Associates, the Shanghai First People's Hospital, and Beijing Tongren Eye Center between July 1, 2013 and March 1, 2017. All OCT imaging was performed as part of routine patient clinical care. There were no exclusion criteria based on age, gender, or race. The dataset contained 84484 OCT images captured from 6277 subjects, which are classified into four categories, i.e. CNV, DME, DRUSEN and NORMAL (examples were shown in Fig. 1). The number of images for the four categories are 37455, 11598, 8866 and 26565, respectively. As far as we know, the database is the only publicly available large OCT dataset for eye disease diagnosis. The work presented in [7] tested their network in two modes, i.e. ''full model'' and ''limited model'', employing Inception V3 network. While the full model was trained using about eighty thousand OCT images, the limited model was trained using one thousand OCT images. The main purpose is to test the performance of the algorithm when large and small numbers of training data are available.

In this paper, we designed three groups of testing data to evaluate our approach. For group 1 we kept 1000 images (250 for each class) for testing and the remaining images for training. For group 2 we employed the same testing dataset as group 1 and selected another 1000 images (250 for each class) for training. While group 1 and 2 are exactly the same as that presented in [7], we further used the third groups

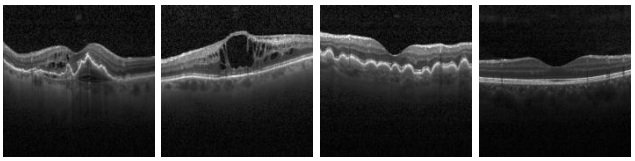

FIGURE 1. Examples of the OCT images (from left to right: CNV, DME, DRUSEN and NORMAL).

TABLE 1. The training and testing datasets.

Group	Class	Training		Testing	
		Images number	Patient number	Image number	Patient number
1	CNV	37205	791	250	129
	DME	11348	709	250	166
	DRUSEN	8616	713	250	142
	NORMAL	26315	3407	250	220
	TOTAL	83484	5620	1000	657
2	CNV	250	250	250	129
	DME	250	250	250	166
	DRUSEN	250	250	250	142
	NORMAL	250	250	250	220
	TOTAL	1000	1000	1000	657
3	CNV	18736	483	18719	437
	DME	5820	443	5778	432
	DRUSEN	4475	427	4391	428
	NORMAL	13293	1901	13272	1726
	TOTAL	42324	3254	42160	3023

to test the algorithm when a reasonable number of training data was available. For group 3, we selected 50% images for training and the remaining 50% for testing. The training and testing images in all three groups were selected from different patients. Table 1 lists the detailed information for the three groups of testing data.

B. METHODS

In this work, we employed SIFT and Gabor as handcrafted features and combine them with deep features for eye disease classification. Three frameworks, namely early, late and full integration frameworks were proposed in this paper to integrate handcrafted features into the DNN for performance enhancement. While early and late integration frameworks simply concatenate original and feature images or the feature maps output by FC layers, full integration network processes the original and handcrafted feature images through different channels, and combine the output of each dense block through the inference. Figure 2 shows the architectures of the proposed networks. One can observe from the figure that the full integration network has a much more complex structure to fuse the handcrafted feature with deep feature at different levels.

1) HANDCRAFTED FEATURES

The curvature and thickness of the retina layers are the key characteristics of eye disease diagnosis [10]. Therefore, two handcrafted features sensitive to the orientation of edges and textures, i.e. SIFT and Gabor, were employed.

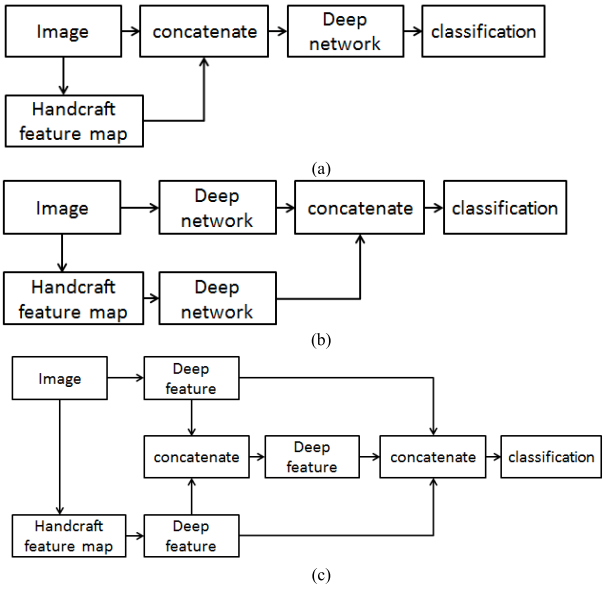

FIGURE 2. The architectures of the proposed networks.

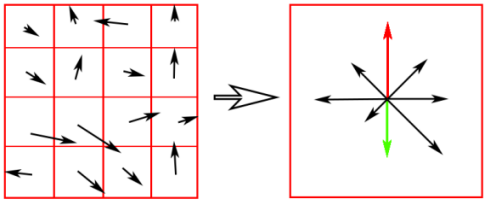

FIGURE 3. SIFT feature extraction.

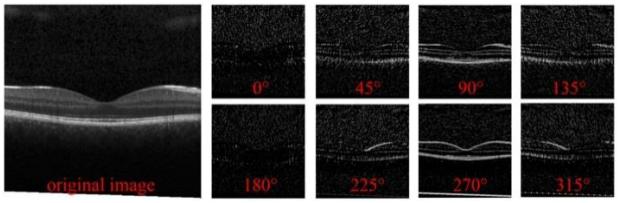

FIGURE 4. The SIFT features of different directions.

a: SIFT FEATURE

The SIFT features [23] are local features representing local appearance of the object at particular interest points, and are invariant to image scale and rotation. They are also robust to changes in illumination, noise, and small changes in viewpoint. In addition to these properties, they are highly distinctive, relatively easy to extract, and thus are widely used in object recognition applications. Fig. 3 shows the SIFT feature extraction process. In this work, the window with 8×8 pixels was employed. The direction vectors of each window was calculated as the SIFT features. The SIFT features at each direction were given in Fig. 4. The retina layers show horizontal structure, therefore, the SIFT features with vertical orientation are more effective.

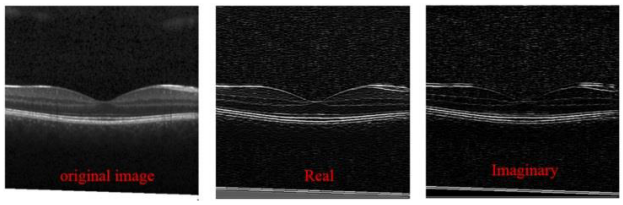

FIGURE 5. The real and imaginary component of Gabor features.

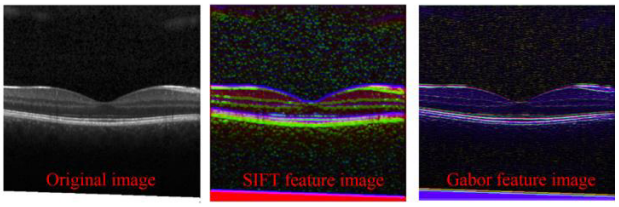

FIGURE 6. The generated SIFT and Gabor feature image for early integration.

b: GABOR FEATURE

The impulse response of Gabor filter is defined by a sinusoidal wave (a plane wave for 2D Gabor filters) multiplied by a Gaussian function [24]. Due to the multiplication-convolution property (Convolution theorem), the Fourier transform of a Gabor filter's impulse response is the convolution of the Fourier transform of the harmonic function (sinusoidal function) and the Fourier transform of the Gaussian function. The filter has a real and an imaginary component representing orthogonal directions. The two components may be formed into a complex number or used individually. The definition of components is given below.

Real component:

$$g(x, y, \lambda, \theta, \psi, \sigma, \gamma) = \exp\left(-\frac{x'^2+\gamma^2 y'^2}{2\sigma^2}\right)\cos\left(2\pi\frac{x'}{\lambda}+\psi\right) \quad (1)$$

Imaginary component:

$$g(x, y, \lambda, \theta, \psi, \sigma, \gamma) = \exp\left(-\frac{x'^2+\gamma^2 y'^2}{2\sigma^2}\right)\sin\left(2\pi\frac{x'}{\lambda}+\psi\right) \quad (2)$$

Complex component:

$$g(x, y, \lambda, \theta, \psi, \sigma, \gamma) = \exp\left(-\frac{x'^2+\gamma^2 y'^2}{2\sigma^2}\right) \times \exp\left(i\left(2\pi\frac{x'}{\lambda}+\psi\right)\right) \quad (3)$$

In these equations, λ represents the wavelength of the sinusoidal factor, θ represents the orientation of the normal to the parallel stripes of a Gabor function, ψ is the phase offset, σ is the standard deviation of the Gaussian envelope and γ is the spatial aspect ratio which specifies the ellipticity of the support of the Gabor function.

We employed the real and imaginary component as the Gabor features (see Fig. 5). In this work, $\lambda = 4, \theta = 90, \psi = 0, \sigma = 2, \gamma = 0.5$.

2) FEATURE INTEGRATION NETWORK

In order to combine handcrafted features and deep learning, three different integration architectures were proposed in this work. The architectures of the proposed networks were given in Fig. 2 (a. early integration; b. late integration and c. full integration). As shown in Fig. 2(a), the early integration method integrated the original and handcrafted feature map at the first stage as the feature image (the SIFT and Gabor feature image given in Fig. 6); and then input it into the DNN. The late integration method, shown in Fig. 2(b), input the original image and the handcrafted feature image into two independent deep networks, concatenated the output feature maps of FC layers, and made the final classification using softmax. The full integration method (shown in Fig. 2(c)) was different from the previous two integration methods, and can be considered as the combination of the both.

a: EARLY INTEGRATION NETWORK

The early integration utilized the original and handcrafted feature images at the very beginning. The SIFT/Gabor images were concatenated with the original image and the concatenation is used as the input of the network. In this work, we concatenated the two SIFT feature images at 90° and 270° with the original image, generated a RGB color image (showed as the middle image in Fig. 6), and used this image as the input of deep network. For Gabor we concatenated the real and imaginary component with the original images and generated the input using the same processing as SIFT (showed as the right image in Fig. 6). The networks employed for early integration were classic DNNs, i.e. VGG-16, DenseNet and Xception.

b: LATE INTEGRATION NETWORK

In late integration network, the information of the original and feature images were integrated at the end of the deep network. Two independent deep networks were employed to extract features of the original and feature images, respectively. The outputs of the two networks were 1D feature vectors, which were concatenated and then fully connected to the classification layer. If the same network is adopted, the late integration network required twice the network parameters when compared to that of the early integration network.

c: FULL INTEGRATION NETWORK

In the full integration network, the information of the original and handcrafted feature image interchanged after every convolutional block. To realize this, the structure of the network used in full integration should be redesigned, and the existing networks cannot be employed directly. In this work, we proposed a DenseNet-based ribcage network (RC) to fully integrate the handcrafted and deep features. Our RC Net is

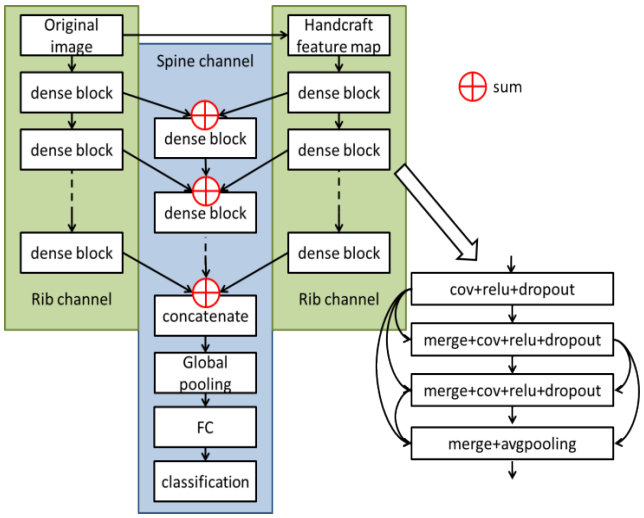

FIGURE 7. The structure of RC Net and dense block.

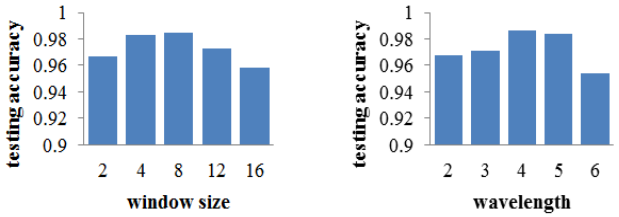

FIGURE 8. The testing accuracy when using SIFT and Gabor feature with different parameters.

an improved version of the ribcage network previous developed as a discriminator in Generative Adversarial Networks (GAN) [25]. As shown in Fig. 7, the RC Net consists of three channels. The first and second channels, referred to as the ''Ribs'', process the original and handcrafted feature images respectively. The third channel, referred to as the ''Spine'', uses the fusion of the feature map of each ''Rib'' channel as input.

The original RC Net employed convolutional blocks (series connection of 2D convolutional layers) for deep feature extraction; and concatenated the output feature maps of the ''rib'' channel as the inputs of the ''spine'' channel. In this work, we replaced the convolutional blocks with the recently proposed dense blocks [22], as the cross layer connections, which can significantly reduce the number of parameters for deep networks. Sum, instead of concatenation, was used to fuse the output feature maps of ''rib'' channels. While concatenation doubled the number of feature maps, sum operation does not increase the complexity of the network. In this work, five and four dense blocks were employed in each ''Rib'' and ''Spine'' channel, respectively. For each dense block, the number of convolutional layers was 3; the growth rate was 12 and the dropout rate was 0.5.

3) DATA AUGMENTATION

Additional training instances are generated by applying random crop (range from 0.8 to 1 image size), zoom in (range from 1 to 1.2) and mirroring to the images of the training set. Random noise with zero mean and 0.001 standard deviation was added to the input images.

4) NETWORK TRAINING

In this work, the networks were implemented using Keras toolbox, and trained with a mini-batch size of 32 using four GPUs (GeForce GTX TITAN X, 12GB RAM). The VGG-16 and Xception were initiated using the pre-trained weights on ImageNet, while the DenseNet and RC Net were initiated using ''He_uniform'' [26]. The initial learning rate is set to 0.001. 'Adam' [27] instead of the traditional stochastic gradient descent (SGD) was employed as optimization algorithm, and to update neural network weights based on training data iteratively. All networks converged after 300 epochs.

III. EXPERIMENTAL RESULTS

A. PARAMETERS OF HANDCRAFTED FEATURES

In this work, we firstly tested the influence of parameters for handcrafted feature extractors, i.e. SIFT and Gabor. The window sizes of 2 to 16 were tested for SIFT and wavelengths of 2 to 6 were tested for Gabor. Other parameters remained the same. Due to the page limit, we only show the results of Xception network with early integration for testing group 2 (1000 images each for training and testing) in Fig. 8. One can observe from the figure that the handcrafted features extracted using the proposed window size and wavelength achieve the best testing accuracy.

B. PERFORMANCE OF NETWORKS WITH AND WITHOUT FEATURE INTEGRATION

Now we can evaluate the performances of the integration of handcrafted features when existing VGG, DenseNet and Xception networks are used. While Fig. 9 lists the performances of VGG, DenseNet and Xception with and without the integration of SIFT and Gabor. In Fig. 9, the blue, red and green bars represent the testing accuracy of the original DNN, DNN integrated with SIFT and DNN integrated with Gabor.

One can observe from the figure that all three DNNs showed better accuracy after integrating handcrafted features. However, the Xception and DenseNet achieved better performances than VGG. There is no significant differences among the integration of SIFT and Gabor features. For testing group 1 and 2, the DNNs with Gabor feature showed the best performance, while the DNN with SIFT feature showed the best performance for testing group 3. When using feature integration, the average performance improvement was 0.38%, 0.93% and 0.44% for testing group 1, 2 and 3, respectively. The integration of handcrafted features produced larger performance improvements for smaller training sets.

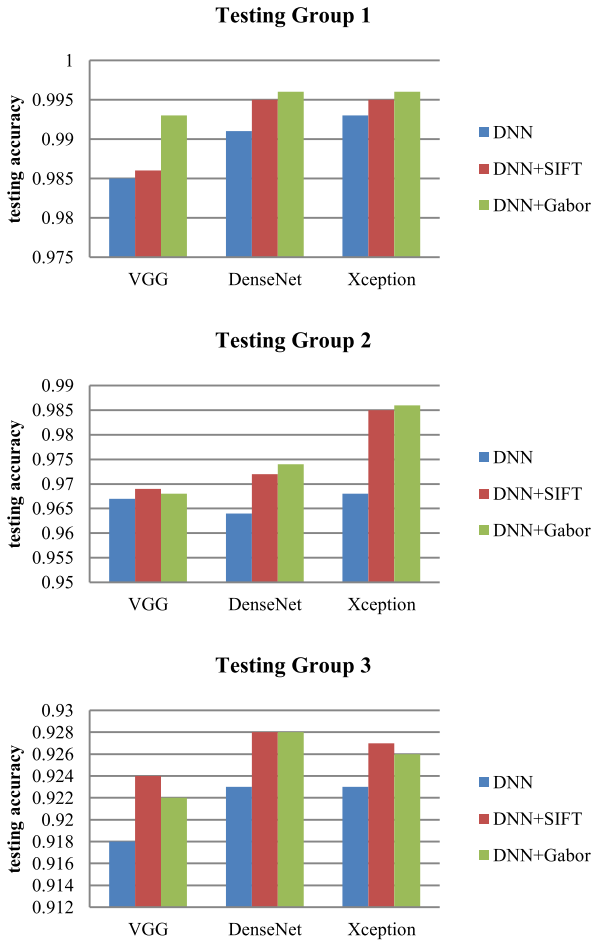

FIGURE 9. The testing accuracy of VGG, DenseNet and Xception with and without handcrafted feature integration.

TABLE 2. The number of parameters and testing accuracy of RC Net with different improvements.

		RC+Cc+Cv (original)	RC+S+Cv	RC+Cc+D	RC+S+D (proposed)
No. of Parameters		31M	3M	1.5M	**625K**
Testing Accuracy	Group 1	0.993	0.993	0.996	**0.996**
	Group 2	0.943	0.948	0.982	**0.988**
	Group 3	0.924	0.928	0.938	**0.941**

C. PERFORMANCE OF THE PROPOSED RC Net

As described in Section 2.2.2.C, the proposed RC Net replaced the convolutional blocks with dense blocks, and using sum instead of concatenation for feature fusion. In this section, we compared performance of the proposed RC Net with the original version in Table 2 and Fig. 10. Since SIFT and Gabor features showed similar performance, we employed Gabor features to evaluate the RC Nets. One can observe from the figure and table that dense block (RC+D) greatly improved the performance of RC Net, and using sum (S) instead of concatenation (Cc) can not only reduce the model parameters, but also improve the network performance. While the proposed RC Net with sum operation only requires 10% of the parameters of the original one, the proposed dense block further reduced the number to 5%. When both dense block and sum operations are used, the proposed RC Net greatly reduced the memory requirement from 31M to 625K. At the same time, 0.3%, 4.5% and 1.7% improvement were achieved using the proposed RC Net for group 1, 2 and 3, respectively. The proposed RC Net is not only much more efficient, but also more accurate.

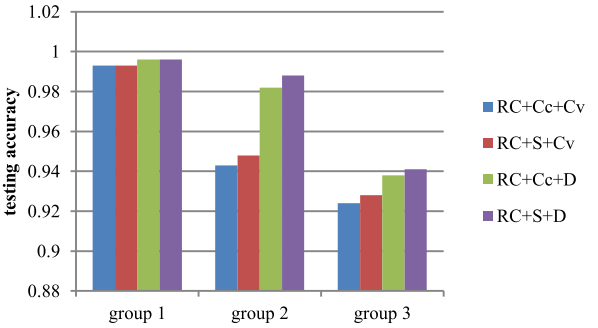

FIGURE 10. The testing accuracy of RC Net with different improvements (Cc: concatenation; Cv: convolutional block; S: sum; D: dense block).

TABLE 3. The comparison of the performances of three integration network (ACC: accuracy; SEN: sensitivity; SPE: specificity).

	Integration	Network	Handcrafted Feature	ACC	SEN	SPE
Group 1	None	Inception V3 [7]	N/A	0.966	0.978	0.974
		Xception	N/A	0.993	0.985	0.995
	Early	Xception	Gabor	0.996	0.991	0.996
	Late	Xception	Gabor	0.996	0.992	0.997
	Full	**RC**	**Gabor**	**0.996**	**0.993**	**0.997**
Group 2	None	Inception V3[7]	N/A	0.934	0.966	0.940
		Xception	N/A	0.968	0.952	0.988
	Early	Xception	Gabor	0.986	0.968	0.993
	Late	Xception	SIFT	0.986	0.969	0.993
	Full	**RC**	**Gabor**	**0.988**	**0.970**	**0.994**
Group 3	None	Inception V3 [7]	N/A	0.919	0.910	0.952
		DenseNet	N/A	0.923	0.913	0.958
	Early	DenseNet	Gabor	0.928	0.918	0.962
	Late	DenseNet	Gabor	0.933	0.921	0.964
	Full	**RC**	**Gabor**	**0.941**	**0.928**	**0.971**

D. COMPARISON OF DIFFERENT INTEGRATION NETWORKS AND THE STATE-OF-THE-ART

We also compared in Table 3 the performances of three different integration network architectures, i.e. early, late and full integration. We only listed the performance of the best model when handcrafted features are integrated or not. As the work in [7] is by far the only work available using the same dataset for eye disease detection, we also include its results for comparison. For dataset group 1 and 2, we directly included the reported results; for group 3, since it was not tested in [7], we replicated the network and tested the performance.

As both group 1 and 2 were used for testing in [7], the results for the two groups in this paper are directly comparable. For group 3, we implemented the same network, i.e. Inception V3 and reported its result in Table 3. As shown in the table, the proposed integration networks generally achieve much better performance than the Inception V3 network proposed in [7]. Take group 1 for example, the integration of Gabor features using the proposed RC Net improved the accuracy of [7] from 96.6% to 99.6%. Similarly, 5.4% and

2.2% improvement was also achieved for group 2 and 3, respectively. When handcrafted feature was not integrated, Xception and DenseNet also achieved better performance than Inception.

For different integration strategies, while the early and late integration showed equal classification performance for all three testing groups, the full integration method showed the best performance. Compared to the deep feature, the integration of handcrafted features achieved an improvement of 0.3%, 2% and 1.8% for testing group 1, 2 and 3, respectively. As only 1000 training samples are available in group 2, the integration of handcrafted features is more useful when limited training samples are available.

IV. CONCLUSION

In this work, we proposed three different feature integration methods: early, late and full integration for eye disease classification using a public OCT image database. Three groups of testing data were used to test the performance of the algorithm when large, small and reasonable numbers of training data are available.

The primary finding of this work is that the performance of DNN can be improved after integrating handcrafted features. It seems that the introduction of handcrafted features can help integrate the human prior knowledge and thus improve the performance of the DNNs. When limited training samples are available, such boosts in performance are more significant. The differences among different handcrafted features, i.e. SIFT and Gabor, are not substantial.

Another finding is that compared to late integration, early integration methods did not show worse performance for this database, but with a substantial saving in parameters and computation time. The full integration method showed better performance on all three testing groups, but with more parameters. The proposed RC Net employed dense block and sum operation to replace the convolutional block and concatenation. This can substantially reduce the model parameters and improve the performance of network.

REFERENCES

[1] M. R. Hee *et al.*, "Optical coherence tomography of age-related macular degeneration and choroidal neovascularization," *Ophthalmology*, vol. 103, no. 8, pp. 1260–1270, 1996.

[2] Y. Li, X. Guo, and K. Zhong, "Optical coherence tomography of exudative age-related macular degeneration and choroidal neovascularization," *Chin. Ophthalmic Res.*, vol. 20, no. 4, pp. 332–334, 2002.

[3] A. H. Rogers, A. Martidis, P. B. Greenberg, and C. A. Puliafito, "Optical coherence tomography findings following photodynamic therapy of choroidal neovascularization," *Amer. J. Ophthalmol.*, vol. 134, no. 4, pp. 566–576, 2002.

[4] M. D. Davis *et al.*, "Comparison of time-domain OCT and fundus photographic assessments of retinal thickening in eyes with diabetic macular edema," *Invest. Ophthalmol. Vis. Sci.*, vol. 49, no. 5, pp. 1745–1752, 2008.

[5] G. Panozzo, E. Gusson, B. Parolini, and A. Mercanti, "Role of OCT in the diagnosis and follow up of diabetic macular edema," *Seminars Ophthalmol.*, vol. 18, no. 2, pp. 74–81, 2003.

[6] N. Jain *et al.*, "Quantitative comparison of drusen segmented on SD-OCT versus drusen delineated on color fundus photographs," *Invest. Ophthalmol. Vis. Sci.*, vol. 51, no. 10, pp. 4875–4883, 2010.

[7] 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.

[8] A. J. Tatham, R. N. Weinreb, L. M. Zangwill, J. M. Liebmann, C. A. Girkin, and F. A. Medeiros, "The relationship between cup-to-disc ratio and estimated number of retinal ganglion cells," *Invest. Ophthalmol. Vis. Sci.*, vol. 54, no. 5, pp. 3205–3214, 2013.

[9] Z. Wu, L. N. Ayton, R. H. Guymer, and C. D. Luu, "Relationship between the second reflective band on optical coherence tomography and multifocal electroretinography in age-related macular degeneration," *Invest. Ophthalmol. Vis. Sci.*, vol. 54, no. 4, pp. 2800–2806, 2013.

[10] A. Eltanboly *et al.*, "A computer-aided diagnostic system for detecting diabetic retinopathy in optical coherence tomography images," *Med. Phys.*, vol. 44, no. 3, pp. 914–923, 2017.

[11] J. Duan, C. Tench, I. Gottlob, F. Proudlock, and L. Bai, "Automated segmentation of retinal layers from optical coherence tomography images using geodesic distance," *Pattern Recognit.*, vol. 72, pp. 158–175, Dec. 2016.

[12] G. Lemaître *et al.*, "Classification of SD-OCT volumes using local binary patterns: Experimental validation for DME detection," *J. Ophthalmol.*, vol. 6, Apr. 2016, Art. no. 3298606.

[13] 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.

[14] C. S. Lee, D. M. Baughman, and A. Y. Lee. (2016). "Deep learning is effective for the classification of OCT images of normal versus age-related macular degeneration." [Online]. Available: https://arxiv.org/abs/1612.04891

[15] N. Antropova, B. Q. Huynh, and M. L. Giger, "A deep feature fusion methodology for breast cancer diagnosis demonstrated on three imaging modality datasets," *Med. Phys.*, vol. 44, no. 10, pp. 5162–5171, 2017.

[16] C. Wang, A. Elazab, J. Wu, and Q. Hu, "Lung nodule classification using deep feature fusion in chest radiography," *Comput. Med. Imag. Graph.*, vol. 57, pp. 10–18, Apr. 2017.

[17] Y. Xie, J. Zhang, S. Liu, W. Cai, and Y. Xia, "Lung nodule classification by jointly using visual descriptors and deep features," in *Medical Computer Vision and Bayesian and Graphical Models for Biomedical Imaging*, Cham, Switzerland: Springer, 2017, pp. 116–125.

[18] Y. Xie, Y. Xia, J. Zhang, D. D. Feng, M. Fulham, and W. Cai, "Transferable multi-model ensemble for benign-malignant lung nodule classification on chest CT," in *Proc. Int. Conf. Med. Image Comput. Computer-Assist. Intervent.*, 2017, pp. 656–664.

[19] H. Zhang, H. Han, J. Cui, S. Shan, and X. Chen, "RGB-D face recognition via deep complementary and common feature learning," in *Proc. 13th IEEE Int. Conf. Autom. Face Gesture Recognit.*, May 2018, pp. 9–15.

[20] K. Simonyan and A. Zisserman. (2014). "Very deep convolutional networks for large-scale image recognition." [Online]. Available: https://arxiv.org/abs/1409.1556

[21] F. Chollet. (2016). "Xception: Deep learning with depthwise separable convolutions." [Online]. Available: https://arxiv.org/abs/1610.02357

[22] G. Huang, Z. Liu, K. Q. Weinberger, and L. van der Maaten, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jun. 2017, pp. 2261–2269.

[23] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Sep. 1999, pp. 1150–1157.

[24] I. Fogel and D. Sagi, "Gabor filters as texture discriminator," *Biological*, vol. 61, no. 2, pp. 103–113, 1989.

[25] A. Arbelle and T. R. Raviv. (2017). "Microscopy cell segmentation via adversarial neural networks." [Online]. Available: https://arxiv.org/abs/1709.05860

[26] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 1026–1034.

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

Authors' photographs and biographies not available at the time of publication.