Computer Methods and Programs in Biomedicine 219 (2022) 106739

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

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

BFENet: A two-stream interaction CNN method for multi-label ophthalmic diseases classification with bilateral fundus images

Xingyuan Ou[a,1], Li Gao[b,1], Xiongwen Quan[a,*], Han Zhang[a], Jinglong Yang[a], Wei Li[a]

[a] College of Artificial Intelligence, Nankai University, Tianjin, China
[b] Ophthalmology, Tianjin Huanhu Hospital, Tianjin, China

ARTICLE INFO

Article history:
Received 15 December 2021
Revised 23 February 2022
Accepted 7 March 2022

Keywords:
Ocular disease classification
Feature enhancement
Patient-level diagnosis
Multi-label
Convolutional neural network

ABSTRACT

Background and objective: Early fundus screening and timely treatment of ophthalmology diseases can effectively prevent blindness. Previous studies just focus on fundus images of single eye without utilizing the useful relevant information of the left and right eyes. While clinical ophthalmologists usually use binocular fundus images to help ocular disease diagnosis. Besides, previous works usually target only one ocular diseases at a time. Considering the importance of patient-level bilateral eye diagnosis and multi-label ophthalmic diseases classification, we propose a bilateral feature enhancement network (BFENet) to address the above two problems.

Methods: We propose a two-stream interactive CNN architecture for multi-label ophthalmic diseases classification with bilateral fundus images. Firstly, we design a feature enhancement module, which makes use of the interaction between bilateral fundus images to strengthen the extracted feature information. Specifically, attention mechanism is used to learn the interdependence between local and global information in the designed interactive architecture for two-stream, which leads to the reweighting of these features, and recover more details. In order to capture more disease characteristics, we further design a novel multiscale module, which enriches the feature maps by superimposing feature information of different resolutions images extracted through dilated convolution.

Results: In the off-site set, the Kappa, F_1, AUC and Final score are 0.535, 0.892, 0.912 and 0.780, respectively. In the on-site set, the Kappa, F_1, AUC and Final score are 0.513, 0.886, 0.903 and 0.767 respectively. Comparing with existing methods, BFENet achieves the best classification performance.

Conclusions: Comprehensive experiments are conducted to demonstrate the effectiveness of this proposed model. Besides, our method can be extended to similar tasks where the correlation between different images is important.

1. Introduction

Retinal damage caused by ophthalmic diseases can result in vision loss and even blindness [1–3]. Among them, diabetic retinopathy, age-related macular degeneration (AMD), cataracts and glaucoma are the most common ocular diseases. According to the study [4], there will be more than 400 million DR patients and 80 million glaucoma patients by 2030. These fundus diseases have become a serious health problem over the world, and timely diagnosis to prevent blindness is extremely urgent. In the early diagnosis of ocular diseases, fundus photography and optical coherence tomography (OCT) are proved to be effective imaging techniques. Compared with expensive OCT images, fundus photography is a cost-effective and non-invasive method for ocular diseases screening. As a result, fundus photographs have become the primary method used by ophthalmological experts to detect diseases such as diabetic retinopathy, AMD, myopia, hypertension, glaucoma and cataract [5]. However, artificial fundus photography inspection is time-consuming and laborious. In most backward areas, the number of ophthalmologists is far from sufficient for manual diagnosis. In order to reduce the pressure of ophthalmologists and improve the accuracy of fundus diagnosis, an automatic diagnosis algorithm for ophthalmic disease screening is in urgent need.

Recently, deep learning models have achieved outstanding performance in many computer vision application [6–8]. Convolutional neural network (CNN) as a commonly used deep learning model, has made great progress and played a significant role in medical imaging field [9–11]. Owing to the powerful ability of high-level feature extraction and represen-tational capability, CNNs have been widely used in ocular disease diagnosing such as glau-

* Corresponding author.
E-mail address: quanxw@nankai.edu.cn (X. Quan).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.cmpb.2022.106739
0169-2607/

coma screening [12], optic disc segmentation [13] and retinal blood vessel segmentation [14]. With respect to fundus disease classification, CNNs have also achieved promising performance. Gulshan et al. classified fundus photography according to diabetic retinopathy grade [15]. Li et al. conducted the classification of glaucomatous optic neuropathy [16]. Transfer learning with pretrained models by ImageNet dataset, such as the ResNet networks, was found to be an effective method in the task of ophthalmic disease classification [17]. By ensembling multi-ple base networks, the model can achieve better classification performance [18]. These studies illustrate the advantage of applying CNNs to achieve specificity and high sensitivity in the classification of ocular diseases.

Although the CNN based models have achieved encouraging performance for fundus diseases screening, there are still some limitations and new challenge. For instance, few works consider the problem of multi-label ophthalmic disease classification with fundus photograph [19]. Since many patients are affected by multiple fundus diseases, it is necessary to improve models to adapt multi-label ocular diseases. Li et al. discovered the coexistence of myopia could cause false-negative prediction for glaucoma classification [20]. Therefore, many existing models with satisfactory results on specific tasks may not be applicable on the actual complex situations. In addition, the majority of existing CNN models just analyze the fundus images of single eye rather than both the left and right eyes for the classification task. But in most clinical scenarios, ophthalmologists diagnose patients based on information from both eyes. Publications have shown that bilateral eyes are correlated closely in terms of ophthalmic disease progression [21], which suggests that ophthalmic patient diagnosis considering the information from bilateral fundus images should be more effective approach.

On a dataset named ODIR which includes images from bilateral eyes and seven normal diseases, some researchers have achieved certain results [22–26]. Since the dataset did not publish their test set, the researchers had to evaluate the results of their own models on the limited training set. To this end, Li et al. released a benchmark dataset named OIA-ODIR for ocular disease recognition [27]. Different from ODIR datasets, the training set and test set of OIA-ODIR dataset are separated and available. Li et al. compared three fusion methods and verified the effectiveness of their model on the test set. Smitha et al. proposed a semi-supervised GAN learning model with enhanced images to fuse photographs from both eyes [28]. However, these models did not enhance the extracted feature information, which decreases the preciseness and robustness of the model.

In this paper, we design a bilateral feature enhancement network (BFENet) to tackle the problem of multi-label ophthalmic disease classification based on binocular fundus images. The input to BFENet is pairs of fundus photographs obtained from both the left and right eyes. The output are the predicted results for different ocular diseases to testing sample. BFENet is composed of four modules as follows. The backbone CNN module is used to extract features from both the single left and right fundus images. Then the multiscale module is used for extraction of multiscale features. Subsequently, we devise the feature enhancement module for feature enhancement and exchange. Finally, the classifier module is used to generate classification results. As the experiments show, BFENet achieves outstanding perfor-mance on a public fundus images dataset. At the same time, we explored the impact of network depth on performance.

In summary, our contributions of this research are as follows: (1) We propose a CNN model with two-stream interactive architecture for the patient-level multi-label ophthalmic disease classification task. Eight types of ophthalmic diseases (Eight types in Fig. 1) can be processed simultaneously through a single network. (2) We design a feature enhancement module by using the attention mechanism to learn the interdependence between local and global information, which enhances the extracted feature under the two-stream interactive architecture. (3) A multiscale module is designed to enrich the feature maps by superimposing feature information of different resolutions images extracted through dilated convolution. (4) Comparing with existing state-of-the-art methods, BFENet achieves the best performance.

The remainder of this paper is organized as follows. Section II discusses related work. Section III presents our multi-label ophthalmic disease classification method. In section IV, we introduce the evaluation criteria and discuss the results. Finally, we conclude the work in Section V.

2. Related works

2.1. CNNs for medical image analysis

In recent years, the application of CNN has made remarkable progress in the field of medical image analysis. Specifically, CNNs are trained as end-to-end feature extractors, which can directly recognize subtle features from fundus photographs without specific domain knowledge or human power. Bravo et al. proposed a method based on transfer learning, using the VGG16 architecture to detect diabetic retinopathy [29]. A six-level cataract grading method focusing on multi-feature fusion was described by Zhang et al. [30], which extracts features from gray level co-occurrence matrix and residual network. Hong et al. developed a deep CNN model with 14 layers, which can diagnose early AMD disease accurately and aid ophthalmologists in fundus screening [31]. However, they have not pay attention to the internal interaction between two branches and multi-scale features which can promote ability of feature representation.

2.2. Multi-label classification in ophthalmic diseases

As for multiple ocular diseases recognition, Koh et al. proposed a private ophthalmic photographs dataset and his model makes use of accelerated robust features and directional gradient pyramid histogram features to solve the classification of diabetic retinopathy, glaucoma and AMD [32]. Chelaramani et al. performed three tasks on a private fundus disease dataset, including 320 fine-grained classification, four common diseases classification and text diagnosis of generation [33]. On a public available dataset named Singapore Malay Eye Study, Chen et al. performed multiple classification of three common ocular diseases on an image by entropic graph regularized probabilistic learning [34]. But the number of diseases they can recognize is still insufficient to satisfy actual needs.

Recently, people have explored to use the left and right fundus images as input, but only using the effective feature extraction is insufficient. Islam et al. [22] proposed a shallow CNN model for ODIR database and their model was trained from scratch. But they did not propose a model for classification of multiple diseases with pair of images simultaneously. Jordi et al. [23] proposed two pre-trained architectures, InceptionV3 and Vgg16, for classification task on ODIR database. They transformed the classification task on ODIR database into a multi-class classification problem. Gour and Khanna [24] proposed a transfer learning based method, which concatenates both pair of fundus images together and input them to a CNN network. He et al. [25,26] proposed a Dense Correlation Network (DCNet), which can exploit the spatial correlations between the pair of fundus photographs. But they did not use local features and global features for contrastive learning to enhance feature information.

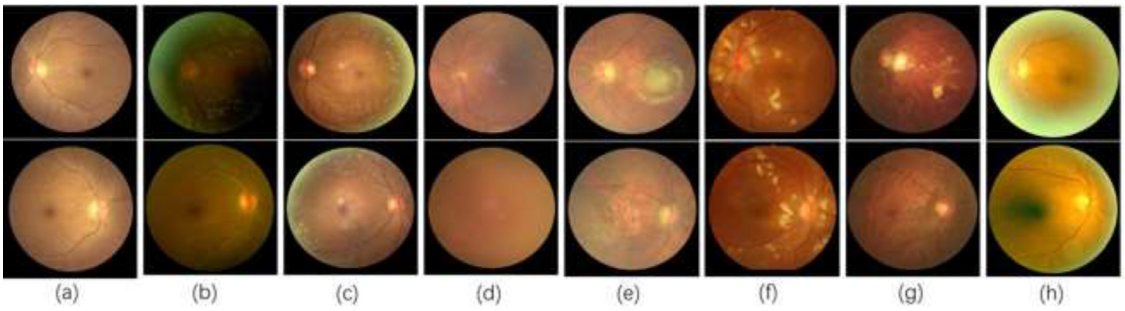

Fig. 1. Examples of fundus photographs in both left (the first row) and right (the second row) eyes. The diagnostic keywords for each column are as follows [left; right] (a) Normal; Normal (b) diabetes; other Abnormalities (c) glaucoma; others (d) cataract; cataract (e) age-related macular degeneration; glaucoma (f) hypertension; normal (g) Myopia; Myopia (h) other Abnormalities; other Abnormalities.

Table 1
Classification results of different backbone CNN architecture.

Backbone CNN	Off-site				On-site			
	Kappa	F_1	AUC	Final	*Kappa*	F_1	*AUC*	*Final*
Vgg16	0.452	0.876	0.871	0.733	0.441	0.874	0.873	0.730
InceptionV3	0.447	0.856	0.877	0.727	0.448	0.858	0.869	0.724
DenseNet	0.391	0.838	0.847	0.692	0.397	0.839	0.846	0.694
ResNet18	**0.506**	**0.883**	**0.894**	**0.761**	**0.482**	**0.878**	**0.889**	**0.750**

2.3. Attention mechanism in ocular diseases classification

Attention mechanism is widely used in computer vision [35] and bioinformatics [36]. Previous works have adopted attention mechanism for ocular disease classification. Wang et al. implemented an attention based CNN model for simultaneously diagnosing diabetic retinopathy and capturing salient regions from small high resolution patches with image-level labels learning based method for the classification of diabetic retinopathy [37]. He et al. proposed category attention blocks and global attention blocks for imbalanced Diabetic Retinopathy grading [38]. But they have not used attention mechanism for two-stream interaction to enhance features.

3. Material and methods

3.1. Network architecture

The architecture of the proposed network is illustrated in Fig. 2a. BFENet includes four major modules: backbone CNN module, multiscale module, feature enhancement module, and the final classifier. In this section, we will explain the four parts in detail as follow.

3.1.1. Backbone CNN module

The backbone CNN module extracts two sets of independent features from the input pairs of fundus photographs. Given both the left and right fundus photographs, P_l and P_r (P_l, P_r $\in R^{h \times w \times 3}$, h and w refer to the height and width of the given fundus photographs, 3 is the three color channels), and the outputs of backbone CNN module are F_l and F_r (F_l, F_r $\in$ $R^{H \times W \times C}$, where H, W and C refer to the height, width, and the number of extracted features). The classification results of different backbone CNN architecture are shown in Table 1. Since ResNet achieves the best result, we adopt it as our backbone CNN module. Resnet-18, Resnet34 and Resnet50 are applied as our backbone CNN module to explore the effects of different network depths.

3.1.2. Multiscale module

To enrich the extracted information, multiscale module is designed to obtain multiscale features. As shown in Fig. 2a, we use a 3×3 dilated convolution [39] and a 1×1 convolution to transfer the feature map F_b from the previous backbone CNN module into F_d and F_c (F_d,F_c $\in R^{H \times W \times C'}$, where $C' = C/2$), convolution with a LeakyRelu activation and dilated convolution without. Compared with convolution layer, dilated convolution can capture different receptive fields information and then get the feature map of different resolution by setting dilation growing rates. Combine the feature map of original resolution obtained by the convolutional layer with the feature set of different resolution and then get the multi-scale information.

$$F_m = Cat(DConv(F_b), LeakyRelu(Conv(F_b))) \quad (1)$$

Where *Cat* represents the concatenation operation, *DConv* and *Conv* denote 3×3 dilated convolution and 1×1 convolution, respectively.

3.1.3. Feature enhancement module

The feature enhancement module is used for highlighting the features of each eye in the global information, so as to facilitate the exchange of complementary information, and recover more details of each independent eye. We use attention mechanism to learn the interdependence between local and global information in the designed interactive architecture for two-stream, which leads to the reweighting of these features, and further enhance feature information.

The feature enhancement module receives two feature sets from the previous layer as input, and outputs two enhanced feature maps by considering the correlation between individual information and global information. The details of the proposed feature enhancement module are shown in Fig. 2b. With the two input feature maps of F_{ml} and F_{mr}, we concatenate them together to get the global information F_{mg} from bilateral fundus images. Inspired by cross parallax attention network [40] and non-local neural networks [41], we firstly transform the two feature maps from each independent fundus image into query (Q_l for the feature of left fundus image and Q_r for the feature of right fundus image, where Q_l, Q_r $\in$ $R^{HW \times C''}$, C'' $=$ $C/4$) by 1×1 convolution. Next, we apply a 1×1 convolution layer on F_{mg} for channel reduction and acquire global feature map F_g. Lastly, we transform F_g into key and value (K and V for the global feature of both eyes, where K $\in$

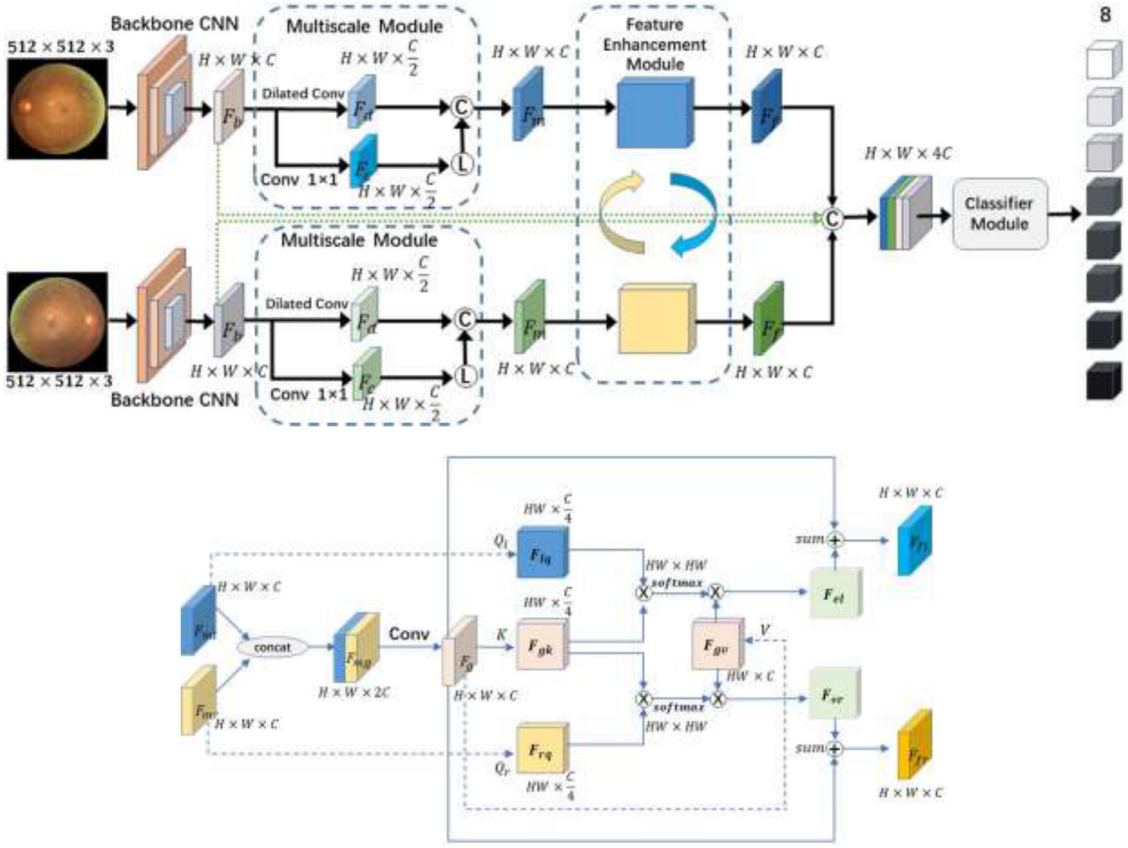

Fig. 2. (a) Overall structure of BFENet. "L"is the LeakyRelu function, "C" is concatenation. (b) Detailed architecture of the proposed feature enhancement module.

$R^{HW \times C'}$ and $V \in R^{HW \times C}$, $C' = C/4$) by 1×1 convolution. Take the inputted images size is 448 and Resnet-50 backbone CNN as example, the H, W and C is 14, 14 and 2048.

$$Q_l = Conv(F_{ml}; \theta_{ml}), \qquad Q_r = Conv(F_{mr}; \theta_{mr}), \tag{2}$$

$$F_g = Conv(Concat(F_{ml}, F_{mr}); \theta_{mg}), \tag{3}$$

$$K = Conv(F_g; \theta_g), \tag{4}$$

$$V = Conv(F_g; \theta_g). \tag{5}$$

Where Conv represents 1×1 convolution, and θ referring to the relevant parameters.

In order to calculate pixel-related correlation weights between the feature map of left and global. As shown in Fig. 2b, the weight assigned to each value is calculated as the inner product of the query with the key. We compute the matrix of F_{el} for the left and F_{er} for the right as:

$$F_{el} = Softmax\left(\frac{Q_l K^T}{\sqrt{d_k}}\right)V \tag{6}$$

$$F_{er} = Softmax\left(\frac{Q_r K^T}{\sqrt{d_k}}\right)V \tag{7}$$

Where d_k represents the dimension of the key.

The last step of feature enhancement module is sum the feature map of F_g and F_{el} to get the enhanced feature map F_{fl}, by the same way we can get F_{fr}.

$$F_{fl} = Sum(F_g, F_{el}) \tag{8}$$

$$F_{fr} = Sum(F_g, F_{er}) \tag{9}$$

Where F_{fl}, $F_{fr} \in R^{H \times W \times C}$ are the outputs of feature enhancement module.

3.1.4. Classifier module

Before input to the final classifier module, the two output feature maps from the feature enhancement module and two feature sets from the backbone CNN module are concatenated. Compared with sum and prod operation, concatenation preserves individual module independent features and provides more information. The classifier module includes three fully-connected layers, the second fully-connected layer with Relu activation and others without. The first two fully-connected layers reduce the dimension of the concatenated features. The exact feature size depends on the utilized backbone CNN. Take the Resnet-50 backbone CNN as example, the dimension of the concatenated feature is 8192. It is reduced to 4096 and 512 by the first two fully-connected layer. The last fully-connected layer reduces the features to eight dimensions, which is equivalent to the classification category. The eight dimension features can be compared with the disease classification labels, and the network loss of the network can be calculated accordingly.

3.2. Transfer Learning

In the multi-label ophthalmic disease classification task, we apply the pre-trained CNN network as a new starting point, where the CNN network with fine-tuning is faster and easier to train than training the network from scratch. The ResNet model is pretrained on the ImageNet dataset, which can divide more than one million natural images into 1000 categories. In this multi-label ophthalmic disease classification task, we use ResNet backbone CNN after transfer learning to classify fundus photographs and compare

Table 2
Distribution of per category patient cases in training and testing datasets.

Labels	N	D	G	C	A	H	M	O
Training Set case	1138	1130	215	212	164	103	174	982
Off-site Testing Set cases	162	163	32	31	25	16	23	136
On-site Testing Set cases	324	327	58	65	49	30	46	275
All cases	1624	1620	305	308	238	149	243	1393

it with the method without transfer learning. The method with transfer learning achieves better classification accuracy.

3.3. Loss function

The cross-entropy loss is widely employed in various classification tasks, but it is not applicable in categories associated task. Thus, according to similar existing research [42], we employ a binary cross-entropy loss $L(y_i, \widehat{y_i})$ in Eq. (10) to solve the multi-label ocular disease classification task.

$$L(y_i, \widehat{y_i}) = -\frac{1}{N}\sum_{i=0}^{N}(y_i \log(\widehat{y_i}) + (1-y_i)\log(1-\widehat{y_i})) \tag{10}$$

where N refers to the total number of categories, y_i is the reference label, and $\widehat{y_i}$ is the predicted label of the network.

3.4. Dataset

We use Ophthalmic Image Analysis-Ocular Disease Intelligent Recognition (OIA-ODIR) dataset [27] to evaluate the BFENet. To our best knowledge, OIA-ODIR dataset is the first internationally multi-type disease detection dataset based on binocular fundus image, which is publicly available and compiled by Shanggong Medical Technology Co. The ground truth of the dataset was annotated by experienced ophthalmologists, and it took over 10 months to complete. Any differences was resolved through negotiation until all annotators reached an agreement. The OIA-ODIR dataset contains 10,000 fundus photographs with eight types of annotations from 5000 clinical patients. At the same time, the data set was divided into three parts, the training set, the off-site test set and the on-site test set, which contain 3500, 500 and 1000 patients respectively. OIA-ODIR is a multi-category and multi-label dataset, which contains 8 ocular classification categories, including normal image (N), diabetes retinopathy (D), glaucoma (G), cataract (C), age-related macular degenerate (A), hypertension (H), myopia (M), and other abnormalities (O). The distribution of the 5000 patient cases per category in training and testing datasets is shown in Table 2. These cases are used to check the effectiveness of our proposed model. The provided training set is split into our training set and our validation set, containing 80% and 20% of the original training set respectively.

3.5. Implementation detail

Since the OIA-ODIR dataset was collected from different hospitals with different cameras, In order to ensure all input images have the same resolution, we firstly resize the original images into the same image resolution of 512 × 512. Then random cropping of 448 × 448 image patches, random horizontal flips and vertical flips are used as data augmentation to reduce overfitting. During the process of testing, we use center cropping to ensure the cropped images are in central position, so as to achieve a better classification result.

The publicly available framework Pytorch [43] was used to implement all our deep neural networks. All the experiments ran on NVIDIA Tesla P100 GPUs. The stochastic gradient decent (SGD) optimizer is applied to train the networks. Initial learning rate is set to 0.005 and decayed based on poly learning rate decay policy $lr = initial_{lr} \times (1 - \frac{iter}{total_{iter}})^{power}$ [44]. According to our experience the power is set as 0.9. The number of iterations during all our experiments was 50 epochs.

4. Experimental results and analysis

This section describes the evaluation criteria at the beginning. Then we conducted two sets of experiments, the first is ablation experiments, the second is the influence of different depths to backbone CNN on the model. Lastly, the experimental results are compared and analyzed.

4.1. Evaluation Criteria

In order to evaluate the performance provided by the proposed method, we choose four evaluation metrics to evaluate the classification performance, such as Kappa score (K in Eq. (11)), F1-score (F1 in Eq. (12)), area under curve (AUC in Eq. (13)), and their average value (Final-score in Eq. (14)). Kappa coefficient is used to check consistency, and the range is -1 to 1. F1-score is the harmonic average value of recall rate and precision, which is higher only when the recall rate and precision are both high. Since Kappa score and F1 score only consider a single threshold of the result, while the output result of the classification network is probabilistic, we use AUC to consider multiple thresholds. All these four evaluation metrics are calculated by official sklearn package.

The kappa coefficient is calculated by follows equation:

$$\begin{aligned} K &= \frac{p_o - p_e}{1 - p_e} \\ p_o &= \frac{\sum_{i=1}^{r} X_{ii}}{N} \\ p_e &= \frac{\sum_{i=1}^{r} X_{i+} \times X_{+i}}{N^2} \end{aligned} \tag{11}$$

where $r = 8$ represents the number of rows in the confusion matrix, N represents the number of all samples, X_{ii} represents the number in column i and row i of the confusion matrix, X_{i+} represents the total number of the i-th row, and X_{+i} represents the total number of the i-th column.

$$F_1 = \frac{2TP}{2TP + FN + FP} \tag{12}$$

$$AUC = \int_{x=0}^{1} TPR(FPR^{-1}(x))dx$$

$$TPR = \frac{TP}{TP + FN}, \quad FPR = \frac{FP}{FP + TN} \tag{13}$$

$$Final = \frac{Kappa + F1 + AUC}{3} \tag{14}$$

where TP, FP, TN, FN, TPR and FPR refer to true positive, false positive, true negative, false negative, true positive rate and false positive rate, respectively.

Table 3
Computational parameters of different network with/without BFE (bilateral feature enhancement).

Backbone	BFE	FLOPs (G)	Params (M)
ResNet18	Without	14.6	11.7
ResNet18	With	31.2	61.5
ResNet34	Without	29.4	21.8
ResNet34	With	60.1	78.2
ResNet50	Without	33.0	25.8
ResNet50	With	67.0	82.6

Table 4
Class-wise accuracy performance of BFENet.

Class	Off-site ACC	On-site ACC
Normal	0.67	0.65
Glaucoma	0.75	0.73
diabetes retinopathy	0.90	0.89
AMD	0.90	0.90
Hypertension	0.94	0.93
Cataract	0.96	0.95
Myopia	0.92	0.90
Other abnormalities	0.62	0.61

4.2. Ablation studies on our proposed methods

4.2.1. Classification results of Resnet50 backbone CNN after transfer learning

First, we compared the backbone CNN model of Resnet50 and Resnet50 after transfer learning in the multi-label ophthalmic disease classification task. The backbone CNN model without transfer learning is widely used in many classification tasks. Secondly, we compared the way of inputting the concatenated left and right images to a CNN network with the way of inputting left and right images to two CNN network separately. The results are shown in Table 5. It can be observed that ResNet50 after ImageNet pre-training without concatenated both images achieves better classifica-tion result. In the result of off-site test set, Kappa, F_1, AUC and Final score after transfer learning increases 0.333, 0.032, 0.091 and 0.152, respectively. In the result of on-site test set, Kappa, F_1, AUC and Final score after transfer learning increases 0.322, 0.029, 0.075 and 0.142, respectively.

4.2.2. The effect of multiscale module on the classifier

A multi-scale feature map is constructed by concatenating the different resolution feature map extracted through the dilated convolution with original resolution feature map. Multi-scale features can enrich the feature information extracted by backbone CNN. As shown in Table 5, in the Off-site test set, compared with the results generated by the complete BFENet, Kappa, F_1, AUC and Final score declined 0.008, 0.003, 0.005 and 0.006 respectively, when the multiscale module is removed. In the On-site test set, compared with the results generated by the complete BFENet, Kappa, F_1, AUC and Final score declined 0.006, 0.004, 0.005 and 0.005 respectively, when the multiscale module is removed..

4.2.3. The effect of feature enhancement module on the classifier

After considering the relationship between global information and independent information, feature enhance-ment module highlight the characteristics of each eye. As shown in Table 5, in the Off-site test set, compared with the results generated by the complete BFENet, Kappa, F_1, AUC and Final score declined 0.016, 0.005, 0.008 and 0.010 respectively, when the feature enhancement module is removed. In the On-site test set, compared with the results generated by the complete BFENet, Kappa, F_1, AUC and Final score declined 0.010, 0.006, 0.012 and 0.009 respectively, when the feature enhancement module is removed.

4.3. Classification performance and computational parameters

Tables 6 and 7 list the detailed results of ResNet CNNs with different depths. Three phenomena can be observed. First of all, among different fusion strategies, the concatenation works best. Element multiplication is more appropriate when deep backbone CNNs is used. Secondly, using deeper backbone CNNs can achieve better performance. Comparing the results in Off-site set of model with ResNet-50 and ResNet-18, the Kappa, F_1, AUC and Final increase by 0.009, 0.005, 0.012 and 0.009, respectively. In On-site test dataset, the Kappa, F_1, AUC and Final increase by 0.011, 0.005, 0.08 and 0.008, respectively. This indicates higher abstraction features have a better fundus disease distinction ability. Finally, the enhancement is limited when ResNet-34 replaced by ResNet-50. Many studies show similar situation that network cannot improve performance by linear increase of network depth [45]. This phenomenon may be caused by three reasons. The first is related to the gradient vanishing issue. The optimization difficulty increases with network depth [45]. The second reason is the reduction of feature reuse, which leads to insufficient use of a large number of features generated by CNNs [46]. Lastly, since the size of training dataset is limited, the network may not be properly trained. Table 4 lists the class-wise accuracy performance of BFENet. It can be observed from the table that the accuracy is higher for the minority class. The reason is that the number of negative samples is much higher than the number of positive samples for these minority classes. Table 3 lists the computational parameters of different models. It can be observed from the table that the introduction of BFE (bilateral feature enhancement) will increase the computational parameters of the model and the computational complexity is highest when ResNet50 is used as the backbone.

Table 5
Ablation experiments performed on the multi-label ocular disease classification.

Test Set Name	Performance Metric	Ablation experiments performed on the multi-label ocular disease classification					
		Resnet50	Resnet50 With Transfer Learning	Resnet50 With Transfer Learning With Concatenated Both Images	Our Method Without Multiscale Module	Our Method Without Feature Enhancement Module	Our Method
Off-site	*Kappa*	0.177	0.510	0.468	0.527	0.519	**0.535**
	F_1	0.853	0.885	0.863	0.889	0.887	**0.892**
	AUC	0.809	0.900	0.871	0.907	0.904	**0.912**
	Final	0.613	0.765	0.734	0.774	0.770	**0.780**
On-site	*Kappa*	0.172	0.494	0.457	0.507	0.503	**0.513**
	F_1	0.848	0.878	0.856	0.882	0.880	**0.886**
	AUC	0.812	0.887	0.862	0.898	0.891	**0.903**
	Final	0.611	0.753	0.725	0.762	0.758	**0.767**

Table 6
Classification results of different backbone CNN.

Backbone CNN	Fusion	Off-site				On-site			
		Kappa	F_1	AUC	Final	*Kappa*	F_1	*AUC*	*Final*
ResNet-18	Sum	0.499	0.882	0.895	0.759	0.480	0.877	0.889	0.749
	Prod	0.484	0.881	0.892	0.752	0.479	0.875	0.880	0.745
	Concat	0.506	0.883	0.894	0.761	0.482	0.878	0.889	0.750
ResNet-34	Sum	0.505	0.879	0.897	0.760	0.484	0.875	0.883	0.747
	Prod	0.503	0.877	0.894	0.758	0.470	0.867	0.880	0.739
	Concat	0.512	0.882	0.898	0.764	0.490	0.875	0.890	0.752
ResNet-50	Sum	0.506	0.884	0.899	0.763	0.487	0.873	0.884	0.748
	Prod	0.510	0.885	0.900	0.765	0.494	0.878	0.887	0.753
	Concat	0.507	0.883	0.898	0.763	0.488	0.874	0.885	0.749

Table 7
Classification results of different backbone CNN with BFE.

Backbone CNN	Off-site				On-site			
	Kappa	F_1	AUC	Final	*Kappa*	F_1	*AUC*	*Final*
ResNet-18	0.526	0.887	0.900	0.771	0.502	0.881	0.895	0.759
ResNet-34	0.531	**0.892**	0.910	0.778	0.510	0.884	0.898	0.764
ResNet-50	**0.535**	**0.892**	**0.912**	**0.780**	**0.513**	**0.886**	**0.903**	**0.767**

Table 8
Comparisons with state-of-the-art methods.

method	Off-site				On-site			
	Kappa	F_1	AUC	Final	Kappa	F_1	AUC	Final
ResNeXt-50	0.460	0.866	0.858	0.728	0.463	0.867	0.850	0.727
SE-ResNet-50	0.432	0.864	0.861	0.719	0.410	0.860	0.857	0.709
SE-ResNeXt-50	0.427	0.866	0.879	0.724	0.422	0.867	0.878	0.722
Inception-v4	0.506	0.879	0.869	0.752	0.451	0.867	0.836	0.718
Li et al [27].	0.449	0.873	0.868	0.730	0.440	0.872	0.871	0.727
Jordi et al [23].	0.426	0.850	0.805	0.693	0.415	0.842	0.800	0.686
Gour and Khanna [24]	0.433	0.853	0.849	0.712	0.420	0.849	0.834	0.701
He et al [25].	0.520	0.886	0.903	0.770	0.500	0.877	0.897	0.758
Ours	**0.535**	**0.892**	**0.912**	**0.780**	**0.513**	**0.886**	**0.903**	**0.767**

4.4. Comparisons with state-of-the-art methods

Compared with the current state-of-the-art results of eight diseases classifications on the OIA-ODIR dataset, our method has achieved the best results on the multi-label classification task. As is shown in Table 8, in the off-site set our method improves the Kappa, F_1, AUC and Final score by 0.015, 0.006, 0.009 and 0.010, respectively. In the on-site set, our method improves the Kappa, F_1, AUC and Final score by 0.013, 0.009, 0.006 and 0.040, respectively. Compared with other method, we use local features and global features for contrastive learning by attention mechanism to enhance feature information under the two-stream interactive architecture. Besides, we obtain multi-scale features to enrich the extracted feature maps.

5. Conclusion

In this article, we propose BFENet, a patient level multi-label ophthalmic diseases classification model for binocular fundus images, bilateral feature enhancement network. The designed feature enhancement module can effectively enhance the disease characteristics of each eye in the global information, recover more details and facilitate complemen-tary information exchange. Multi-scale module can extract feature information of different resolutions fundus images and improve classification performance. Both of the proposed modules can be widely applied in various backbone networks and trained in end-to-end manner. The designed ablation experiment verify the effectiveness of the above modules. For backbone of BFENet, the improvement from linear increase of network depth in this task is limited. On complex multi-label ophthalmic disease classification tasks, BFENet achieves the best results and can aid in clinical diagnosis.

The current method has several limitations, which should be addressed in our future research. Firstly, since only patient-level ophthalmic diseases category labels are provided, we cannot compare the results of our proposed method with those image-level classifications. Secondly, the distribution of different categories cases is seriously unbalanced. Oversampling cases with small samples may be an effective way to solve this problem.

The proposed method has good scalability and can be easily extended beyond the classification of ophthalmic diseases. It can be adapted to similar tasks, such as breast cancer and thorax disease diagnosis, which need bilateral information for consideration. Moreover, the method could also be applied to multimodal image analysis filed, where the correlation between different modal images is important for the tasks.

Declaration of Competing Interest

None.

Acknowledgement

This research was supported by the National Natural Science Foundation of China through Grants (No. 61973174).

References

[1] S.R. Flaxman, R.R.A. Bourne, S. Resnikoff, et al., Global causes of blindness and distance vision impairment 1990–2020: a systematic review and meta-analysis, Lancet Glob. Health 5 (12) (2017) e1221–e1234.

[2] J.D. Steinmetz, R.R.A. Bourne, P.S. Briant, et al., Causes of blindness and vision impairment in 2020 and trends over 30 years, and prevalence of avoidable blindness in relation to VISION 2020: the Right to Sight: an analysis for the global burden of disease study, Lancet Glob. Health 9 (2) (2021) e144–e160.

[3] L. Kong, M. Fry, M. Al-Samarraie, et al., An update on progress and the changing epidemiology of causes of childhood blindness worldwide, J. Am. Assoc. Pediatr. Ophthalmol. Strabismus 16 (6) (2012) 501–507.

[4] A.W. Stitt, T.M. Curtis, M. Chen, et al., The progress in understanding and treatment of diabetic retinopathy, Prog. Retin. Eye Res. 51 (2016) 156–186.

[5] R. Bernardes, P. Serranho, C. Lobo, Digital ocular fundus imaging: a review, Ophthalmologica 226 (4) (2011) 161–181.

[6] T. Hu, H. Qi, Q. Huang, & Y. Lu, See better before looking closer: Weakly supervised data augmentation network for fine-grained visual classification. 2019 arXiv preprint arXiv:1901.09891.

[7] C. Zhang, G. Lin, F. Liu, et al., Canet: Class-agnostic segmentation networks with iterative refinement and attentive few-shot learning, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 5217–5226.

[8] Q. Guan, Y. Huang, Y. Luo, et al., Discriminative Feature Learning for Thorax Disease classification in Chest X-ray Images, IEEE Trans. Image Process. 30 (2021) 2476–2487.

[9] G. Litjens, T. Kooi, B.E. Bejnordi, et al., A survey on deep learning in medical image analysis, Med. Image Anal. 42 (2017) 60–88.

[10] W. Zhao, J. Yang, Y. Sun, et al., 3D deep learning from CT scans predicts tumor invasiveness of subcentimeter pulmonary adenocarcinomas, Cancer Res. 78 (24) (2018) 6881–6889.

[11] J. Yang, H. Deng, X. Huang, et al., Relational learning between multiple pulmonary nodules via deep set attention transformers, in: Proceedings of the IEEE 17th International Symposium on Biomedical Imaging (ISBI), IEEE, 2020, pp. 1875–1878.

[12] U. Raghavendra, H. Fujita, S.V. Bhandary, et al., Deep convolution neural network for accurate diagnosis of glaucoma using digital fundus images, Inf. Sci. 441 (2018) 41–49.

[13] H. Fu, J. Cheng, Y. Xu, et al., Joint optic disc and cup segmentation based on multilabel deep network and polar transformation, IEEE Trans. Med. Imaging 37 (7) (2018) 1597–1605.

[14] T.B. Sekou, M. Hidane, J. Olivier, & H. Cardot, From Patch to Image Segmentation using Fully Convolutional Networks–Application to Retinal Images. 2019 arXiv preprint arXiv:1904.03892.

[15] V. Gulshan, L. Peng, M. Coram, et al., Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs, JAMA 316 (22) (2016) 2402–2410.

[16] Z. Li, Y. He, S. Keel, et al., Efficacy of a deep learning system for detecting glaucomatous optic neuropathy based on color fundus photographs, Ophthalmology 125 (8) (2018) 1199–1206.

[17] S.P.K. Karri, D. Chakraborty, J. Chatterjee, Transfer learning based classification of optical coherence tomography images with diabetic macularedema and dry agerelated macular degeneration, Biomed. Opt. Express 8 (2) (2017) 579–592.

[18] F. Li, H. Chen, Z. Liu, et al., Deep learning-based automated detection of retinal diseases of retinal diseases using optical coherence tomography images, Biomed. Opt. Express 10 (12) (2019) 6204–6226.

[19] D.S.W. Ting, C.Y.L. Cheung, G. Lim, et al., Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes, JAMA 318 (22) (2017) 2211–2223.

[20] Z. Li, Y. He, S. Keel, et al., Efficacy of a deep learning system for detecting glaucomatous optic neuropathy based on color fundus photographs, Ophthalmology 125 (8) (2018) 1199–1206.

[21] F.L. Ferris, M.D. Davis, T.E. Clemons, et al., A simplified severity scale for agerelated macular degeneration: AREDS Report No. 18, Arch. Ophthalmol. 123 (11) (2005) 1570–1574 (Chicago, Ill.: 1960).

[22] M.T. Islam, S.A. Imran, A. Arefeen, et al., Source and camera independent ophthalmic disease recognition from fundus image using neural network, in: Proceedings of the IEEE International Conference on Signal Processing, Information, Communication & Systems (SPICSCON), IEEE, 2019, pp. 59–63.

[23] C.C. Jordi, N.D.R. Joan Manuel, V.R. Carles, Ocular Disease Intelligent Recognition Through Deep Learning Architectures, Universitat Oberta de Catalunya, Barcelona, Spain, 2019.

[24] N. Gour, P. Khanna, Multi-class multi-label ophthalmological disease detection using transfer learning based convolutional neural network, Biomed. Signal Process. Control 66 (2021) 102329.

[25] J. He, C. Li, J. Ye, et al., Multi-label ocular disease classification with a dense correlation deep neural network, Biomed. Signal Process. Control 63 (2021) 102167.

[26] C. Li, J. Ye, J. He, et al., Dense correlation network for automated multi-label ocular disease detection with paired color fundus photographs, in: Proceedings of the IEEE 17th International Symposium on Biomedical Imaging (ISBI), IEEE, 2020, pp. 1–4.

[27] N. Li, T. Li, C. Hu, K. Wang, H. Kang, A Benchmark of ocular disease intelligent recognition: one shot for multi-disease detection, in: Proceedings of theThird Bench Council International Symposium, 2020, pp. 177–193.

[28] A. Smitha, P. Jidesh, Classification of multiple retinal disorders from enhanced fundus images using semi-supervised GAN, SN Comput. Sci. 3 (1) (2022) 1–11.

[29] M.A. Bravo, P.A. Arbeláez, Automatic diabetic retinopathy classification, in: Proceedings of the International Conference on Medical Information Processing and Analysis. International Society for Optics and Photonics, 10572, 2017.

[30] H. Zhang, K. Niu, Y. Xiong, et al., Automatic cataract grading methods based on deep learning, Comput. Methods Programs Biomed. 182 (2019) 104978.

[31] J.H. Tan, S.V. Bhandary, S. Sivaprasad, et al., Agerelated macular degeneration detection using deep convolutional neural network, Future Gener. Comput. Syst. 87 (2018) 127–135.

[32] J.E.W. Koh, E.Y.K. Ng, S.V. Bhandary, et al., Automated detection of retinal health using PHOG and SURF features extracted from fundus images, Appl. Intell. 48 (5) (2018) 1379–1393.

[33] S. Chelaramani, M. Gupta, V. Agarwal, et al., Multi-task learning for fine–grained eye disease prediction, in: Proceedings of the Asian conference on pattern Recognition, Springer, Cham, 2019, pp. 734–749.

[34] X. Chen, Y. Xu, L. Duan, et al., Multiple ocular diseases classification with graph regularized probabilistic multi-label learning, in: Proceedings of the Asian Conference on Pattern Recognition, Springer, Cham, 2014, pp. 127–142.

[35] J. Fu, H. Zheng, T. Mei, Look closer to see better: Recurrent attention convolutional neural network for fine-grained image recognition, in: Proceedings of the CVPR, 2017, pp. 4476–4484.

[36] C. Jin, Z. Shi, H. Zhang, Y. Yin, Predicting lncRNA–protein interactions based on graph autoencoders and collaborative training, in: Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2021.

[37] Z. Wang, Y. Yin, J. Shi, W. Fang, W. Li, X. Wang, Zoom-innet: Deep mining lesions for diabetic retinopathy detection, in: Proceedings of the MICCAI, 2017, pp. 267–275.

[38] A. He, T. Li, N. Li, K. Wang, H. Fu, CABNet: category attention block for imbalanced diabetic retinopathy grading, IEEE Trans. Med. Imaging 40 (1) (2021) 143–153.

[39] F. Yu, & V. Koltun, Multi-scale context aggregation by dilated convolutions. 2015 arXiv preprint arXiv:1511.07122.

[40] C. Chen, C. Qing, X. Xu, et al., Cross parallax attention network for stereo image super-resolution, IEEE Trans. Multimed. (2021).

[41] X. Wang, R. Girshick, A. Gupta, et al., Non-local neural networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 7794–7803.

[42] Z.M. Chen, X.S. Wei, P. Wang, et al., Multi- label image recognition with graph convolutional networks, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 5177–5186.

[43] N. Ketkar, Introduction to pytorch, in: Deep Learning with Python, Apress, Berkeley, CA, 2017, pp. 195–208.

[44] W. Liu, A. Rabinovich, & A.C. Berg, Parsenet: Looking wider to see better. 2015 arXiv preprint arXiv:1506.04579.

[45] Srivastava, K. Rupesh, Klaus Greff, Schmidhuber. Jürgen, Training very deep networks, Advances in neural information processing systems 28 (2015).

[46] S. Zagoruyko, & N. Komodakis, Wide residual networks. 2016 arXiv preprint arXiv:1605.07146.