# Automated retinopathy of prematurity screening using deep neural network with attention mechanism

Peng, Yuanyuan, Zhu, Weifang, Chen, Feng, Xiang, Daoman, Chen, Xinjian

Yuanyuan Peng, Weifang Zhu, Feng Chen, Daoman Xiang, Xinjian Chen, "Automated retinopathy of prematurity screening using deep neural network with attention mechanism," Proc. SPIE 11313, Medical Imaging 2020: Image Processing, 1131321 (10 March 2020); doi: 10.1117/12.2548290

Event: SPIE Medical Imaging, 2020, Houston, Texas, United States

# Automated retinopathy of prematurity screening using deep neural network with attention mechanism

Yuanyuan Peng[1#], Weifang Zhu[1 #], Feng Chen[3]，Daoman Xiang[3,]*，Xinjian Chen[1,2,]*

[1]School of Electronics and Information Engineering, Soochow University, Suzhou, 215006, China
[2]State Key Laboratory of Radiation Medicine and Protection, Soochow University, Suzhou, 215123, China
[3]Guangzhou Women and Children Medical Center, Guangzhou, 510623, China

## ABSTRACT

Retinopathy of prematurity (ROP) is an ocular disease which occurs in premature babies and is considered as one of the largest preventable causes of childhood blindness. However, insufficient ophthalmologists are qualified for ROP screening, especially in developing countries. Therefore, automated screening of ROP is particularly important. In this paper, we propose a new ROP screening network, in which pre-trained ResNet18 is taken as backbone and a proposed attention block named Complementary Residual Attention Block (CRAB) and Squeeze-and-Excitation (SE) block as channel attention module are introduced. Our main contributions are: (1) Demonstrating the 2D convolutional neural network model pre-trained on natural images can be fine-tuned for ROP screening. (2) Based on the pre-trained ResNet18, we propose an improved scheme combining which that effectively integrates attention mechanism for ROP screening. The proposed classification network was evaluated on 9794 fundus images from 650 subjects, in which 8351 are randomly selected as training set according to subjects and others are selected as testing set. The results showed that the performance of the proposed ROP screening network achieved 99.17% for accuracy, 98.65% for precision, 98.31% for recall, 98.48% for F1 score and 99.84% for AUC. The preliminary experimental results show the effectiveness of the proposed method.

**KEYWORDS:** Retinopathy of prematurity, neural networks, attention mechanism, fundus image, automated ROP screening

## 1. INTRODUCTION

Retinopathy of Prematurity (ROP) is a vascular proliferative disease affecting premature and low birth-weight (less than 1500g) infants. According to [1][2][3][4], abnormal retinas of prematurity include five stages of ROP and a type of ancillary illness called plus disease, which are shown in Figure 1.

The vast majority of automated or semi-automated methods for ROP diagnosis are focused on plus disease which is defined by the abnormality of vessels. For example, a system called “ROPTool” has been proposed in [5] to assist ophthalmologists in diagnosing plus disease and “i-ROP” [6] was is a system designed to grade plus disease into three types: normal, pre-plus, and plus. In recent years, several studies have used ImageNet pre-trained DNNs for the screening of ROP. For example, Wang et al used Inception-V2 pre-trained on ImageNet to recognize the existence and severity of ROP[7] and Zhang et al used VGG16 pre-trained on ImageNet to automated screening of ROP[8]. Similar to [7] and [8], in this paper, We adopt a pre-trained ResNet18[9] for ROP screening. In order to further improve the feature extraction ability in high-level stage without noteworthy increase of complexity and computation of network, the attention mechanism[10] is introduced, which can help model optimize intermediate features.

*Corresponding author: E-mail: xjchen@suda.edu.cn, # indicates these authors contributed equally to this work

Medical Imaging 2020: Image Processing, edited by Ivana Išgum, Bennett A. Landman,
Proc. of SPIE Vol. 11313, 1131321 · © 2020 SPIE · CCC
code: 1605-7422/20/$21 · doi: 10.1117/12.2548290

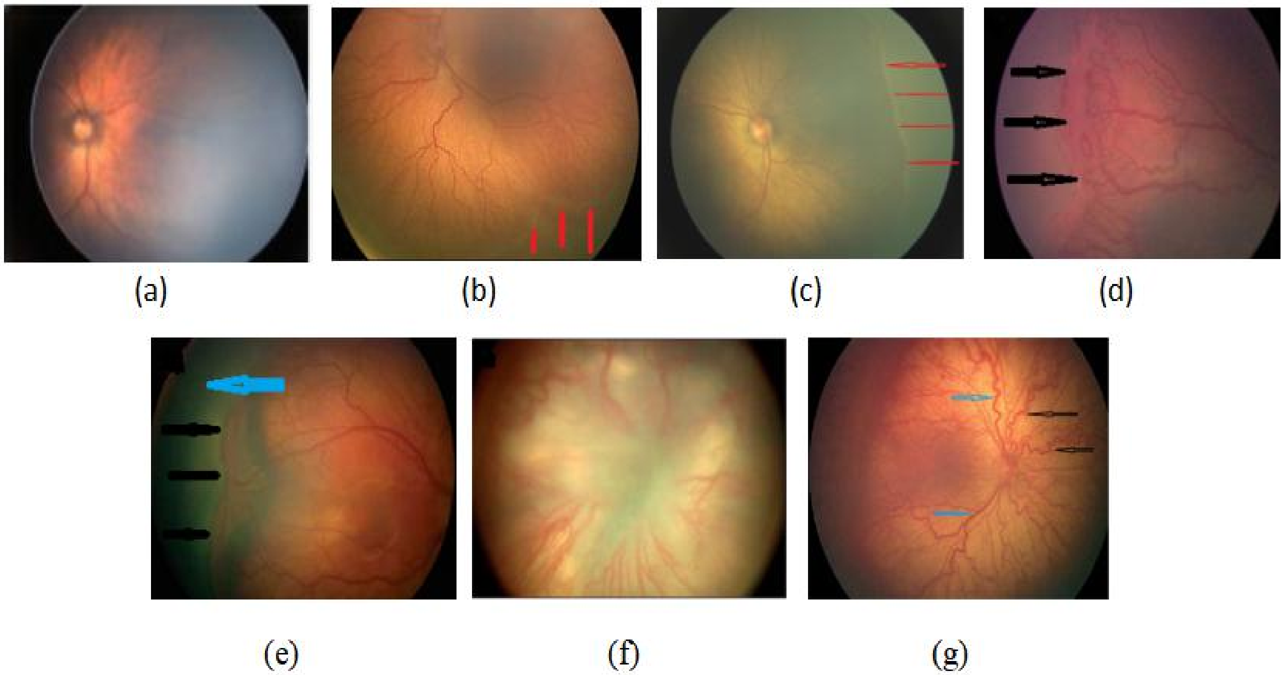


Figure 1. Examples of normal retina, different stages of ROP and retina with plus disease. (a) Normal. (b) Stage 1.(c) Stage 2. (d) Stage 3. (e) Stage 4. (f) Stage 5. (g) Plus disease.

## 2. METHODS

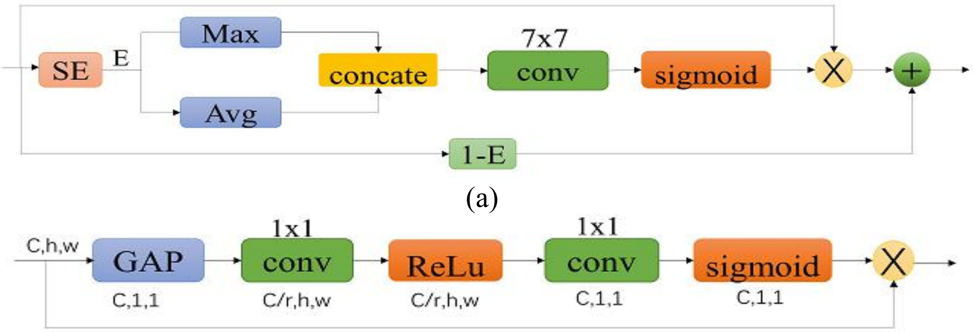
In this section, the proposed method is described as three parts: structure of ResNet18, the proposed attention block and structure of the proposed deep network.

### 2.1 Structure of ResNet18

ResNet18 is a network with 17 convolutional layers, 1 max-pooling, 1 avg-pooling layer, 17 BatchNorm2d layers, 1 fully connected layer, and a softmax output layer. All 2D convolution kernels are 3 × 3 with stride 1 or 2 except for the first convolutional layer, in which convolution kernel is 7 × 7 with stride 2 in spatial dimensions. Max-pooling kernel is 3 × 3 with stride 2 in order to reduce the number of parameters and enhance robustness.

### 2.2 The proposed attention block

Inspired by Convolutional Block Attention module (CBAM)[11], we propose a new attention block named Complementary Residual Attention Block (CRAB), which combines channel and spatial attention mechanism (shown in Figure 2(a)). The difference between our CRAB and CBAM is that we add the complementary residual connection and replace the first half of CBAM with SE-block[12]. In this way, the importance of channel and spatial direction can be obtained by learning, the network can focus on key information without losing other secondary information in our task. In order to enhance important channel information and suppress irrelevant information, we also introduce channel attention module named SE-block from [7], which is shown in Figure 2(b). In Figure 2, r and c represent compression ratio and channel numbers, respectively. In our experiments, the compression ratio r is set to 32.



(a)

(b)

Figure 2 Attention block. (a) CRAB. (b) SE-block

### 2.3 Structure of the proposed deep network

The architecture of our network is shown in Table 1. As shown in Table 1, the shape and operation with specific parameter settings of a residual building block are listed inside the brackets and the number of stacked blocks in a stage is presented outside. The inner brackets following with CRAB and SE-block indicate the compression ratio and the output dimension.

Table 1. The diagram of different network.

| Output size | Backbone | Backbone+CRAB | Backbone+SE | Backbone+All |
|---|---|---|---|---|
| 160 × 120 | Conv, 7 × 7, 64, stride=2 | | | |
| 80 × 60 | Maxpool, 3 × 3, 64, stride=2 | | | |
| | $\begin{bmatrix}\text{conv},3\times3,64\\ \text{conv},3\times3,64\end{bmatrix}\times2$ | | | |
| 40 × 30 | $\begin{bmatrix}\text{conv},3\times3,128\\ \text{conv},3\times3,128\end{bmatrix}\times2$ | | | |
| 20 × 15 | $\begin{bmatrix}\text{conv},3\times3,256\\ \text{conv},3\times3,256\end{bmatrix}\times2$ | $\begin{bmatrix}\text{conv},3\times3,256\\ \text{conv},3\times3,256\end{bmatrix}\times2$<br>CRAB,[32,256] | $\begin{bmatrix}\text{conv},3\times3,256\\ \text{conv},3\times3,256\end{bmatrix}\times2$ | $\begin{bmatrix}\text{conv},3\times3,256\\ \text{conv},3\times3,256\end{bmatrix}\times2$<br>CRAB,[32,256] |
| 10 × 8 | $\begin{bmatrix}\text{conv},3\times3,512\\ \text{conv},3\times3,512\end{bmatrix}\times2$ | $\begin{bmatrix}\text{conv},3\times3,512\\ \text{conv},3\times3,512\end{bmatrix}\times2$ | $\begin{bmatrix}\text{conv},3\times3,512\\ \text{conv},3\times3,512\end{bmatrix}\times2$<br>SE - block,[32,512] | $\begin{bmatrix}\text{conv},3\times3,512\\ \text{conv},3\times3,512\end{bmatrix}\times2$<br>SE - block,[32,512] |
| 1 × 1 | Average pool, 2-d fc, softmax | | | |

## 3. RESULTS

### 3.1 Datasets and preprocessing

The ROP fundus images (640 × 480 ×3) were acquired using RetCam3 from the Guangzhou Women and Children Medical Center, which are labeled as normal or ROP by three ophthalmologists including a chief physician and two attending physicians.

8351 (4752 normal and 3599 ROP) fundus images from 550 subjects are used in training stage, which are divided as training dataset and validation dataset according to the ratio of 7:3. 1443 (850 normal and 593 ROP) fundus images from 100 subjects are used for model testing. In order to reduce the computational cost, all fundus images are downsampled to 320 ×240 × 3 using bilinear interpolation. To prevent over-fitting and enhance the generalization ability of the model, online data augmentation has been performed, including random rotation of $30^{\circ}$, horizontal flipping, vertical flipping and affine transformations.

### 3.2 Parameter settings

The encoder of our proposed model is based on pre-trained ResNet18. The implementation of the proposed network is based on the public platform PyTorch and NVIDIA Tesla K40 GPU with 12GB memory. We train the model with back-propagation algorithm by minimizing the cross-entropy cost function:

$$L = -\frac{1}{M}\sum_{j=1}^{M} y_j^T \ln(a_j^L) \tag{1}$$

Where $a_j^L$ denotes the jth output of the network after applying the softmax function. The cross-entrop y represents the similarity between the true distribution of labels and the approximated distribution of the network. Adam is used as the optimizer to minimize the cost function. Both initial learning rate and weight decay are set to 0.0001 to optimize the network. The batch size and epoch are set to 64 and 30, respectively. During training, all networks are trained with identical optimization schemes an d we save the best model on validation set.

### 3.3 Evaluation metrics

To quantitatively evaluate the performance of our method, we compare classification results of the original ResNet18 (trained from scratch and pre-trained on ImageNet) and the improved ResNet18 according to the following five metrics: accuracy, precision, recall, F1 score and area under curve (AUC). The accuracy, precision, recall and F1 score are defined as:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{2}$$

$$Precision = \frac{TP}{TP+FP} \tag{3}$$

$$Recall = \frac{TP}{FP+FN} \tag{4}$$

$$F1-score = \frac{2*P*R}{P+R} \tag{5}$$

Where TP, FP, TN and FN represent true positive, false positive, true negative and false negative, respectively.

### 3.4 Results

Figure 3 shows the validation accuracy curves of ResNet18 from scratch and pre-trained ResNet18, which indicates that transfer learning is effective in our ROP screening. We apply the class activation map to visualize the heat map of class activation. As shown in Figure 4, the visualization results demonstrate that our model can extract the potential features for ROP images.

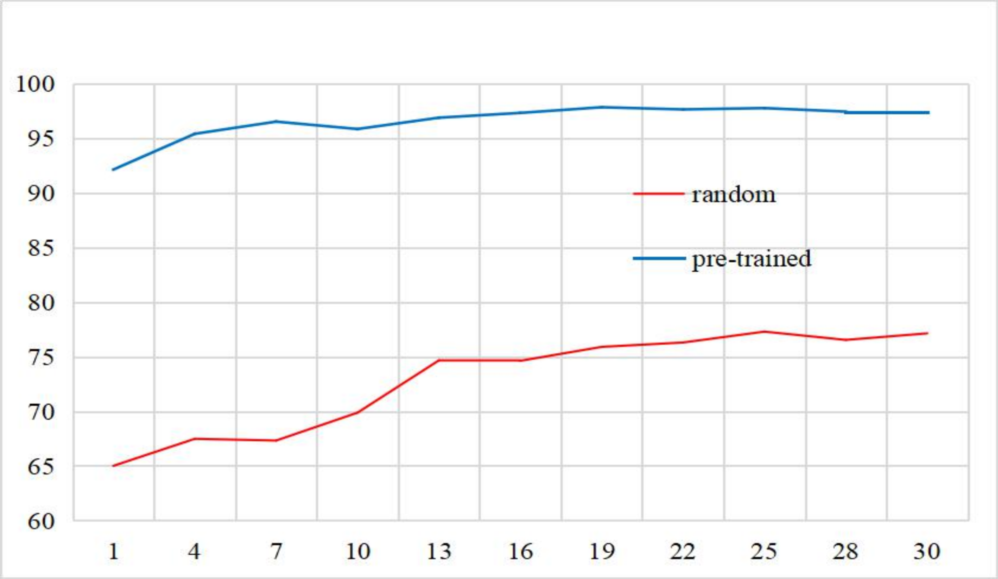


Figure 3. Validation accuracy curves.

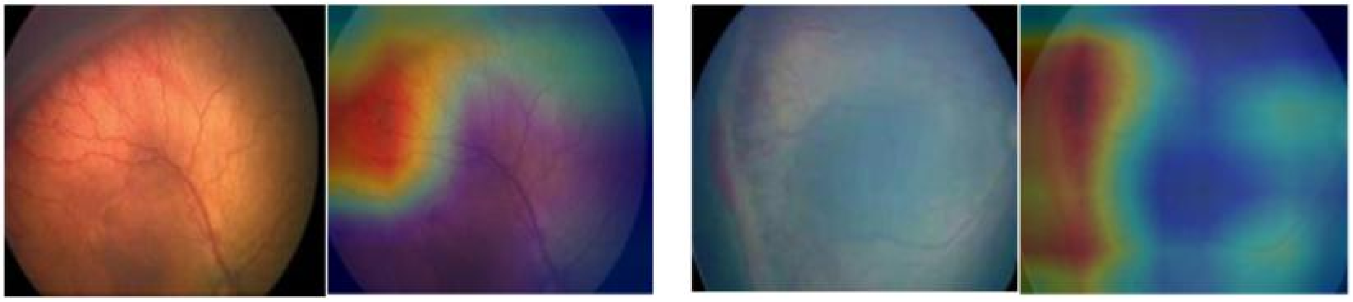

(a) (b) (c) (d)

Figure 4. Heat map of class activation. (a) and (c) Original ROP fundus images. (b) and (d) Corresponding heat map. The red regions in (b) and (d) represent the primary focus of the network.

In order to evaluate the effectiveness of CRAB and SE-block in the pre-trained ResNet18, a series of ablation studies are conducted. We call the fine-tuned ResNet18 as 'Backbone'. As shown in Table 2, compared to Backbone, Backbone with CRAB (denoted as Backbone+CRAB) improves the accuracy, precision, recall, F1 score and AUC by 0.84%, 0.52%, 1.52%, 1.02% and 0.02%, respectively. Table 2 also shows the effectiveness of SE-block, which improves the most performances of classification. The proposed network (denoted as Backbone+All) has the highest recall with 98.31%, which is clinically important.

To further illustrate the effectiveness of the proposed method, we also compare our method with the automated screening of ROP system [8], which is based on VGG16 pre-trained on ImageNet and optimized by the Adam algorithm. As shown in table 3, the proposed method outperform Zhang et al.' method on all metrics. In terms of recall, the performance of the proposed method is 3.04% higher than that of Zhang et al.' method. And the parameters of our method is much smaller than Zhang et al.' method.

Table 2. Classification Results.

| Methods | Accuracy | Precision | Recall | F1 score | AUC | Parameters (M) |
|---|---|---|---|---|---|---|
| ResNet18_Scratch | 76.51% | 70.35% | 74.03% | 72.14% | 84.28% | 11.177538 |
| Backbone | 98.19% | 99.30% | 96.29% | 97.78% | 99.78% | **11.177538** |
| Backbone+CRAB | 99.03% | **99.82%** | 97.81% | **98.80%** | 99.80% | 17.60416 |
| Backbone+SE | 98.96% | 98.31% | 97.81% | 98.06% | 99.77% | 11.193922 |
| Backbone+All | **99.17%** | 98.56% | **98.31%** | 98.48% | **99.84%** | 17.620546 |

Table 3. Comparison of the proposed method with Zhang et al.' method

| Methods | Accuracy | Precision | Recall | F1 score | AUC | Parameters (M) |
|---|---|---|---|---|---|---|
| Zhang et al. | 97.43% | 98.43% | 95.27% | 96.82% | 99.68% | 134.263738 |
| Proposed | **99.17%** | **98.56%** | **98.31%** | **98.48%** | **99.84%** | **17.620546** |

## 4. CONCLUSIONS

In this paper, we proposed a pre-trained ResNet18 based on network for ROP screening. Firstly, we introduce the pre-trained ResNet18 as backbone and our workthe results demonstrates that the pre-trained ResNet18 can be used effectively in automated screening of ROP. Secondly, we propose a new attention block named Complementary Residual Attention Block (CRAB) inspired by Convolutional Block Attention module (CBAM), which combines channel and spatial attention mechanism. In this way, the network can focus on key information without losing other secondary information in our task and output semantically rich feature map. Finally, we also introduce channel attention module named SE-block in order to enhance important channel information and suppress irrelevant information. The experimental results demonstrate the effectiveness and practicability of the proposed method. The proposed method provides a promising technology which can assist pediatric ophthalmologists in ROP screening. Automated ROP grading is our further research in the near future.

## 5. ACKNOWLEDGEMENTS

This work was supported in part by the National Key R&D Program of China under Grant 2018YFA0701700, and in part by the National Basic Research Program of China (973 Program) under Grant 2014CB748600, the National Natural Science Foundation of China (NSFC) under Grant 61622114 and 81401472.

## 6. REFERENCE

1. Flynn J T. An international classification of retinopathy of prematurity: clinical experience. Ophthalmology, 1985, 92(8): 987-994.
2. Committee for the Classification of Retinopathy of Prematurity, An international classification of retinopathy of prematurity, Arch. Ophthalmol., 1984,vol. 102, no. 8, pp. 1130–1134.
3. ICROP Committee for Classification of Late Stages ROP, An international classification of retinopathy of

prematurity, II: The classification of retinal detachment, Arch. Ophthalmol., 1987, vol. 105, no. 7, pp. 906-912.

4. J. P. Campbell et al., “Plus disease in retinopathy of prematurity: A continuous spectrum of vascular abnormality as a basis of diagnostic variability,” Ophthalmology, , 2016, vol. 123, no. 11, pp. 2338–2344.

5. Wallace D K, Zhao Z, Freedman S F. A pilot study using “ROPtool” to quantify plus disease in retinopathy of prematurity. Journal of American Association for Pediatric Ophthalmology and Strabismus, 2007, 11(4): 381-387.

6. E. Ataer-Cansizoglu *et al.*, “Computer-based image analysis for plus disease diagnosis in retinopathy of prematurity: Performance of the ‘i-ROP’ system and image features associated with expert diagnosis,” Transl. Vis. Sci. Technol., 2015,vol. 4, no. 6, p. 5.

7. Hu J, Chen Y, Zhong J, et al. Automated analysis for retinopathy of prematurity by deep neural networks. IEEE transactions on medical imaging, 2018, 38(1): 269-279.

8. Zhang Y, Wang L, Wu Z, et al. Development of an Automated Screening System for Retinopathy of Prematurity Using a Deep Neural Network for Wide-Angle Retinal Images. IEEE Access, 2018, 7: 10232-10241.

9. He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition, 2016: 770-778.

10.Cho K.Van Merrienboer B. Gulcehre C. et al. Learning Phase Representations using RNN Encoder-Decoder for Statistical Machine Translation.[J]. Computer Science, 2014.

11. Woo S, Park J, Lee J Y, et al. Cbam: Convolutional block attention module. Proceedings of the European Conference on Computer Vision, 2018: 3-19.

12. Hu J, Shen L, Sun G. Squeeze-and-excitation networks. Proceedings of the IEEE conference on computer vision and pattern recognition, 2018: 7132-7141.