J. Vis. Commun. Image R. 59 (2019) 327–333

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

journal homepage: www.elsevier.com/locate/jvci

Iterative fusion convolutional neural networks for classification of optical coherence tomography images ☆

Leyuan Fang [a], Yuxuan Jin [a], Laifeng Huang [a], Siyu Guo [a], Guangzhe Zhao [b,*], Xiangdong Chen [c,*]

[a] College of Electrical and Information Engineering, Hunan University, Changsha, China
[b] College of Electrical and Information Engineering, Beijing University of Civil Engineering and Architecture, Beijing, China
[c] The First Hospital of Hunan University of Chinese Medicine, Department of Ophthalmology, Changsha, China

ARTICLE INFO

Article history:
Received 30 August 2018
Revised 8 November 2018
Accepted 12 January 2019
Available online 19 January 2019

Keywords:
Classification
Convolutional neural network (CNN)
Deep learning
Optical coherence tomography (OCT)
Retinal

ABSTRACT

Optical coherence tomography (OCT) can achieve the high-resolution 3D tomography imaging of the retina, which is crucial for the diagnosis of retinal diseases. Currently, the classification of retinal OCT images is mainly conducted by ophthalmologists, which is time consuming and subjective. In this paper, we propose an iterative fusion convolutional neural network (IFCNN) method for the automatic classification of retinal OCT image. In the convolutional neural network (CNN), different convolutional layers contain feature information from different scales. Therefore, the proposed network adopts an iterative fusion strategy, which iteratively combines features in current convolutional layer with those of all previous layers in the CNN network, and thus can jointly utilize the features of different convolutional layers to achieve accurate classification of OCT images. Experimental results on a real retinal OCT dataset and a musculoskeletal radiographs dataset demonstrate the superiority of the proposed method over the traditional CNN and several well-known OCT classification methods.

1. Introduction

The retina plays a crucial role in the acquisition of visual information by the human eye. The macula is located in the center of the retina and is responsible for highly sensitive and accurate vision. Serious maculopathy can cause central vision loss or even blindness. Typical macular degenerations include the age-related macular degeneration (AMD), diabetic macular edema (DME), and choroidal neovascularization (CNV) [1]. AMD is one of the leading causes of blindness among the elderly in the world. About 15% of people over the age of 60 in the world are patients with age-related macular degeneration [2,3]. The common clinical feature of AMD is the presence of the drusen and asymptomatic deposition of extracellular material located between the retinal pigment epithelium (RPE) and Bruch's intracellular collagen layer [3]. The late stage of AMD, also known as CNV [4], usually results in irreversible macular damage, such as subretinal and intraretinal effusion, RPE detachment, and fibrosis scarring. DME is another common form of retinal disease, which is usually caused by diabetes, uveitis, retinal vein occlusion, and cataract surgery. The DME usually manifests as thickening of the retina, showing different degree of low-reflective cavity in the retina [5,6].

Optical coherence tomography (OCT) [7–9] is a non-invasive, high resolution, and 3D tomography imaging technique, which has been widely used to diagnose the retinal diseases [10]. Examples of retinal macular lesions in spectral-domain OCT (SD-OCT) images are shown in Fig. 1. In clinical analysis, ophthalmologists often make diagnostic decision on corresponding disease based on the information of the morphology, thickness, and brightness of the retinal membrane structure in the OCT image. This manual assessment process is not only time-consuming, but is also often subjective. Therefore, it is very important to develop techniques that can automatically recognize the retinal OCT images for the efficient diagnosis and remote treatment of retinal diseases.

In the past two decades, numerous retinal OCT classification algorithms have been developed [11–19], which usually consists of image preprocessing, feature extraction and classifier design. Image preprocessing includes denoising [20,21], layer flatten [22], and retinal alignment [23]. The typical feature extraction techniques includes LBP [11–13], HOG [14], and SIFT [16]. The classifier used in the OCT classification includes SVM [11,12,14,16], Bayesian classifier [24], and Random forest [25]. Recently, the neural network of deep learning [23,24] has demonstrated to be a very powerful tool in the fields of image processing and computer vision

☆ This article is part of the Special Issue on TIUSM.

* Corresponding authors.
E-mail address: zhaoguangzhe@bucea.edu.cn (G. Zhao).

https://doi.org/10.1016/j.jvcir.2019.01.022
1047-3203/

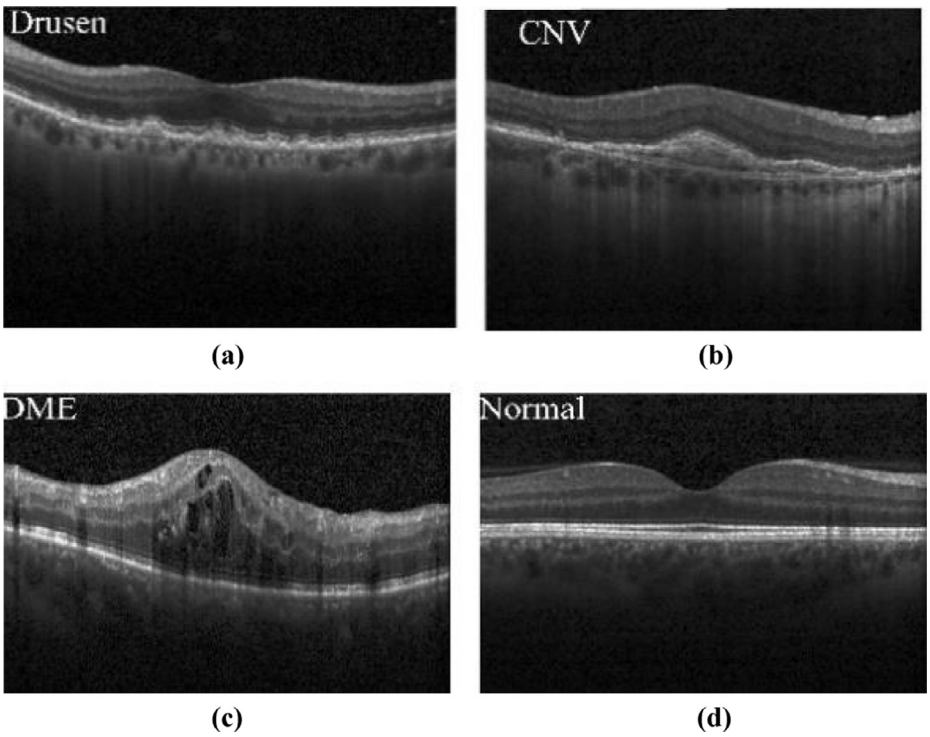

Fig. 1. SD-OCT B-Scans of the retina showing Drusen, CNV, DME and Normal macular, respectively.

[26]. One of the typical deep learning models [25] is the convolutional neural network (CNN), which can automatically learn a series of hierarchical features from a huge amount of training data sets. Recently, the CNN framework has also been extended to retinal OCT image analysis, such as retinal layer segmentation [27–29] and OCT image classification [30,31]. In the CNN model, different convolutional layers contain feature information from different scales. The traditional CNN-based OCT image classification method [29,30] uses only the features of the final convolutional layer for the classification without considering the information from previous convolutional layers. In this paper, we propose a novel iterative fusion convolutional neural network (IFCNN) method for retinal OCT image classification. To exploit the information among different convolutional layers, the proposed IFCNN method introduces an iteration layer fusion strategy. Specifically, we iteratively combine feature in current convolutional layer with those of all previous layers in the CNN network. Therefore, the feature information of all layers are iteratively combined, which can achieve better classification.

The remaining of this paper is organized as follows. Section 2 briefly reviews the convolutional neural networks. Section 3 introduces the proposed IFCNN classification method. Section 4 describes the experimental results on the real acquired clinical OCT data set. Section 5 concludes this paper and suggests some future works.

2. Related work

2.1. CNN model

Recently, deep neural network models have been successfully applied to many image classification problems [32]. As one of the typical deep model, the CNN is an end-to-end feature learning method that can automatically learn a hierarchical of features from a given training sample. As shown in Fig. 2, the CNN mainly includes a series of convolution layers, pooling layers, fully connected layers and softmax layers. The convolutional layer is one of the most important parts in CNN [33], which convolves multiple convolution kernels with the input image to create different feature maps. Specifically, shallow convolution layers with narrow fields of view can extract local information while deeper layers can capture global information with larger fields of view. Each neuron in the convolution layer senses the local area of the previous layer and takes this local area as its input. Assuming that x_i^{l-1} is the output of the ith neuron of the $l-1$ convolutional layer, the output of the lth neuron is:

$$x = \delta\left(\sum_{m=1}^{N} x_{i,m}^{l-1} * w_{i,j,m}^{l} + b_j^l\right), \tag{1}$$

where N is the local input size of the current neuron, $w_{i,j,m}^{l}$ and b_j^l are the connection weights and offsets of the two neurons, respectively.

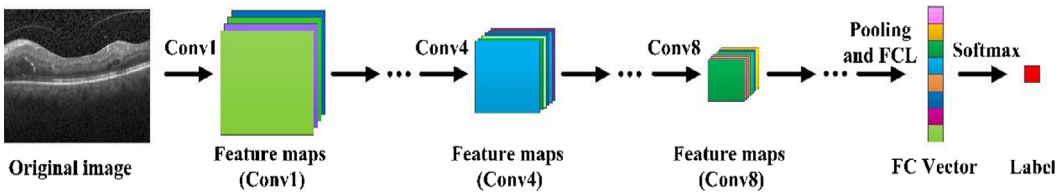

Fig. 2. The CNN model.

These weights and biases are continuously updated and optimized during the network training process. $\sigma(g)$ is a neural ReLU nonlinear unit usually using a meta-activation function: $\sigma(x) = max(0, x)$.

The pooling layer performs independent operations on each output feature map of the convolutional layer, and generally uses two operations: average pooling or maximum pooling. After the pooling operation, the resolution of the feature map is reduced, while the effective features are still retained, thus greatly reducing the network parameters. The fully connected layer is similar to the convolution layer and is also composed of many neurons. However, each neuron in the fully connected layer has a connection relationship with all neurons in the previous layer. The softmax layer is the last layer structure of CNN. The output of this layer is the probability that the input image belongs to C categories, and the function is to perform category prediction based on the features extracted by the network as the decision layer (DL). The difference between the predicted and actual tags of the input image can be quantified using the cross-entropy loss function [24]:

$$Loss = -\sum_{m=1}^{M}\sum_{c=1}^{C} y_c \log(p_c | I_{m,w,b}) \tag{2}$$

where C is the total number of categories, $I_m(m = 1, \cdots, M)$ is the input image. y_c and p_c represent the prior probability and the predicted probability of the input image belonging to category c, respectively. The process of CNN training is to optimize the weight parameters of the convolutional layer and the fully connected layer by minimizing the loss function, so that the entire network can automatically learn suitable feature representations for the target data.

Convolutional Neural Networks (CNN) have evolved from traditional artificial neural networks and achieved great success in image classification, target detection, and image segmentation [34–36]. Different from the traditional artificial neural network, the convolutional neural network uses the local connection and weight sharing strategy, which greatly reduce the parameters of the network and the complexity of the network model, thus making the deep network be easier to optimize. CNN combines the two processes of feature extraction and feature classification. Compared with the traditional manual feature extraction and classifier design method, the CNN is an end-to-end feature learning method that can automatically learn a hierarchical of features from a given training sample.

2.2. CNN based OCT classification

In ophthalmology, CNN has recently been used to automatically detect diabetic retinopathy, retinal segmentation, and quantification of macular fluid in OCT images from fundus photographs. For retinal OCT classification, CNN is expected to extract valid advanced features from the original OCT image. Formally, given the input OCT image $\boldsymbol{x}$, the output of the classification network can be calculated by a series of convolution, pooling, and FC operations:

$$\boldsymbol{a_k}(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{W}\boldsymbol{x} + \boldsymbol{b}), \tag{3}$$

where f denotes a composite function with input $\boldsymbol{x}$, which is obtained by multiple linear and nonlinear operations, and $\boldsymbol{a_k}$ denotes the class score of class k. W and b are the weight and deviation parameters of the network, respectively. As described in Section 2-A, the softmax function is used to generate a probability distribution of the output of the image x belonging to each class:

$$\boldsymbol{p}(\boldsymbol{k}|\boldsymbol{x}) = \frac{\boldsymbol{e}^{\boldsymbol{a_x}}}{\sum_{\boldsymbol{k}=1}^{\boldsymbol{K}} \boldsymbol{e}^{\boldsymbol{a_x}}}, \tag{4}$$

where K is the number of output classes. CNN typically trains backpropagation rules and classical stochastic gradient descent (SGD) algorithms by minimizing the cross entropy error between the probability output and the single thermal reference label:

$$\boldsymbol{L_{CLS}} = -\frac{1}{\boldsymbol{m}}\sum_{\boldsymbol{i}=1}^{\boldsymbol{m}}\sum_{\boldsymbol{k}=1}^{\boldsymbol{K}} \boldsymbol{I}(\boldsymbol{l_i} = \boldsymbol{k}) \log \boldsymbol{p}(\boldsymbol{k}|\boldsymbol{x_i}) + \boldsymbol{\lambda}\|\boldsymbol{W}\|_2, \tag{5}$$

where m is the number of samples for each small batch during training, $\boldsymbol{l_i}$ is the class label for image $\boldsymbol{x_i}$, and $\boldsymbol{I}(\cdot)$ is the index function. If $\boldsymbol{l_i}$ is equal to k, it is equal to k, and $\|\boldsymbol{W}\|_2$ is the n-norm of $\boldsymbol{L_2}$. The network weight is used as the regularization and attenuation factor $\boldsymbol{\lambda}$. Once the optimization is complete, the trained CNN can be used to predict the label of the test image $\boldsymbol{x}^*$ based on the maximum probability $\boldsymbol{p}(\boldsymbol{k}|\boldsymbol{x}^*)$:

$$\boldsymbol{Class}(\boldsymbol{x}^*) = \underset{\boldsymbol{k}=1,2,\cdots,\boldsymbol{K}}{\boldsymbol{arg\ max}}\ \boldsymbol{p}(\boldsymbol{k}|\boldsymbol{x}^*), \tag{6}$$

3. Methods

As the convolutional layer increases, CNN can use multiple convolution kernels to automatically extract features from fine to coarse. Fig. 3 shows some of the feature maps extracted from three different convolutional layers for OCT image. It can be observed that the feature map extracted in the shallow convolution layer (e.g., feature maps of Conv1) can reflect the detailed structure information of the input image. By contrast, feature maps from deeper convolutional layers (e.g., feature maps of Conv8) can represent more abstract information of the input image. To effectively utilize the information from all the convolutional layers for the automatic classification of retinal OCT images, in this paper, we

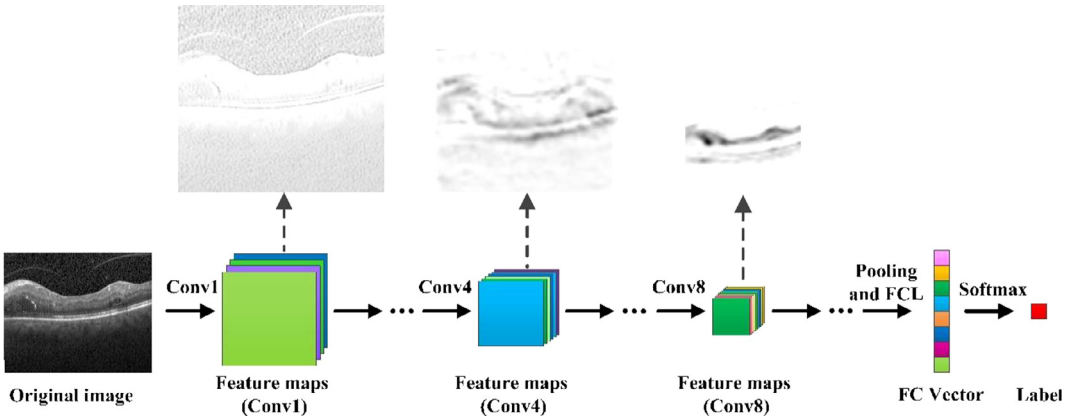

Fig. 3. Illustration of the convolutional feature maps extracted by CNN for OCT image.

propose an iterative fusion convolutional neural network (IFCNN) method. The proposed method adopts an iterative strategy to effectively combine the characteristics of the current convolutional layer with the features of all previous layers in the CNN network, so that the characteristics of different convolutional layers can be jointly used to achieve accurate classification of OCT images. The structure of the proposed IFCNN method is shown in Fig. 4. Detailed descriptions of the proposed OCT image classification network are given as follows.

The combination of information from different layers in the network can be defined as fusion [37]. The iterative fusion can be defined as the iterative combination of all the previous layers with the current layer. We divide the network's stacking blocks into several stages. The deeper stages encode abstract and coarse features, while the shallow stages capture low-level and fine features. Therefore, the fusion begins at the shallowest scale and then iteratively merges deeper scales. Through the integration of different stages, the shallow features are gradually refined.

Our method consists of two branches: the basic branch and the fusion branch. The basic branch is used to extract features of multiple layers. It can be easily established by one of off-the-shelf CNN architectures, such as AlexNet [27], VGG16 [38], and GoogleNet [39]. In this paper, we use VGG16 as the basic network. The fusion branch is used to iteratively integrate multiple layers of information. Our network consists of several identical modules, which are described below.

Let X^{l-1} be the feature maps of the l-1 layer of the basic branch with input image x, X^{l} be the feature maps of the l layer of the basic branch generated by the convolution and pooling operations with X^{l-1}. Y^{l-1} and Y^{l} represent the feature maps of the l-1 and l layers of the fusion branch, respectively. Specifically, Y^{l} is formulated as:

$$Y^{l} = \begin{cases} \text{concat}\left(X^{l-1}, X^{l}\right) & l = 1 \\ \text{concat}\left(Y^{l-1}, X^{l}\right) & l \neq 1 \end{cases} \tag{7}$$

Since X^{l} and Y^{l-1} have different spatial dimensions, and cannot directly concatenate, we perform a convolution of stride 2 on Y^{l-1} before fusing them. The outputs of last module in fusion branch are fed into the fully connected layer and the categorical probability distribution on each class can be generated by the softmax layer, as described in Section 2.

4. Experimental results

4.1. Clinical datasets

The dataset utilized in this experiment consists of 84,484 OCT B-scans obtained from 4686 patients at the Shiley Eye Institute of the University of California San Diego (UCSD) and is publicly available at: https://data.mendeley.com/datasets/rscbjbr9sj/2. All of the images are divided into four categories, namely drusen, CNV, DME and normal, the numbers are 8866, 37,455, 11,598 and 26,565, respectively. These images (acquired by Heidelberg Engineering, Spectralis OCT, Germany) were all selected from the retrospective cohort of adult patients, with no exclusion criteria based on age, sex, or ethnicity. According to the manufacturer's software and instructions, the final OCT scans are obtained by a horizontal foveal cut from original standard image format. More details about the dataset can be found in [40].

4.2. Experimental settings

Convergence is determined by periodically calculating the performance of an independent validation set (960 B-scans randomly selected from the lesion level data set) during training. At each iteration, the small batch size is set to one B-scan. The epoch is chosen to 10 and the initial learning rate is set to 10^{-4}. To train IFCNN, we used the Adam optimizer [41] to optimize the network with a small batch size of 24 and an initial learning rate of 10^{-4}. The training process will be ended after 10 epochs. No data augmentation strategies are adopted for training the IFCNN network. During the training of IFCNN, we adjusted the size of the original OCT image to 224 × 224 as input.

In our experiments, we used cross-validation to evaluate the classification performance of the proposed IFCNN method. Cross-validation involves dividing the dataset into 10 folds. Each verification involves one training and the remaining are used as the verification folds. We select the folds in the sequence way, instead of randomly choosing them. The classification performance was evaluated based on the following assessment metrics: accuracy, specificity and sensitivity, each reported value is averaged on the individual categories and all category labels. In our classification problems, the sensitivity of the independent class is the prediction accuracy, and the specificity is defined in the same way for each class, where the negative sample is a sample that is not in the class under consideration.

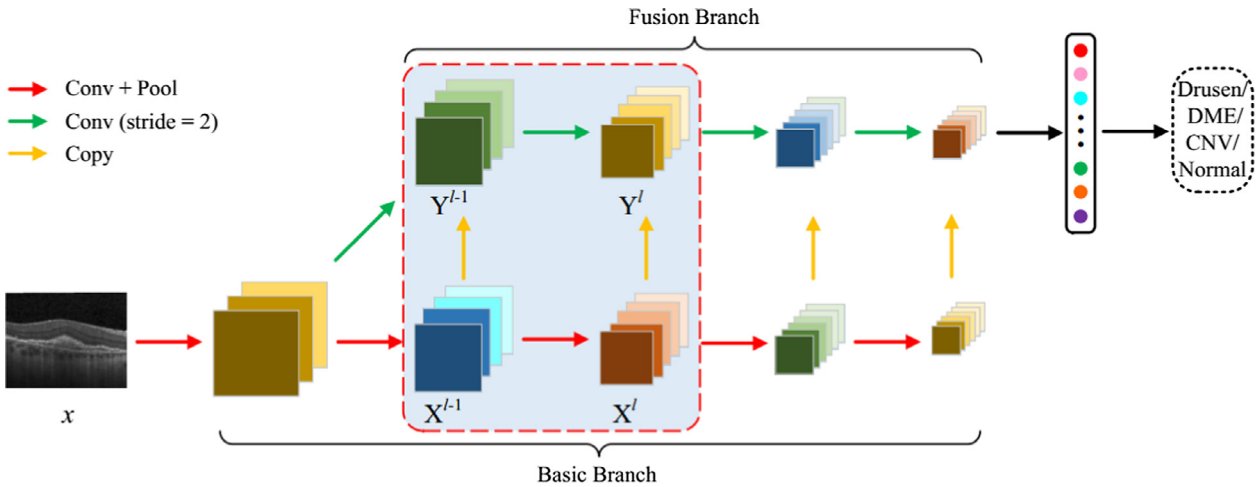

Fig. 4. The structure of the proposed IFCNN method for retinal OCT image classification.

The proposed IFCNN method is implemented using the Tensorflow framework [42] and the NVIDIA Cuda v8.0 and cuDNN v5.1 acceleration libraries, and is encoded in Python and MATLAB. All experiments were performed on an Ubuntu OS machine with CPU Intel Core i7-7700@3.60 GHz, GPU NVIDIA GeForce 1080Ti and 16 GB RAM.

4.3. Compared methods

The proposed IFCNN method is compared with three well-known classification methods: HOG-SVM [43], transfer learning [40] and VGG16 [38]. The HOG-SVM method extracts feature vectors from each B-scan using multi-scale histograms of directed gradient (HOG) descriptors and trains multiple binary SVMs for OCT classification. The other two approaches are based on deep leaning. The transfer learning method is based on the InceptionV3 architecture pre-trained on the ImageNet dataset [27] and is fine-tuned by freezing all convolutional layers and only retraining the final fully connected layer and identifying OCT images between the drusen, CNV, DME and normal macula. VGG16 is a general network for comparison purposes, consisting of multiple convolution and pooling layers and three fully connected layers. In our experiments, the parameters of VGG16 were set to be the same as IFCNN, as described in Section 4-B.

4.4. Results on UCSD datasets

Table 1 gives the quantitative results for the different methods on the UCSD dataset with 10-fold cross-validation. It can be observed that the three methods based on deep learning obtain better performance than the HOG-SVM method. In addition, by introducing an iterative fusion strategy, IFCNN has made significant improvements in all quantitative indicators compared to the transfer learning and VGG16 networks. Specifically, when using one-tenth of OCT data for training, our overall accuracy reached 87.3%. For each individual class of the Drusen, CNV, DME and Normal, the gain (sensitivity) of IFCNN relative to the counterpart without the iterative fusion strategy is about 11.8%, 2.8%, 8.1%

Table 1
Classification Results (in percentage) on UCSD dataset. Each of the first four rows gives the classification metrics for individual class Dursen, CNV, DME and Normal, respectively. The best results in the this table are labeled bold.

Methods	Classes	Sensitivity	Specificity	Accuracy	OA
HOG-SVM [43]	Drusen	29.8 ± 3.7	**96.9** ± 0.6	90.7 ± 0.2	77.8 ± 0.5
	CNV	87.1 ± 2.3	84.5 ± 1.3	85.2 ± 0.5	
	DME	53.3 ± 2.3	**96.6** ± 0.3	91.9 ± 0.4	
	Normal	90.8 ± 1.1	88.1 ± 1.2	89.6 ± 0.6	
Transfer Learning [40]	Drusen	58.7 ± 5.8	92.2 ± 1.8	88.8 ± 1.1	80.3 ± 1.1
	CNV	81.5 ± 4.8	92.9 ± 1.9	87.8 ± 1.1	
	DME	70.7 ± 3.7	94.6 ± 1	91.3 ± 0.5	
	Normal	89.8 ± 2.6	94 ± 1	92.7 ± 0.3	
VGG16 [38]	Drusen	65 ± 6	91.5 ± 2.4	88.8 ± 2.2	82.2 ± 2.5
	CNV	85.1 ± 3.6	94.8 ± 2	90.5 ± 1.8	
	DME	73.8 ± 6.3	96.2 ± 1.2	93.2 ± 1.4	
	Normal	87.5 ± 3.9	94.2 ± 1.7	92.1 ± 1.2	
IFCNN	Drusen	**76.8** ± 7.2	94.9 ± 1.9	**93** ± 1.7	87.3 ± 2.2
	CNV	**87.9** ± 4.3	**96** ± 1.7	**92.4** ± 1.3	
	DME	**81.9** ± 6.8	96.3 ± 2	**94.4** ± 1	
	Normal	**92.2** ± 4.7	**96** ± 1.6	**94.8** ± 1.2	

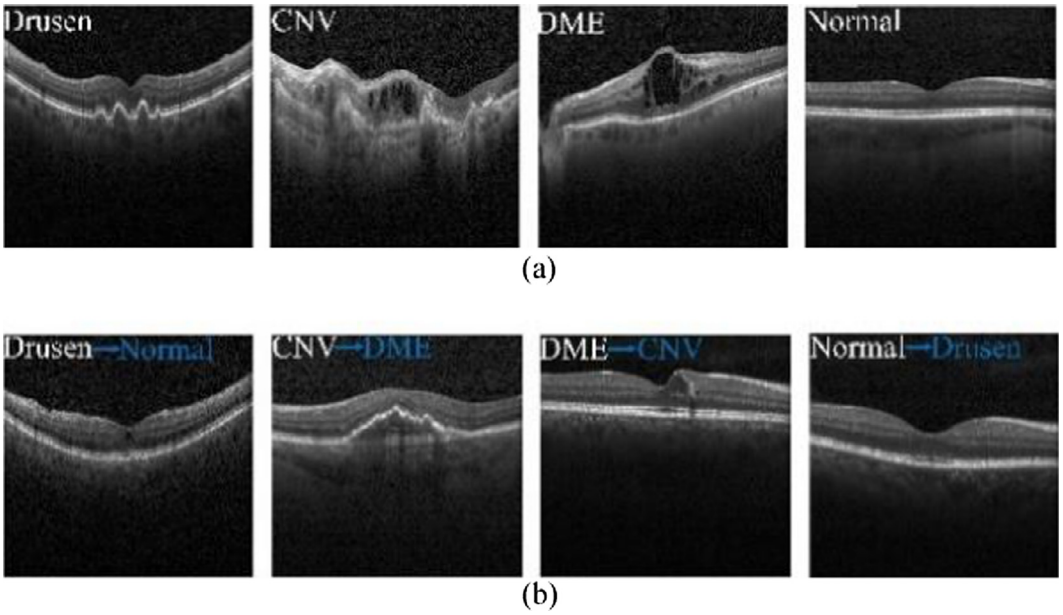

Fig. 5. Examples of classification results of the proposed IFCNN. (a) Correctly classification; (b) Misclassified cases.

and 4.7%, respectively, indicating the effectiveness of using an iterative fusion strategy for CNN-based OCT classification. In addition, IFCNN is also superior to the transfer learning method, and improvement on the overall accuracy is about 7%. This means that when only a very limited number of samples are available, the iterative fusion strategy can effectively utilize the information from different layers and enable the IFCNN to achieve high performance. Fig. 5 demonstrates some correctly classified (a) and misclassified (b) examples of the proposed IFCNN method. The white characters denote the true classes and the blue characters represent the misclassified classes.

4.5. Results on musculoskeletal radiographs (MURA) datasets

Another dataset is from the open source MURA dataset [44] of the Stanford University Wu Enda team and is can be downloaded at: https://stanfordmlgroup.github.io/projects/mura. All images were derived from 40,895 musculoskeletal X-rays from 14,982 cases. In 14,982 cases, there were 9067 normal upper musculoskeletal and 5915 upper extremity abnormal musculoskeletal X-rays, including the shoulder, tibia, elbow, forearm, wrist, palm and fingers. Each study was manually labeled as normal or abnormal by a radiologist. We evaluated the proposed IFCNN method on the MURA [44] dataset. As described in Section 4.2, the experimental parameters of IFCNN remain unchanged. As introduced in this dataset, 36,808 images are used for the training and 3197 images are used for the testing. The classification results are shown in Table 2. It can be observed that our proposed IFCNN is superior to the VGG16 method on the MURA datasets. For each individual class of the elbow, finger, forearm, hand, humerus, shoulder and wrist, the gains (sensitivity) of the proposed IFCNN method over the VGG16 method are about 3.7%, 2.8%, 4.9%, 2.9%, 7.6%, 5.2% and 4.4%, respectively, indicating that the effectiveness of the IFCNN method.

Table 2
Classification results (in percentage) of the IFCNN and VGG16 methods on MURA Datasets. The best results in this table are labeled in bold.

Category	VGG16 [38]	IFCNN
Elbow	75.7	**79.4**
Finger	61.6	**64.4**
Forearm	72.8	**77.7**
Hand	67.1	**70.0**
Humerus	67.4	**75.0**
Shoulder	65.5	**70.7**
Wrist	73.1	**77.5**
Total	69.1	**73.4**

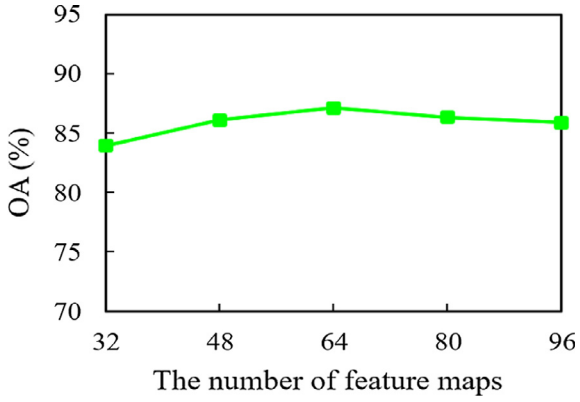

Fig. 6. Effect of different numbers of feature maps for IFCNN on UCSD dataset.

4.6. Effect of different numbers of feature maps

The effects of number of feature maps (described as the first convolutional layer) of our IFCNN method are analyzed on the UCSD dataset with cross validation. The parameters of IFCNN remain to be the same as in Section 4.2. In this experiment, the number of feature maps varied between 32, 48, 64, 80, and 96. The results are shown in Fig. 6. It can be observed that as the number of feature maps increases to 64, the performance of IFCNN method generally improve. In addition, if the number of feature maps further increases, the performance will degrade to some extent while still greatly creating more computational cost.

5. Conclusions

In this paper, we presented a novel iterative fusion convolutional neural network (IFCNN) method for retinal OCT image classification. To exploit the information among different convolutional layers, the proposed IFCNN method introduces an iteration layer fusion strategy. Specifically, we iteratively combine feature in current convolutional layer with those of all previous layers in the CNN network. Therefore, the feature information of all layers is iteratively combined, which can achieve better classification. Compared with several well-known classification methods, the proposed IFCNN method can make full use of hierarchical information among features of different convolutional layers. The classification results have shown that the proposed method outperform the other compared methods.

In this paper, we utilized the iterative fusion strategy for improving retinal OCT image classification. Therefore, in our future work, we will apply the proposed iterative fusion CNN model to OCT images from other pathologies.

Conflict of interest

There is no conflict of interest.

Acknowledgments

This work was supported in part by the National Natural Science Foundation under Grant No. 61771192, 61471167, and 61462089, the National Natural Science Foundation for Young Scientist of China under Grant No. 61501180, China Postdoctoral Science Foundation funded Project No. 2017T100597, and National Natural Science Foundation of Hunan Province Grant No. 2018JJ3077.

References

[1] A. Yuan, P.K. Kaise, Emerging therapies for the treatment of neovascular age related macular degeneration, Semin. Ophthalmol. 26 (3) (2011) 149–155.
[2] M. Thomas, S.S. Mousa, S.A. Mousa, Comparative effectiveness of aflibercept for the treatment of patients with neovascular age-related macular degeneration, Clin. Ophthalmol. 7 (11) (2013) 495–501.
[3] V.C.M. Del, P.L. Gehlbach, Ppar-α ligands as potential therapeutic agents for wet age-related macular degeneration, PPAR Res. 2008 (Special) (2008) 821592.
[4] K.B. Freund, L.A. Yannuzzi, J.A. Sorenson, Age-related macular degeneration and choroidal neovascularization, Am. J. Ophthalmol. 115 (6) (1993) 786–791.
[5] J.B. Jonas, I. Kreissig, B. Kamppeter, et al., Intravitreal triamcinolone acetonide for the treatment of intraocular edematous and neovascular diseases, Ophthalmology 83 (6) (2010) 645–663.
[6] F.E. Hirai, M.D. Knudtson, B.E.K. Klein, et al., Clinically significant macular edema and survival in type 1 and type 2 diabetes, Am. J. Ophthalmol. 145 (4) (2008) 700–706.
[7] D. Huang, E.A. Swanson, C.P. Lin, et al., Optical coherence tomography, Science 254 (5035) (1991) 1178–1181.
[8] C.A. Puliafito, M.R. Hee, C.P. Lin, et al., Imaging of macular diseases with optical coherence tomography, Ophthalmol. 102 (2) (1995) 217–229.

[9] W. Drexler, J.G. Fujimoto, State-of-the-art retinal optical coherence tomography, Prog. Retinal Eye Res. 27 (1) (2008) 45–88.
[10] P.H. Tomlins, R.K. Wang, Theory, developments and applications of optical coherence tomography, J. Phys. D Appl. Phys. 38 (15) (2005) 2519–2535.
[11] Y.Y. Liu, M. Chen, H. Ishikawa, et al., Automated macular pathology diagnosis in retinal OCT images using multi-scale spatial pyramid and local binary patterns in texture and shape encoding, Med. Image Anal. 15 (5) (2011) 748–759.
[12] G. Lemaître, M. Rastgoo, J. Massich, et al., Classification of SD-OCT volumes using local binary patterns: experimental validation for DME detection, J. Ophthalmol. 2016 (6) (2016) 3298606.
[13] A. Albarrak, F. Coenen, Y. Zheng, Age-related macular degeneration identification in volumetric optical coherence tomography using decomposition and local feature extraction, Proc. Med. Image Understand. Anal., 2013, pp. 59–64.
[14] P.P. Srinivasan, L.A. Kim, P.S. Mettu, et al., Fully automated detection of diabetic macular edema and dry age-related macular degeneration from optical coherence tomography images, Biomed. Opt. Exp. 5 (10) (2014) 3568–3577.
[15] L. Fang, C. Wang, S. Li, et al., Automatic classification of retinal three-dimensional optical coherence tomography images using principal component analysis network with composite kernels, J. Biomed. Opt. 22 (11) (2017) 11–16.
[16] Y. Sun, S. Li, Z. Sun, Fully automated macular pathology detection in retina optical coherence tomography images using sparse coding and dictionary learning, J. Biomed. Opt. 22 (1) (2017) 12–21.
[17] W. Yu, Y. Zhang, Z. Yao, et al., Machine learning based detection of age-related macular degeneration (AMD) and diabetic macular edema (DME) from optical coherence tomography (OCT) images, Biomed. Opt. Exp. 7 (12) (2016) 4928–4940.
[18] S. Farsiu, S.J. Chiu, R.V. O'Connell, et al., Quantitative classification of eyes with and without intermediate age-related macular degeneration using optical coherence tomography, Ophthalmology 121 (1) (2014) 162–172.
[19] A.F.G. Venhuizen, B.V. Ginneken, B. Bloemen, et al., Automated age-related macular degeneration classification in OCT using unsupervised feature learning, Proc. SPIE Med. Imag., 2015, pp. 94141I-94141I.
[20] L. Fang, S. Li, D. Cunefare, et al., Segmentation based sparse reconstruction of optical coherence tomography images, IEEE Trans. Med. Imag. 36 (2) (2016) 407–421.
[21] L. Fang, S. Li, Q. Nie, et al., Sparsity based denoising of spectral domain optical coherence tomography images, Biomed. Opt. Exp. 3 (5) (2012) 927–942.
[22] J. Shi, J. Malik, Normalized cuts and image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 22 (8) (1997) 888–905.
[23] A.N. Kuo, R.P. Mcnabb, S.J. Chiu, et al., Correction of ocular shape in retinal optical coherence tomography and effect on current clinical measures, Am. J. Ophthalmol. 156 (2) (2013) 304–311.
[24] Y. Lecun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436.
[25] G.E. Hinton, R.R. Salakhutdinov, Reducing the dimensionality of data with neural networks, Science 313 (5786) (2006) 504–507.
[26] O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional networks for biomedical image segmentation, in: MICCAI, 2015, pp. 234–241.
[27] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, in: Proc. Adv. Neural Inform. Process. Syst., 2012, pp. 1097–1105.
[28] L. Fang, C. Wang, S. Li, et al., Automatic segmentation of nine retinal layer boundaries in OCT images of non-exudative AMD patients using deep learning and graph search, Biomed. Opt. Exp. 8 (5) (2017) 2732–2744.
[29] A.G. Roy, A. Katouzian, C. Wachinger, et al., ReLayNet: retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional networks, Biomed. Opt. Exp. 8 (8) (2017) 3627–3642.
[30] Y. Xu, K. Yan, J. Kim, et al., Dual-stage deep learning framework for pigment epithelium detachment segmentation in polypoidal choroidal vasculopathy, Biomed. Opt. Exp. 8 (9) (2017) 4061–4076.
[31] L.H.N. Sheck, Deep learning is effective for classifying normal versus age-related macular degeneration OCT images, Ophthalmol. Retina 2 (2) (2017) 322–327.
[32] R. Rasti, H. Rabbani, A. Mehridehnavi, et al., Macular OCT classification using a multi-scale convolutional neural network ensemble, IEEE Trans. Med. Imag. 37 (4) (2017) 1024–1034.
[33] J. Zhu, L. Fang, P. Ghamisi, Deformable convolutional neural networks for hyperspectral image classification, IEEE Trans. Geosci. Remote lett. PP (99) (2018) 1–5.
[34] S. Ren, K. He, R. Girshick, et al., Faster R-CNN: towards real-time object detection with region proposal networks, Proc. Adv. Neural Inform. Process. Syst., vol. 39, 2015, pp. 91–99.
[35] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, Proc. Adv. Neural Inform. Process. Syst., vol. 60, 2012, pp. 1097–1105.
[36] E. Shelhamer, J. Long, T. Darrell, Fully convolutional networks for semantic segmentation, IEEE Trans. Med. Imag. PP (99) (2014), 1-1.
[37] L. Fang, N. He, S. Li, et al., Extinction profiles fusion for hyperspectral images classification, IEEE Trans. Geosci. Remote Sens. 56 (3) (2018) 1803–1815.
[38] K. Simonyan, A. Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition, (Online). Avaliable: https://arxiv.org/abs/1409.1556, 2014.
[39] C. Szegedy, W. Liu, Y. Jia, et al., Going deeper with convolutions, in: Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit., 2015, pp. 1–9.
[40] D.S. Kermany, M. Goldbaum, W. Cai, et al., Identifying medical diagnoses and treatable diseases by image-based deep learning, Cell 172 (5) (2018) 1122–1131.e9.
[41] D. Kingma, J. Ba, Adam: A Method for Stochastic Optimization, (Online). Available: https://arxiv.org/abs/1412.6980, 2014.
[42] M. Abadi, P. Barham, J. Chen, et al., TensorFlow: a system for large-scale machine learning, (Online). Available: https://arxiv.org/abs/1605.08695, 2016.
[43] G.M. Comer, J.A. Izatt, L.A. Kim, et al., Fully automated detection of diabetic macular edema and dry age-related macular degeneration from optical coherence tomography images, Biomed. Opt. Exp. 5 (10) (2014) 3568–3577.
[44] P. Rajpurkar, J. Irvin, A. Bagul, et al., MURA: Large Dataset for Abnormality Detection in Musculoskeletal Radiographs, https://arxiv.org/abs/1712.06957, 2017.