Multimedia Tools and Applications
https://doi.org/10.1007/s11042-021-11208-0

Automated detection of retinopathy of prematurity by deep attention network

Baiying Lei[1] · Xianlu Zeng[2] · Shan Huang[1] · Rugang Zhang[1] · Guozhen Chen[1] · Jinfeng Zhao[2] · Tianfu Wang[1] · Jiantao Wang[2] · Guoming Zhang[2]

Received: 5 September 2020 / Revised: 13 January 2021 / Accepted: 28 June 2021
Published online: 04 September 2021

Abstract
Retinopathy of prematurity (ROP) is a retinal vascular proliferative disease principally observed in infants born prematurely with low birth weight. ROP is the leading cause of childhood blindness. Early screening and timely treatment are crucial in preventing ROP blindness. Previous ROP diagnosis lacks clear understanding of the underlying factors and properties that supports the final decision. For this reason, a deep convolutional neural network (DCNN) is developed for automated ROP detection using wide-angle retinal images. Specifically, we first choose ResNet50 as our base architecture and improve the ResNet by adding a channel and a spatial attention module. Then, we utilize a class-discriminative localization technique (i.e., gradient-weighted class activation mapping (Grad-CAM)) to visualize the trained models and realize pathological structure localization. The efficacy of the proposed network is evaluated on two test datasets. Our method obtains a sensitivity of 94.84 % and a specificity of 99.49 % on test set 1 while a sensitivity of 98.03 % and a specificity of 94.55 % on test set 2. Also, the model successfully detects the pathological structures of ROP (e.g., demarcation lines or ridges) in the retina images.

Keywords Retinopathy of prematurity · Deep learning · Attention mechanism · Automatic detection

1 Introduction

Retinopathy of prematurity (ROP) is a vascular proliferative blindness-causing retinal disease affecting infants with low birth weight, especially premature infants. It is the leading cause of

Baiying Lei and Xianlu Zeng contributed equally to this work.

✉ Jiantao Wang
1021235537@qq.com

✉ Guoming Zhang
13823509060@163.com

Extended author information available on the last page of the article

childhood blindness worldwide [17, 26], causing a significant psychological burden on the child and the family. Globally, in 2010, it is estimated that 184,700 infants of 14.9 million premature infants developed any stage of ROP, about 20,000 of them became blind or severely visually impaired from ROP [26]. The international classification of Retinopathy of Prematurity (ICROP) was published by an international group of ROP experts first in 1984 and later expanded in 1987 to facilitate the development of clinical treatment and further understanding of this disease [20, 21]. Some clarifications and changes were made to the original ICROP in the International Committee of Retinopathy of Prematurity in 2005 [22]. The ICROP was based on several key points in describing the ROP. These include (1) location of the ROP symptom by three zones with each zone is centered on the optic disc of the retina. (2) Five stages of ROP is used to describe the severity of disease. (3) Plus disease is characterized by presenting sufficient vascular dilatation and tortuosity and occurs in active ROP. (4) Pre-plus disease and aggressive posterior ROP (AP-ROP) and so on.

The progression of ROP is very quickly and the effective treatment window is short. If the disease progresses to the late stage traction retinal detachment, the visual acuity prognosis after retinal surgery is usually poor [3]. According to the Early Treatment for Retinopathy of Prematurity (ETROP) trial [13, 45], early treatment has shown to be beneficial to improve the visual acuity of the high-risk infants with ROP. Early detection of a disease that with high probability of progression to retinal detachment and treat it with peripheral retinal laser can reduce the risk of blindness [3]. Therefore, early detection and timely intervention are extremely crucial to prevent ROP blindness.

With the development of China's second-child policy and the increase number of pregnant women in senior age, the percentages of premature and low birth weight babies keep increasing [39]. In addition, the development of the neonatal intensive care unit (NICU) causes an increase in the survival rate of premature and low birth weight infants significantly. Meanwhile, the incidence rate of the ROP is increased, which causes an increase of ROP screening yearly [19, 38]. The current Chinese screening guidelines recommend a minimum examination of all infants with birth weights (BWs) of 2000 g or less or estimated gestational age (GA) at the birth of 32 weeks or less. Although the screening criteria have high sensitivity to detect infants in the need of treatment, the implementation guidelines cause many unnecessary examinations, and only a small percentage of infants screened will meet the treatment criteria [10]. However, ROP screening generally requires professional ophthalmology knowledge, which is time consuming. Also, there is a shortage of professional ophthalmologists in many underdeveloped regions. Since the current telemedicine cannot solve this problem [19], many premature infants become blind due to the lack of timely screening and early treatment. The existing ophthalmologists highly rely on binocular indirect ophthalmoscopes and wide-angle digital retinal imaging systems (RetCam3) in the ROP diagnosis of the premature infants [27], but the classification guidelines provide only qualitative signs rather than quantitative descriptions [18]. Thus, clinical diagnosis mainly depends on the subjective experience of the ophthalmologist on symptoms. As a result, there is disagreement in different experts for the same examination and the ROP diagnosis decisions differ even by one expert made on different days. This uncertainty has been reported when diagnosing the stage of ROP and the existence of plus disease [6, 14]. Hence, the objective assessment is greatly necessary with the computer-aided diagnosis (CAD). Over the past decade, many methods were proposd to assist the clincians with CAD by using diverse data modality [8, 24, 29, 56]. For instance, Diaz et al. proposed a representation of the joint dynamical and static handwritting information to assist clinicians to diagnose Parkinson's disease [8]. To assist ophthalmologists in reducing the

subjectivity and increase the ROP diagnostic accuracy, CAD for automatic ROP screening is developed. The RetCam3 can objectively observe and record the fundus image of premature infants [48]. Some researchers proposed some traditional methods to realize the purpose of assisting the clinicians to make an objective decision by the way of CAD [25, 51]. However, the traditional methods are limited to extract deep high-level features for fundus images and insufficient to improve the classification performance of normal and ROP images. For the reason that the fundus images are challenging and the lesion features are not obvious. As shown in Fig. 1, we can observe that the constrat between the lesion and normal areas is low with inhomogeneous illumination. Besides, the salient objects are relative small to the whole image. Therefore, the method that can extract deep high-level is needed, which provides the opportunity for automatic screening of ROP by machine learning, especially deep learning method.

It is known that deep convolutional neural networks (DCNNs) have the powerful ability to process image datasets. They have signifiantly improved the performance on computer vision tasks [12, 28, 32] based on their rich representation power. As networks move deeper in a data-driven manner, they extract different levels of features from low to high. Also, plenty of studies have explored DCNNs in a range of highly complex medical domains applications, including classification of skin cancer [9], diagnosis of Alzheimer's disease [31], breast cancer detection [53], recognition of gastric infections [35], etc. In ophthalmology, DCNNs have also been used to automatically diagnose cataracts [33], glaucoma [36], aged macular degeneration [2] and diabetic retinopathy [15]. In addition, in the field of ROP, Brown et al. [1] used DCNNs including a preprocessing network similar to U-Net architecture and a classification network similar to Inception V1 architecture, which achieved a sensitivity of 93 % and a specificity of 94 % in a test set of 100 retinal images for the diagnosis of plus disease. Zhang et al. [54] studied the automatic ROP screening system and obtained a level closed to the human experts. Wang et al. [46] proposed an automated ROP detection system and dividing the ROP detection into ROP identification (Id-Net) and grading (Gr-Net). The Id-Net is used to classify the presence of the ROP disease, and the Gr-Net is responsible to identify its severity. However, the previous methods still have the limitation of understanding the underlying factors and properties that support the final decision, such as the pathological structures (e.g., demarcation lines, ridges) in the retina.

To improve the performance of DCNNs, recently, many studies have focused on some factors of the networks. The most important of which are depth, width, and cardinality. VGGNet [42] indicated that stacked blocks with the same shape can give reasonable results. Also, ResNet [16] built a deep architecture by stacking the same residual blocks along with skip connection. GoogLeNet [43] indicated that expand the width of a network can improve the performance of a model. Xception [7] and ResNeXt [49] proposed to increase the

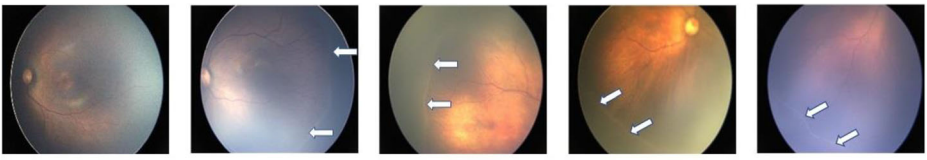

Fig. 1 Fundus images from Retcam3. (**a**) represents normal fundus image and (**b**)-(**e**) indicate ROP fundus images with different lesion locations and appearances. There is a definite structure demarcation line or ridge that separates the avascular retina from the vascularized retina in ROP

cardinality of the network. They indicated that the cardinality is a major factor in improving the performance of a network, it not only reduced the total number of parameters but also produced a more powerful representation. In addition to these factors, attention mechanism has been studied extensively in the literature [11, 40, 47, 55]. Attention mechanism can increase the representation power of DCNNs by focusing on important features and inhibiting unnecessary ones. In this study, the attention mechanism is exploited to help DCNNs to pay more attention to the ROP pathological structures.

In this paper, we present a novel DCNN for automatic ROP detection that aims to improve the interpretability of the model while achieving a satisfactory performance. We take advantage of the residual learning and attention mechanism to train a DCNN with a set of fundus images in which the attention mechanism is used to strengthen the feature representation ability of DCNN, making it focuses more on semantically meaningful parts (i.e., demarcation lines or ridges) in fundus images. Except for classification, another task of our study is to locate the pathological structures. This is similar to weakly supervised localization. We apply a class-discriminative localization technique named gradient- weighted class activation mapping (Grad-CAM) [41] to localize the pathological structure of ROP, which is generally a demarcation line or ridge. We test our proposed method in two self-collected ROP dataset and achieve remarkable performances for the ROP detection, which show potential application in assisting the ophthalmologist for ROP screening.

2 Methodology

The proposed method consists of several parts: the basic architecture for classification is based on residual learning; the channel attention and spatial attention (short as CASA) are applied to strengthen the feature representation ability. Note that the CASA module is integrated into every residual block, and the Grad-CAM is utilized to localize the pathological structures. The overall architecture is shown in Fig. 2. Specifically, we adopt the ResNet50 model as the backbone to extract the high-level features, which takes the original images as input of the whole network. For each residual module, the CASA is integrated together to enforce the network to extract the features related to the nidus. After obtaining the discriminative features, the Grad-CAM method is used to localize the pathological structure to provide the interpretability of the learned features of ROP with visualization.

2.1 Basic architecture

The feature representation ability of DCNN can be increased by stacking blocks. An ordinary block is defined as:

$$y = F(x, W) \tag{1}$$

where x and y denote the input and output of the block respectively, and $F(\cdot)$ denotes the underlying mapping function, which is learned by the parameter set W of these stacked layers. However, as the depth increases, the networks are more difficult to train due to vanishing gradients. To ease the training of networks, a residual learning framework (ResNet) [16] is presented. The residual learning is defined as:

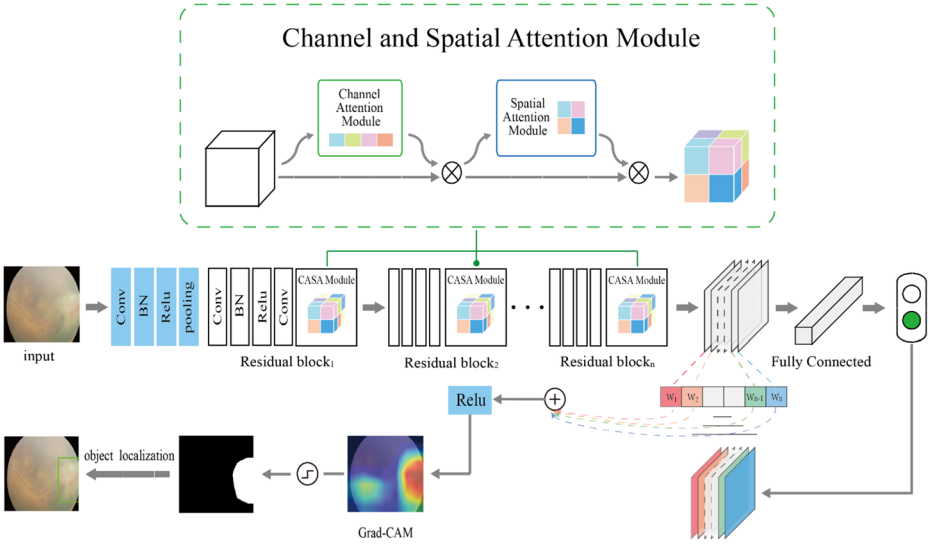

Fig. 2 Architecture of the proposed network for ROP detection. The ResNet50 is adopted as the backbone, which is integrated with channel and spatial attention module, to extract the discriminative features related to the lesion of ROP. Then the obtained features are processed with Grad-CAM method to visualize the learned features by the proposed method and locate the pathological structures

$$y = F(x, W) + x \tag{2}$$

The framework of residual learning can be implemented through the shortcut connection that performs an element-wise summation to the input and output. The ResNet was shown to be an effective backbone architecture in addressing many image analysis problems [4, 5, 23, 30]. In this study, we choose ResNet50 as our basic architecture and the structure of the used ResNet50 is shown in Fig. 4.

2.2 Channel and spatial attention module

The CASA module is similar to [47]. The implementation of the attention process can be divided into two steps: Firstly, It provides an intermediate feature map $\mathbf{F} \in \mathbb{R}^{\mathbf{C}\times\mathbf{H}\times\mathbf{W}}$ as input, $\mathbf{F}$ is 3D tensors with dimensionality of H height, W width and C channel. And the channel attention sub-module generates a 1D channel attention map $\mathbf{M_C} \in \mathbb{R}^{C\times1\times1}$, thereby obtaining a feature map $\mathbf{F}' \in \mathbb{R}^{C\times H\times W}$ refined by the channel attention map. Secondly, the spatial attention sub-module generates a 2D spatial attention map$\mathbf{M_S} \in \mathbb{R}^{1\times H\times W}$, and then obtains the final refined output $\mathbf{F}'' \in \mathbb{R}^{C\times H\times W}$. The overall attention process can be summarized as:

$$0pt\mathbf{F}' = \mathbf{M_C}(\mathbf{F}) \otimes \mathbf{F}\mathbf{F}'' = \mathbf{M_S}\left(\mathbf{F}'\right) \otimes \mathbf{F}' \tag{3}$$

where $\otimes$ denotes the element-wise multiplication. In order to ensure that the two matrix dimensions of the multiplication are the same, during the calculation of the multiplication process, the attention map values would be broadcasted correspondingly: channel attention map values are broadcasted along the spatial dimension, and spatial attention map values are broadcasted along the channel dimension. The computation

processes of each attention map are expounded in Fig. 3. The detailed function of each attention module is described as below (Fig. 4).

Channel attention module Since each channel of the feature map is considered as a feature detector [52], the channel attention concentrates on 'what' makes sense for a provided input image. We take advantage of the relationship between the channels of feature to generate a channel attention map. In order to save computational resources for calculating channel attention, we squeezed the spatial dimension of the input feature map. In detail, we first apply both global average pooling and global max pooling operations to aggregate spatial information of a feature map separately. In this way, it generates two different spatial context expressions: $\mathbf{F}^c_{avg}$ and $\mathbf{F}^c_{max}$, which represent global average pooled feature and global max pooled feature, respectively. Then, both features are forwarded to two shared fully connected layers in parallel. Finally, we incorporate the output feature vectors adopting element-wise summation to generate our channel attention map $\mathbf{M}_C \in \mathbb{R}^{C\times 1\times 1}$. In short, the channel attention map is obtained by the following calculation:

$$\begin{aligned}\mathbf{M_c}(\mathbf{F}) &= \sigma(FC_s(GAvgPool(\mathbf{F})) + FC_s(GMaxPool(\mathbf{F}))) \\ &= \sigma(\mathbf{W_1}(\mathbf{W_0}(\mathbf{F}^c_{avg})) + \mathbf{W_1}(\mathbf{W_0}(\mathbf{F}^c_{max})))\end{aligned} \tag{4}$$

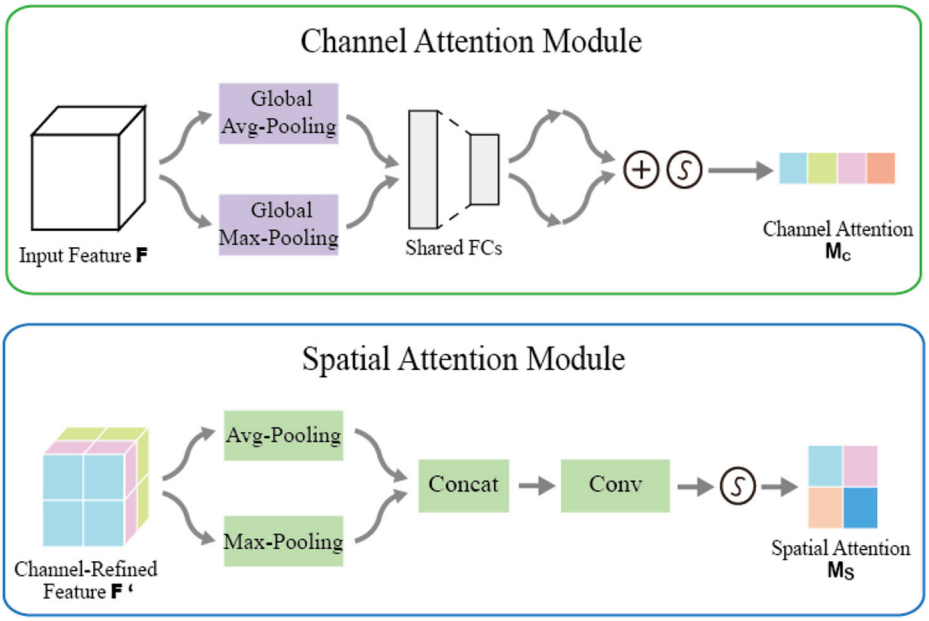

Fig. 3 Illustration of the used channel and spatial attention module. The input features are processed with global average-pooling and global max-pooling operation, which are imported into two shared fully connected layer respectively and then are concatenated to gain the channel attention maps with the sigmoid function. For the spatial attention module, the obtained channel-refined features are used as the input features, which is similar to the channel attention module, where two pooling methods are utilized to handle the features. Then the obtained feature are concatenated and dealt with a convolutional layer, and finally the spatial attention maps are gained after a sigmoid function

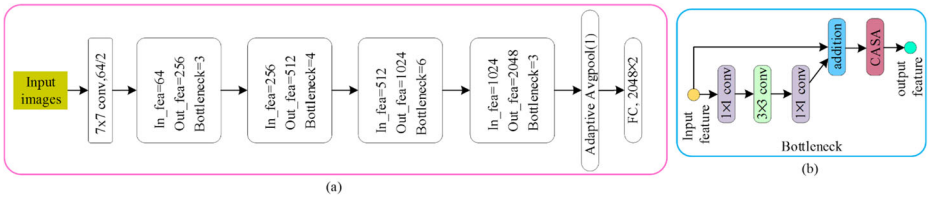

Fig. 4 Structure of the used ResNet-50 model and the corresponding bottleneck. (**a**) represents the ResNet-50 model, where In_fea and Out_fea denote the channel number of the input features and output features respectively, and FC indicates the fully connected layer. (**b**) indicates the used bottleneck in ResNet-50 model, where the CASA module is embed into each bottleneck

where σ denotes the sigmoid function, FC_s denotes two shared fully connected layers, *G AvgPool* and *GMaxPool* denote global average pooling and global max pooling, respectively. To reduce the number of parameters, the first fully connected layer activation size is set to $\mathbb{R}^{C/r\times1\times1}$, where r is the reduction ratio. $\mathbf{W}_0 \in \mathbb{R}^{C/r\times C}$ and $\mathbf{W}_1 \in \mathbb{R}^{C\times C/r}$ are the two shared fully connected layers weights, respectively.

Spatial attention module Different from the channel attention, the spatial attention focuses on 'where' is the large amount of information, which is a supplement to the channel attention. We produce a spatial attention map by exploiting the relationship between the spatial of feature. In detail, we apply both average-pooling and max-pooling operations to generate two 2D maps: $\mathbf{F}^s_{avg} \in \mathbb{R}^{1\times H\times W}$ and $\mathbf{F}^s_{max} \in \mathbb{R}^{1\times H\times W}$, which denote average-pooled feature and max-pooled feature across the channel [50], respectively. Then, we concatenate them and apply a convolution operation to generate our 2D spatial attention map $\mathbf{M}_S \in \mathbb{R}^{1\times H\times W}$, which encodes both relevant and irrelevant information. In short, the spatial attention map is obtained by the following calculation:

$$\begin{aligned}\mathbf{M_c}(\mathbf{F}) &= \sigma(f^{7\times7}(AvgPool(\mathbf{F})); f^{7\times7}(MaxPool(\mathbf{F})))\\ &= \sigma(\mathbf{f}^{7\times7}([\mathbf{F}^s_{avg}; \mathbf{F}^s_{max}]))\end{aligned} \tag{5}$$

where σ denotes the sigmoid function and $f^{7\times7}$ denotes a convolution operation with a filter kernel size of 7 × 7. *AvgPool* and *MaxPool* denote average-pooling and max-pooling along the channel axis, respectively.

2.3 Pathological structures localization via Grad-CAM

The previous studies have shown that Grad-CAM can visualize the regions of input that are 'important' for predictions from these models – or visual explanations [41]. Based on it, we localize the pathological structure (e.g., demarcation line or ridge) by adopting Grad-CAM.

Given an input image, we first obtain class prediction from our network, which will serve as a diagnostic result. Next, we generate Grad-CAM localization maps for the predicted class, which is binarized with the appropriate threshold. This leads to the connected segments of pixels and we draw our bounding rectangle around the largest contour. In general, for an image predicted to be ROP, the rectangular framed area is the pathological structures. Accordingly, we can explain ROP by providing the pathological structures.

3 Experimental setup and results

3.1 Dataset

The fundus images (obtained by RetCam2 or RetCam3) are collected from the premature infants screened in the Shenzhen Screening for Retinopathy of Prematurity Cooperative Group. Diagnosis of ROP demands to inspect the fundus of premature infants from different views when using RetCam3. A standard 10-views photograph for an infant's eye is performed in an examination. Figure 5 presents an examination of the left eye in stage 3 of ROP. In this examination, the ridge in the fundus images can be observed in images other than (a), (b) and (g). In order to facilitate the detection of pathological structures, we train our model based on images, namely, we determine if an image has the characteristics of ROP. However, the actual clinical ROP screening is performed in terms of each infant rather than each image independently. Therefore, we use two test datasets to evaluate the effectiveness of our method, one is based on the images and the other is based on cases. Each infant is split into two cases according to the left or right eyes since the healthy states of each eye could be different.

The data is labeled by five pediatric ophthalmologists in this work, two are senior experts (chief physicians) who have about 20 years of clinical experience in ROP screening and treatment, two are attending physicians that have about 10 years of clinical experience, and one is junior ophthalmologist that have about 3 years of clinical experience. All the pediatric ophthalmologists individually labeled the images or cases as positive or negative (ROP or Normal). If there are distinctions, a consensus which reached through group discussion is used as the final label.

The datasets used in this study are shown in Table 1. All infants' birth weight $\leq$ 2000 g or gestational age $\leq$ 36.5 weeks. Note that the ratio of the two classes (1:10) in the test set 1 approximates the natural distribution of ROP screening results in our study. Train set and test set 1 are the datasets used in [54] (its work is image-based, without counting the number of infants and cases). The last row of Table 1 shows the number of cases and images on the test set 2. Note that in the original data collected, the number of images per case may exceed 10 which result in multiple images at a single view due to factors such as the infant eye movement in the examination. We manually remove redundant or blurred images. Finally, the number of images in each case is less than or equal to 10 in test set 2.

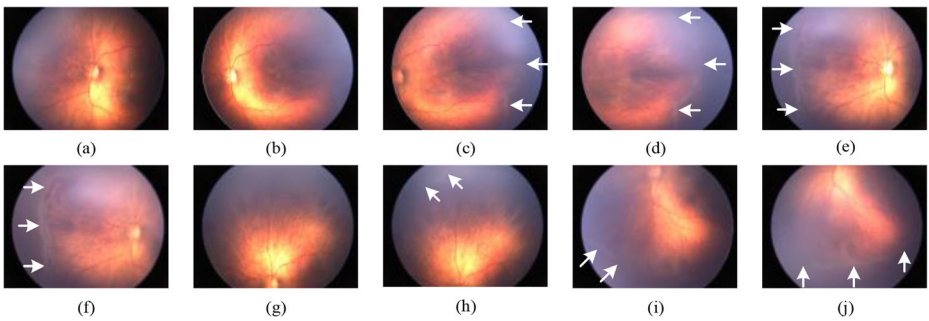

Fig. 5 Standard 10-views fundus images in an examination of the left eye. (**a**) disc-centered image; (**b**) macular-centered image; (**c**) temporal side image with the optic disc; (**d**) temporal side image without the optic disc; (**e**) nasal side image with the optic disc; (**f**) nasal side image with optic disc but close to serrated edge; (**g**) top image with the optic disc; (**h**) top image without the optic disc; (**i**) bottom image with the optic disc; (**j**) bottom image without the optic disc. An obvious ridge can be seen in the images at the marked arrows

Table 1 Summaries of datasets used in this study

	Normal			ROP		
	infants	cases	images	infants	cases	images
Train set	--	--	9711	--	--	8090
Test set1	--	--	155	--	--	1587
Test set2	103	202	1927	61	152	1491

3.2 Evaluation metrics

To evaluate the performance of our method, we use the accuracy (ACC), sensitivity (SEN, recall), specificity (SPEC), precision (positive predictive value (PPV)), and F1 score (F1, the harmonic mean of precision and recall) as evaluation metrics, which are defined as:

$$\mathrm{ACC} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FN} + \mathrm{TN} + \mathrm{FP}} \tag{6}$$

$$\mathrm{SEN} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{7}$$

$$\mathrm{SEPC} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}} \tag{8}$$

$$\mathrm{PPV} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{9}$$

$$\mathrm{F1} = \frac{2}{1/\mathrm{precision} + 1/\mathrm{recall}} = \frac{2\mathrm{TP}}{2\mathrm{TP} + \mathrm{FP} + \mathrm{FN}} \tag{10}$$

where TP, TN, FP, and FN denote the number of true positive, true negative, false positive and false negative, respectively. In addition, area under the receiver operating characteristic curve (AUC) is also used to evaluate the performance of ROP detection.

3.3 Implementation details

To save computational resources, the original images are resized from 1600 × 1200 × 3 to 320 × 240 × 3 and we adopted the data augmentation (i.e., horizontal flip, color jitter). Our method is implemented in the platform of Pytorch [37] using two NVIDIA TITAN X GPU with 12 GB RAM. Cross-entropy is used as cost function. The adaptive moment estimation (Adam) is utilized for optimization and weights update. The learning rate is initially set to 0.0001, then reduced by 0.9 decay when the train loss converges. The weight decay is set to 0.0001 and a mini batch size of 64 is used.

3.4 The composition of the attention module

For the experiment exploring the composition of the attention module, we use the test set 1 to measure the performance of different methods and adopt ResNet50 as base architecture. In this part of experiment, we demonstrate our module design process, which is split into three parts. We first evaluate the effectiveness of the channel attention module, then the spatial attention module. Finally, we verify the combinations of the channel attention module and spatial attention module.

Channel attention module We experimentally confirm that joint using both global average pooling and global max pooling results in finer attention performance. We compare three ways of aggregating spatial information in channel attention: global average pooling, global max pooling, and joint use of both global pooling. Experimental results with different pooling operations are shown in Table 2. We can see that both global average pooling and global max pooling are meaningful, also, joint using both global average pooling and global max pooling achieves the best performance. As a brief conclusion, we use both global average pooling and global max pooling in our channel attention module in the following experiment.

Spatial attention module Spatial attention requires a 2D descriptor that encodes channel information on every pixel of all spatial locations. We experimentally verify that the ways of encoding channel information are utilized to gain the 2D descriptor. Similar to channel attention, we compare three pooling ways: average-pooling, max-pooling, and joint use of both pooling (the pooling both across the channel axis). Then, a convolutional layer is used to reduce the channel dimension to 1. We also explore the effect of convolutional kernel size with the following four settings: kernel sizes of 1, 3, 5 and 7.

Table 3 shows the experimental results. We can observe that joint using both average-pooling and max-pooling with kernel size of 7 achieves the best accuracy. As a brief conclusion, we use both average-pooling and max-pooling across the channel axis and a convolution layer with a filter kernel size of 7 × 7 as our spatial attention module.

Combinations of the channel and spatial attention module We compare five different combinations of the channel attention module and spatial attention module: single channel attention (ResNet50_CA), single spatial attention (ResNet50_SA), sequential spatial-channel attention (ResNet50_SA + CA), sequential channel-spatial attention (ResNet50_CA + SA), and parallel use of both attention modules (ResNet50_CA & SA in parallel). In the case of ResNet50_CA & SA in parallel, two attention modules are applied in parallel and the outputs are added. As the channel attention module and spatial attention module have different capabilities, the ways of the combination may affect the overall performance.

Table 2 The performance of different channel attention methods on the test set 1 (%)

Description (channel)	Acc	Sen	Spec	Pre	F1	AUC
ResNet50	98.15	90.32	98.92	89.17	89.74	99.27
ResNet50 + GAvgPool	98.21	92.90	98.73	87.88	90.28	**99.64**
ResNet50 + GMaxPool	98.32	92.90	98.86	88.89	90.85	99.41
ResNet50 + GAvgPool & GMaxPool	**98.67**	**94.19**	**99.11**	**91.25**	**92.70**	99.53

Bold entries indicate the effectiveness of GAvgPool and GmaxPool

Table 3 The performance of different channel attention methods on the test set 1 (%)

Variable	Description (spatial)	Acc	Sen	Spec	Pre	F1	AUC
--	ResNet50	98.15	90.32	98.92	89.17	89.74	99.27
	ResNet50 + AvgPool	98.27	92.90	98.80	88.34	90.57	99.33
kernel size = 1	ResNet50 + MaxPool	98.27	87.74	99.30	92.52	90.07	99.05
	ResNet50 + AvgPool & MaxPool	98.33	92.26	98.92	89.38	90.79	99.35
	ResNet50 + AvgPool	98.27	90.97	98,99	89.81	90.38	99.52
kernel size = 3	ResNet50 + MaxPool	98.33	90.32	99.11	90.91	90.61	99.37
	ResNet50 + AvgPool & MaxPool	98.38	90.97	99.11	90.97	90.97	99.55
	ResNet50 + AvgPool	98.27	92.90	98.80	88.34	90.57	99.55
kernel size = 5	ResNet50 + MaxPool	97.86	90.97	98.54	85.98	88.64	98.11
	ResNet50 + AvgPool & MaxPool	98.33	87.74	**99.37**	**93.15**	90.37	99.16
	ResNet50 + AvgPool	98.21	90.97	98.99	89.24	90.10	99.02
kernel size = 7	ResNet50 + MaxPool	98.33	90.32	99.11	90.91	90.61	99.46
	ResNet50 + AvgPool & MaxPool	**98.50**	**93.55**	98.99	90.06	**91.77**	**99.56**

Bold entries show the effectiveness of the combination of AvgPool and MaxPool in the channel attention module

The experimental results on different attention combining methods on the test set 1 are shown in Table 4. From the results, we can find that the ResNet50_CA + SA obtains the best performance, so we choose it as our final method. Note that all the combining methods using two attentions outperform that only with the channel attention or spatial attention independently. It shows that utilizing both channel and spatial attention is crucial to enhance performance. It is also important to note that the performance of the network with CA is better than the one with the SA module, which indicates that channel information plays a more important role in classification compared with the spatial information.

3.5 The location of the attention module

We believe that the different locations of attention module will affect the final performance. In this section, we discuss the influence of attention module placement to find the optimal location. Three different locations are compared, respectively: P1: Embedding attention modules in each residual connection (taking ResNet50 as an example, 16 attention modules are added in total); P2: Attention modules are immediately followed by each residual block (4 attention modules are added in total); P3: After all residual blocks, add one attention module (1 attention module in total) before global average pooling. The P1 method used in the experiment of attention module composition is discussed earlier in this section. The experimental

Table 4 The performance of different combination methods of the channel and spatial attention module in test set 1 (%)

Method	ACC	SEN	SPEC	PPV	F1	AUC
ResNet50	98.15	90.32	98.92	89.17	89.74	99.27
ResNet50_CA	98.67	94.19	99.11	91.25	92.70	99.53
ResNet50_SA	98.50	93.55	98.99	90.06	91.77	99.56
ResNet50_SA + CA	98.79	93.55	99.30	92.95	93.25	**99.60**
ResNet50_CA + SA (Ours)	**99.08**	**94.84**	**99.49**	**94.84**	**94.53**	99.36
ResNet50_CA & SA in parallel	98.90	94.84	99.30	93.04	93.93	99.50

Bold entries indicate that our proposed ResNet50_CA + SA achieves the best performance

Table 5 The performance of the attention module at different added locations in test set 1 (%)

Location	Acc	Sen	Spec	Pre	F1	AUC
P1	**99.08**	**94.84**	**99.49**	**94.84**	**94.53**	99.36
P2	98.50	91.61	99.18	91.61	91.61	99.07
P3	98.21	90.97	98.92	89.24	90.10	99.32

Bold entries indicate that in the P1 mode, the attention module embedded in each residual connection can obtain the optimal performance

results of the addition locations of different attention modules in test set 1 are shown in Table 5. The attention modules here are composed of the sequential connection channel attention-spatial attention module combination method obtained in the previous section. According to the results, in the P1 mode, the attention module embedded in each residual connection can obtain the optimal performance. According to Table 4, although the performance of the other two positions is slightly better than ResNet50, it is not as good as the single channel attention module or the single spatial attention module. Therefore, the final model composition of the method in this paper is ResNet50_CA + SA, and the attention module is embedded in each residual connection.

3.6 Performance evaluation in image-level dataset

We compare our network with several other typical classification networks on test set 1. We compare the following networks: VGG16 [42], Inceptionv4 [44], Xception [7] and ResNeXt [49]. The experimental results are illustrated in Table 6. It can be observed that our network performs better than other approaches.

Our target is to achieve a satisfactory performance while enhancing the interpretability of the ROP screening system. Namely, providing the underlying factors and properties that support the final decision. Figure 6 illustrates the visualization results with Grad-CAM by ResNet50 and ResNet50-CA + SA. It indicates that the attention regions (the highlights in Grad-CAM) learned by both models. Our model demonstrates a stronger attention capability that covers the regions of the pathological structure comparing to ResNet50. In addition, the pathological structure localization results by ResNet50-CA + SA are visualized in bottom row of Fig. 7. We can see that it can detect the demarcation line or ridge that separates the avascular retina from the vascularized retina approximately. Note that given the 3^{th} column in the sample, the area it positioned does not have a demarcation line or ridge, but some dilation and tortuosity of retinal vessels, that is, the model may have potential application in diagnosing the plus disease.

Table 6 Summaries of the model performance in test set 1 (%)

Method	ACC	SEN	SPEC	PPV	F1	AUC
ResNet50	98.15	90.32	98.92	89.17	89.74	99.27
VGG16	97.23	76.77	99.24	90.84	83.22	98.45
Inceptionv4	98.56	91.61	99.24	92.21	91.91	99.37
Xception	98.04	90.32	98.80	88.05	89.17	99.67
ResNeXt101	98.44	**95.48**	98.73	88.10	91.64	**99.86**
Ours	**99.08**	94.84	**99.49**	**94.84**	**94.53**	99.36

Bold entries indicate that our proposed model has relatively good performance on data set 1 and data set 2

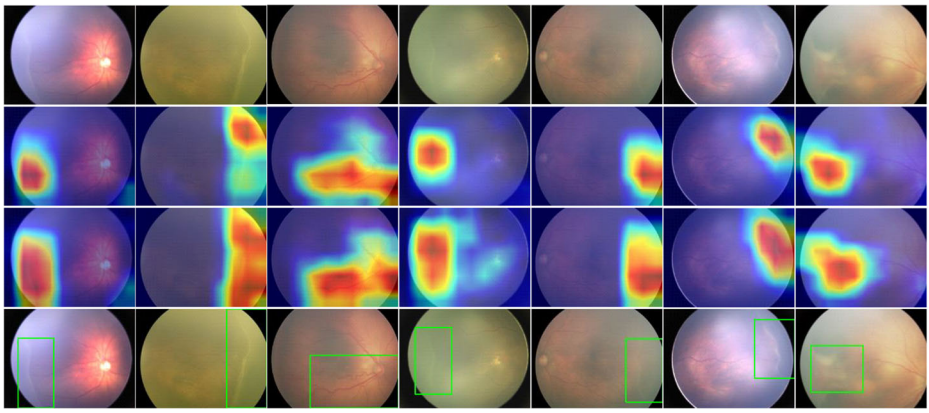

Fig. 6 Visualization results of Grad-CAM and pathological structures localization. Top row: images with ROP. Second row: Grad-CAM by ResNet50. Third row: Grad-CAM by Resnet50_CA + SA. Bottom row: the pathological structures localization results by the green rectangle. The Grad-CAM visualization is calculated for the last convolutional outputs

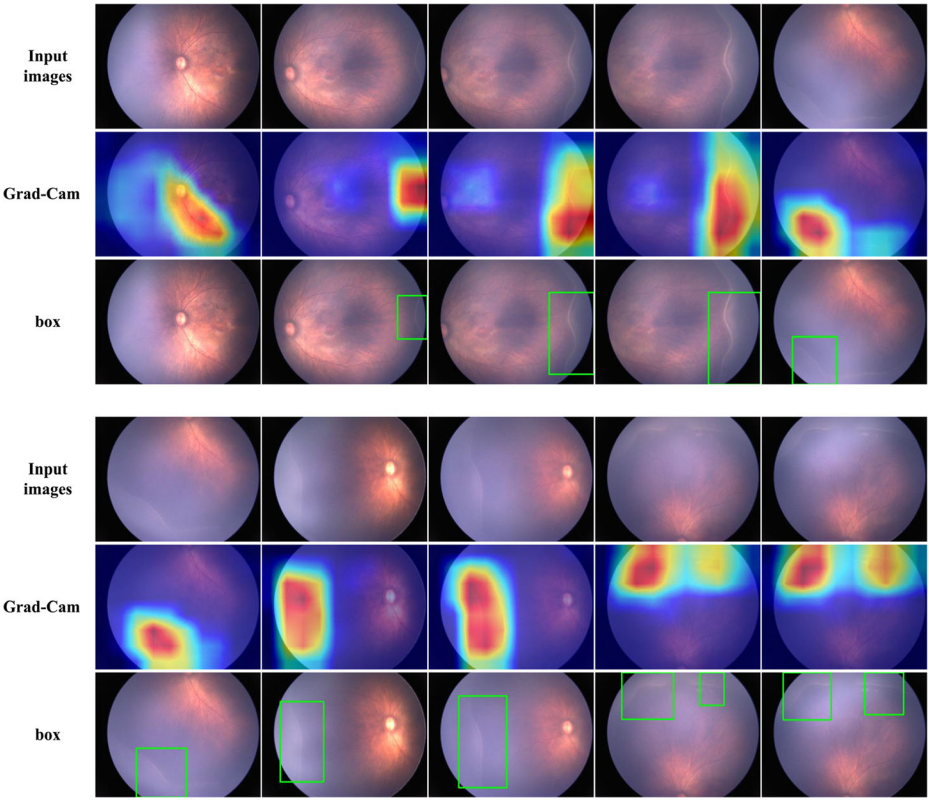

Fig. 7 Visualization results of a case. Top row: the standard 10-views fundus images. Middle row: Grad-CAM by our method. Bottom row: the pathological structures localization results

3.7 Performance evaluation in case-level dataset

Since the actual clinical ROP screening is performed upon each case rather than each image independently. We apply our method to test the model in the test set 2. One case is predicted to be ROP as long as one of the images in the case is labeled as to be ROP. We compare our model with several other typical classification networks to show the effectiveness of our method again. The results are presented in Table 7. It shows that our method still achieves good performance in test set 2, which proves that our method has good robustness. Figure 6 illustrates a case visualization results, among them, the pathological structures of the image predicted to be ROP can be found. We also notice that some predictions are normal images, the network pays more attention to the optic disc area. The main reason may be that the optic disc area is the most obvious features for normal image.

4 Discussion

In this paper, we propose a DCNN framework for automated ROP detection using wide-angle retinal images. Firstly, we adopt the residual learning as the basic architecture for classification. Secondly, the channel and spatial attention module are integrated to enhance the DCNN feature representation capabilities. A better performance is achieved by focusing on important features and suppressing unimportant ones. Finally, the Grad-CAM is used to visualize the trained models and locate the pathological structures. The experiment results show that DCNN can be trained through large-scale datasets to learn the features for ROP diagnosis. Our method can automatically detect ROP in retinal fundus images with high sensitivity and accuracy and without the expert-specified features. Our method obtains promising performance and can provide the interpretable and underlying factors and properties (detecting the demarcation lines or ridges) that support the final decision.

The t-distributed stochastic neighbor embedding (t-SNE) [34] is used to visualize high-level features learned by our method in 2 dimension. Figure 8 illustrates the visualization results. Each point on the scatter plot corresponds to an individual retinal fundus image (left) or a separate case (right), in which similar ones based on their features are closer to each other than dissimilar ones. The ground truth labels are only used for visualization to represent the different clusters (represented in different colors on the scatter plot). The t-SNE shows that different categories (Normal or ROP) can be qualitatively separated, which indicates that the features learned by our method are effective.

Table 7 Summaries of the model performance in test set 2 (%)

Method	ACC	SEN	SPEC	PPV	F1
ResNet50	93.50	97.37	90.59	88.62	92.79
VGG16	90.68	82.24	97.03	95.42	88.34
Inceptionv4	95.76	98.03	94.06	92.55	95.21
Xception	94.07	96.71	92.08	90.18	93.33
ResNeXt101	95.20	92.76	**97.03**	**95.92**	94.31
Ours	**96.05**	**98.03**	94.55	93.13	**95.51**

Bold entries indicate that our proposed model has relatively good performance on data set 1 and data set 2

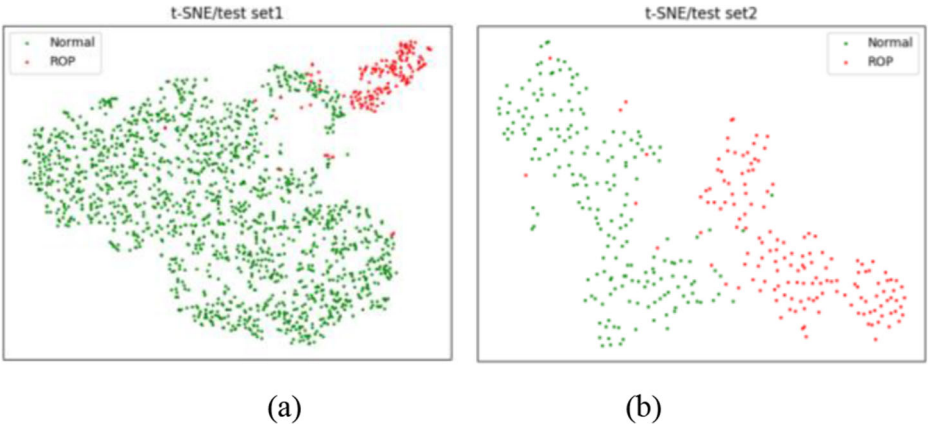

Fig. 8 t-SNE visualization of features extracted from an intermediate layer (the global average pooling) of our method. (**a**) test set 1 and (**b**) test set 2

Although our method has achieved a good performance, there are still some wrong predictions. Figure 9 illustrates the confusion matrix of our model on two test sets. We can conclude from Fig. 8, a total of 16 images are classified erroneously in test set 1, including 8 FP and 8 FN images; and a total of 14 cases are misclassified for test set 2, including 11 FP and 3 FN samples. Samples of these misclassifications are presented to the pediatric ophthalmologist to identify possible reasons for further DCNN improvements.

For the test set 1, Fig. 10 shows several typical FN and FP samples. Among the 8 FN samples, 2 samples are misclassified due to the vagueness in the marginal region of the image. The other 6 samples are all associated with the plus disease (such as Fig. 10(a)), which is defined by the dilation and tortuosity of retinal vessels. We observe that the softmax score of plus disease for the ground-truth class is between 0.3 and 0.7, it indicates that the effect of network on the classification of plus lesions is not ideal. Such FN samples can be reduced by providing more training data of typical non-plus and plus disease. Among the 8 FP samples, 4 are due to blurred images, and the other 4 are due to optical artifacts (e.g., Fig. 10(b) and (c), the rectangular frame with artifact area). Both of these factors are caused by imaging, and a feasible solution might be data cleaning or expanding the data sets.

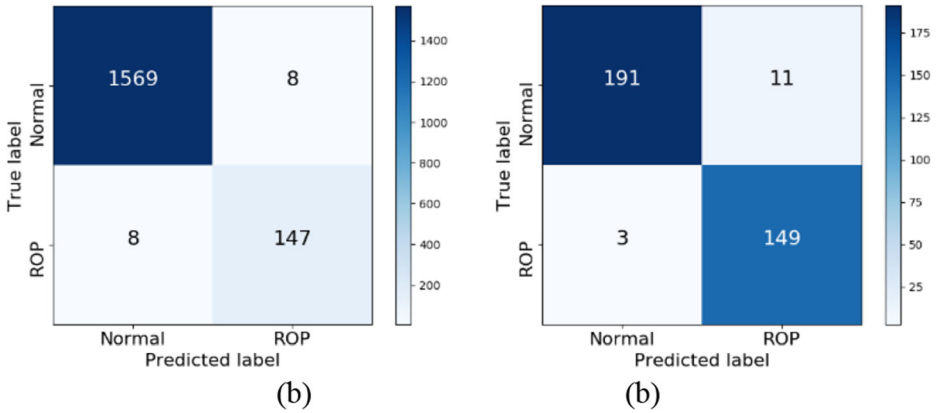

Fig. 9 The confusion matrix of our methods on two test sets. (**a**) test set 1 and (**b**) test set 2

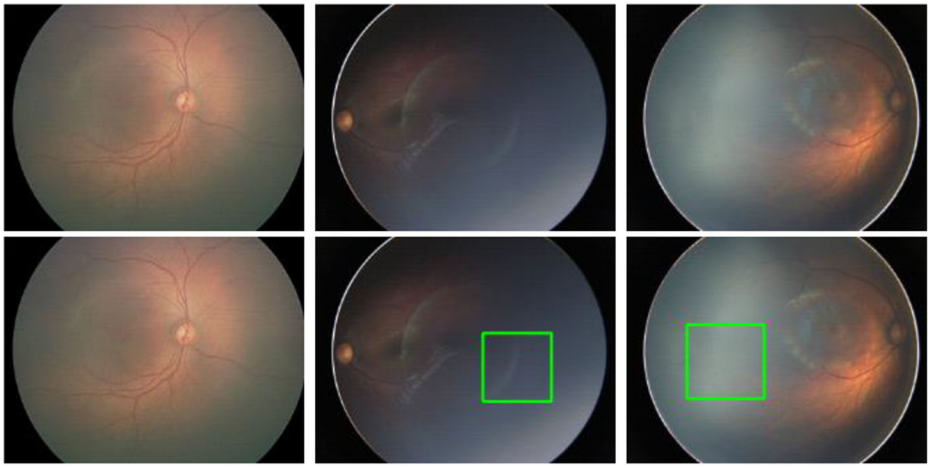

Fig. 10 Misclassified image samples. Top row: the fundus images. Bottom row: the pathological structures localization results that are predicted to be ROP. (**a**) A FN sample, which shows the insignificant characteristics of the plus disease. (**b**) and (**c**) are FP samples with the artifact, which may be misrecognized as a "ridge" of ROP by the DCNN

For the test set 2, due to the high sensitivity of the method upon the images, when it performs upon the cases, there will be fewer FN samples since in a case there will usually be multiple images with ROP features, and individual misjudgments will not affect the final prediction. Conversely, FP samples will be more. All the 3 FN samples are misclassified due to

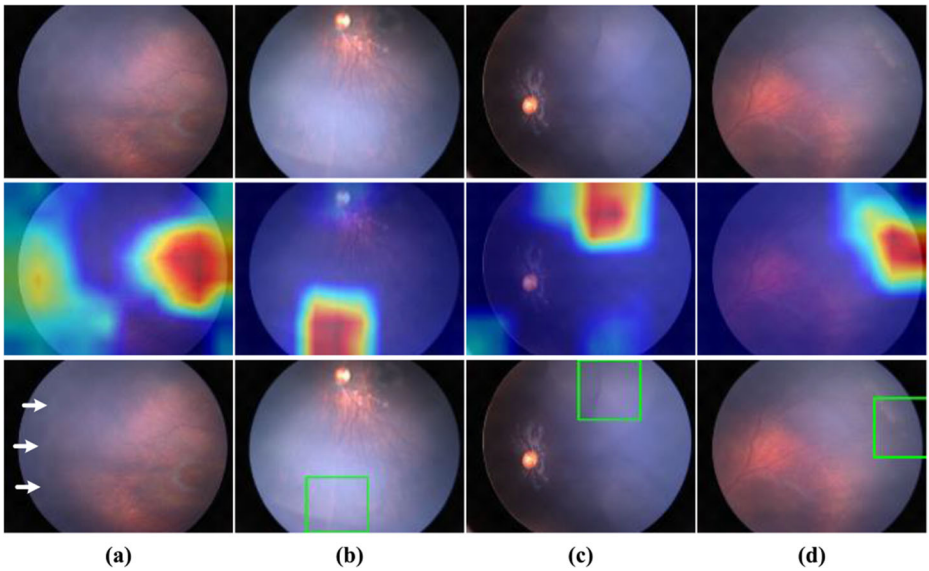

Fig. 11 Typical images of cases of misclassification. Top row: the fundus images. Middle row: Grad-CAM by our method. Bottom row: the pathological structures localization results that are predicted as ROP. (**a**) A FN sample, A 'demarcation line' can be seen in the images at the marked arrows. (**b**), (**c**) and (**d**) are FP samples, where (**d**) is exudative vitreoretinopathy

Table 8 Computation complexity comparison of ResNet50 and ours

Methods	Parameters(M)	FLOPs(G)
ResNet50	23.51	4.12
Ours	26.03	4.14

vagueness in the marginal region and the "demarcation line" is particularly fuzzy. Figure 11(a) shows an example where the network focused on the macular rather than the area with 'demarcation line'. Among the 11 FP samples, most of them are blurred images or optical artifacts (Fig. 11(b) and (c)). It is worth noting that one of the cases is found to be exudative vitreoretinopathy (Fig. 11(d)). The sample is labeled as normal, but the network regards it as ROP. Each of our case criteria should contain images of 10 views, but there are more than 10 in practice due to blurred images or incorrect angles. We manually remove redundant or blurred images, but there are still slipped image through the net.

A comparison of the parameters and FLOPs (Floating-point Operations) of our method and the base architecture (e.g., ResNet50) are shown in Table 8. Since computational resources are limited in reality, we need to reduce the computation complexity of the method while improving performance. We can find that the CASA module only adds a few parameters and FLOPs. The reason is that the main source of CASA module parameters is the shared two-layer fully connected layer, and we reduce the first fully connected layer activation size in order to reduce the number of parameters. The parameters and FLOPs of the CASA module are comparable to the entire network.

5 Conclusions

In this work, we propose a DCNN framework for automatic screening ROP in the wide-angle retinal image. First, the channel and spatial attention module are designed to achieve better performance by focusing on important features and suppressing unnecessary ones. Then, the Grad-CAM is used to visualize trained models while locating the pathological structure. Experiments show our network obtains satisfactory performance and can provide the underlying factors and properties (demarcation lines or ridges) that explain the final decision. There are still some limitations in this study. Firstly, we should study the method of image quality intelligent control to reduce the interference of blurred images. Secondly, we only classify ROP and Normal, nevertheless, the ROP diagnoses are more complicated in clinical application. The threshold disease, pre-threshold disease, aggressive posterior ROP (AP-ROP), as well as zones, stages and plus diseases of ROP [22] are needed for further study. In future work, we will investigate these scenarios and extend the method to diagnose the AP-ROP, plus disease and identify the stage independently. Thirdly, some localization results can not completely cover the pathological structure areas, more accurate pathological structures localization methods should be studied in the future.

Acknowledgements This work was supported partly by Shenzhen Key Medical Discipline Construction Fund (No. SZXK038), Shenzhen Fund for Guangdong Provincial High-level Clinical Key Specialties (No.SZGSP014), Shenzhen-Hong Kong Co-financing Project (No.SGDX20190920110403741), and Guangdong Basic and Applied Basic Research Foundation (No. 2019A1515111205).

Declarations

Competing interests We wish to confirm that there have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Brown JM, Campbell JP, Beers A, Chang K, Ostmo S, Chan RP et al (2018) Automated diagnosis of plus disease in retinopathy of prematurity using deep convolutional neural networks. JAMA Ophthalmol 136(7): 803–810
2. Burlina PM, Joshi N, Pekala M, Pacheco KD, Freund DE, Bressler NM (2017) Automated grading of age-related macular degeneration from color fundus images using deep convolutional neural networks. JAMA Ophthalmol 135(11):1170–1176
3. Chen Y, Feng J, Gilbert C, Yin H, Liang J, Li X (2015) Time at treatment of severe retinopathy of prematurity in China: recommendations for guidelines in more mature infants. PLoS One 10(2):e0116669
4. Chen L-C, Papandreou G, Schroff F, Adam H (2017) Rethinking atrous convolution for semantic image segmentation. arXiv preprint arXiv:1706.05587
5. Chen L-C, Zhu Y, Papandreou G, Schroff F, Adam H (2018) Encoder-decoder with atrous separable convolution for semantic image segmentation. Proc IEEE Europ Conf Comput Vis, pp 801–818
6. Chiang MF, Jiang L, Gelman R, Du YE, Flynn JT (2007) Interexpert agreement of plus disease diagnosis in retinopathy of prematurity. Arch Ophthalmol 125(7):875–880
7. Chollet F (2017) Xception: Deep learning with depthwise separable convolutions. In: Proc IEEE Conf Comput Vis Pattern Recognit, pp 1251–1258
8. Diaz M, Ferrer MA, Impedovo D, Pirlo G, Vessio G (2019) Dynamically enhanced static handwriting representation for Parkinson's disease detection. Pattern Recogn Lett 128:204–210
9. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM et al (2017) Dermatologist-level classification of skin cancer with deep neural networks. Nature 542(7639):115–118
10. Early Treatment For Retinopathy Of Prematurity Cooperative Group (2003) Revised indications for the treatment of retinopathy of prematurity: results of the early treatment for retinopathy of prematurity randomized trial. Arch Ophthalmol 121:1684
11. Fu J, Liu J, Tian H, Fang Z, Lu H (2018) Dual attention network for scene segmentation. In: Proc IEEE Conf Comput Vis Pattern Recognit, pp 3146–3154
12. Girshick R, Donahue J, Darrell T, Malik J (2014) Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proc IEEE Conf Comput Vis Pattern Recognit, pp 580–587
13. Good WV, Hardy RJ, Dobson V, Palmer EA, Phelps DL, Tung B et al (2010) Final visual acuity results in the early treatment for retinopathy of prematurity study. Arch Ophthalmol 128(6):663–671
14. Gschließer A, Stifter E, Neumayer T, Moser E, Papp A, Pircher N et al (2015) Inter-expert and intra-expert agreement on the diagnosis and treatment of retinopathy of prematurity. Am J Ophthalmol 160(3):553–560.e3
15. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A et al (2016) Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA 316(22):2402–2410
16. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proc IEEE Conf Comput Vis Pattern Recognit, pp 770–778
17. Hellström A, Smith LEH, Dammann O (2013) Retinopathy of prematurity. Lancet 382(9902):1445–1457
18. Hu J, Chen Y, Zhong J, Ju R, Yi Z (2018) Automated analysis for retinopathy of prematurity by deep neural networks. IEEE Trans Med Imaging 38(1):269–279
19. Hutchinson AK, Melia M, Yang MB, VanderVeen DK, Wilson LB, Lambert SR (2016) Clinical models and algorithms for the prediction of retinopathy of prematurity: a report by the American Academy of Ophthalmology. Ophthalmology 123(4):804–816
20. I. C. f. t. C. o. R. o. Prematurity (1984) An international classification of retinopathy of prematurity. Arch Ophthalmol 102:1130–1134
21. I. C. f. C. o. L. S. ROP (1987) An international classification of retinopathy of prematurity: II. The classification of retinal detachment. Arch Ophthalmol 105(7):906–912
22. I. C. f. t. C. o. R. o. Prematurity (2005) The international classification of retinopathy of prematurity revisited. Arch Ophthalmol 123(7):991–999
23. Jia X, Shen L, Zhou X, Yu S (2016) Deep convolutional neural network based HEp-2 cell classification. In: (2016) 23rd International Conference on Pattern Recognition (ICPR), pp 77–80

24. Khan MA, Kadry S, Alhaisoni M, Nam Y, Zhang Y, Rajinikanth V et al (2020) Computer-aided gastrointestinal diseases analysis from wireless capsule endoscopy: A framework of best features selection. IEEE Access 8:132850–132859
25. Khan MA, Arshad H, Nisar W, Javed MY, Sharif M (2021) An integrated design of fuzzy C-means and NCA-based multi-properties feature reduction for brain tumor recognition. In: Signal and Image Processing Techniques for the Development of Intelligent Healthcare Systems. Springer, Berlin, pp 1–28
26. Kim SJ, Port AD, Swan R, Campbell JP, Chan RP, Chiang MF (2018) Retinopathy of prematurity: a review of risk factors and their clinical significance. Surv Ophthalmol 63(5):618–637
27. Kimyon S, Mete A (2018) Comparison of bevacizumab and ranibizumab in the treatment of type 1 retinopathy of prematurity affecting zone 1. Ophthalmologica 240(2):1–7
28. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. Adv Neural Inf Process Syst:1097–1105
29. Liaqat A, Khan M, Sharif M, Mittal M, Saba T, Manic K et al (2020) Gastric tract infections detection and classification from wireless capsule endoscopy using computer vision techniques: a review. Curr Med Imaging 16(10):1229–1242
30. Lin T-Y, Goyal P, Girshick R, He K, Dollár P (2017) Focal loss for dense object detection. In: Proc IEEE Int Conf Comput Vis, pp 2980–2988
31. Liu M, Cheng D, Yan W (2018) Classification of Alzheimer's disease by combination of convolutional and recurrent neural networks using FDG-PET images. Front Neuroinform 12:35
32. Long J, Shelhamer E, Darrell T (2015) Fully convolutional networks for semantic segmentation. In: Proc IEEE Conf Comput Vis Pattern Recognit, pp 3431–3440
33. Long E, Lin H, Liu Z, Wu X, Wang L, Jiang J et al (2017) An artificial intelligence platform for the multihospital collaborative management of congenital cataracts. Nat Biomed Eng 1(2):0024
34. Lvd M, Hinton G (2008) Visualizing data using t-SNE. J Mach Learn Res 9(11):2579–2605
35. Majid A, Khan MA, Yasmin M, Rehman A, Yousafzai A, Tariq U (2020) Classification of stomach infections: A paradigm of convolutional neural network along with classical features fusion and selection. Microsc Res Tech 83(5):562–576
36. Masumoto H, Tabuchi H, Nakakura S, Ishitobi N, Miki M, Enno H (2018) Deep-learning classifier with an ultrawide-field scanning laser ophthalmoscope detects glaucoma visual field severity. J Glaucoma 27(7): 647–652
37. Paszke A, Gross S, Chintala S, Chanan G, Yang E, Z. DeVito, et al (2017) Automatic differentiation in pytorch
38. Quinn GE, Gilbert C, Darlow BA, Zin A (2010) Retinopathy of prematurity: an epidemic in the making. Chin Med J (Engl) 123(20):2929–2937
39. Rao J, Fan D, Wu S, Lin D, Zhang H, Ye S et al (2018) Trend and risk factors of low birth weight and macrosomia in south China, 2005–2017: a retrospective observational study. Sci Rep 8(1):3393
40. Roy AG, Navab N, Wachinger C (2019) Recalibrating fully convolutional networks with spatial and channel "Squeeze and Excitation" blocks. IEEE Trans Med Imaging 38(2):540–549
41. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D (2017) Grad-cam: Visual explanations from deep networks via gradient-based localization. In: Proc IEEE Conf Comput Vis Pattern Recognit, pp 618–626
42. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv:preprint arXiv:1409.1556
43. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D et al (2015) Going deeper with convolutions. In: Proc IEEE Conf Comput Vis Pattern Recognit, pp 1–9
44. Szegedy C, Ioffe S, Vanhoucke V, Alemi AA (2017) Inception-v4, inception-resnet and the impact of residual connections on learning. Proc AAAI Conf Artif Intell
45. Ting DS, Wu W-C, Toth C (2018) Deep learning for retinopathy of prematurity screening. Br J Ophthalmol
46. Wang J, Ju R, Chen Y, Zhang L, Hu J, Wu Y et al (2018) Automated retinopathy of prematurity screening using deep neural networks. EBioMedicine 35:361–368
47. Woo S, Park J, Lee J-Y, So Kweon I (2018) Cbam: Convolutional block attention module. In: Proc IEEE Europ Conf Comput Vis, pp 3–19
48. Wu C, Petersen RA, Vanderveen DK (2006) RetCam imaging for retinopathy of prematurity screening. J AAPOS 10(2):107–111
49. Xie S, Girshick R, Dollár P, Tu Z, He K (2017) Aggregated residual transformations for deep neural networks. In: Proc IEEE Conf Comput Vis Pattern Recognit, pp 1492–1500
50. Zagoruyko S, Komodakis N (2016) Paying more attention to attention: Improving the performance of convolutional neural networks via attention transfer. arXiv:preprint arXiv:1612.03928
51. Zahoor S, Lali IU, Khan M, Javed K, Mehmood W (2020) Breast cancer detection and classification using traditional computer vision techniques: a comprehensive review. Curr Med Imaging 16(10):1187–1200

52. Zeiler MD, Fergus R (2014) Visualizing and understanding convolutional networks. In: Proc IEEE Europ Conf Comput Vis, pp 818–833
53. Zhang Y, Wang L, Wu Z, Zeng J, Chen Y, Tian R et al (2018) Development of an automated screening system for retinopathy of prematurity using a deep neural network for wide-angle retinal images. IEEE Access 7:10232–10241
54. Zhang H, Goodfellow I, Metaxas D, Odena A (2018) Self-attention generative adversarial networks. In: International conference on machine learning, pp 7354–7363
55. Zheng X, Chen W, You Y, Jiang Y, Li M, Zhang T (2020) Ensemble deep learning for automated visual classification using EEG signals. Pattern Recognit 102:107147
56. Zheng X, Chen W, You Y, Jiang Y, Li M, Zhang T (2020) Ensemble deep learning for automated visual classification using EEG signals. Pattern Recognit 102:107147

Affiliations

Baiying Lei[1] · Xianlu Zeng[2] · Shan Huang[1] · Rugang Zhang[1] · Guozhen Chen[1] · Jinfeng Zhao[2] · Tianfu Wang[1] · Jiantao Wang[2] · Guoming Zhang[2]

[1] School of Biomedical Engineering, Health Science Center, National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, Guangdong Key Laboratory for Biomedical Measurements and Ultrasound Imaging, Shenzhen University, Nanhai Ave 3688, Shenzhen, Guangdong 518060, China

[2] Shenzhen Key Ophthalmic Laboratory, Health Science Center, Shenzhen Eye Hospital, Shenzhen University, The Second Affiliated Hospital of Jinan University, Shenzhen, China