PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

Automated zone recognition for retinopathy of prematurity using deep neural network with attention mechanism and deep supervision strategy

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

Yuanyuan Peng, Weifang Zhu, Feng Chen, Xinjian Chen, "Automated zone recognition for retinopathy of prematurity using deep neural network with attention mechanism and deep supervision strategy," Proc. SPIE 11596, Medical Imaging 2021: Image Processing, 115961S (15 February 2021); doi: 10.1117/12.2580809

SPIE.

Event: SPIE Medical Imaging, 2021, Online Only

Automated zone recognition for retinopathy of prematurity using deep neural network with attention mechanism and deep supervision strategy

Yuanyuan Peng[1], Weifang Zhu[1], Feng Chen[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 the main cause of blindness in children worldwide. The severity of ROP can be reflected by staging, zoning and plus disease. Specially, some studies have shown that zone recognition is more important than staging. However, due to the subjective factors, ophthalmologists are often inconsistent in their recognition of zones according to fundus images. Therefore, automated zones recognition of ROP is particularly important. In this paper, we propose a new ROP zones recognition network, in which pre-trained DenseNet121 is taken as backbone and a proposed attention block named Spatial and Channel Attention Block (SACAB) and deep supervision strategy are introduced. Our main contributions are: (1) Demonstrating the 2D convolutional neural network model pre-trained on natural images can be fine-tuned for automated zones recognition of ROP. (2) Based on pre-trained DenseNet121, we propose two improved schemes which effectively integrate attention mechanism and deep supervision learning for ROP zoning. The proposed method was evaluated on 662 retinal fundus images (82 zone I, 299 zone II, 281 zone III) from 148 examinations with 5-fold cross validation strategy. The results show that the performance of the proposed ROP zone recognition network achieves 0.8852 for accuracy (ACC), 0.8850 for weighted F1 score (W_F1) and 0.8699 for kappa. The preliminary experimental results show the effectiveness of the proposed method.

KEYWORDS: Retinopathy of prematurity, Neural Networks, Deep Learning, Attention Mechanism, Deep Supervision Strategy, Fundus Image, Automated ROP Zone Recognition

Medical Imaging 2021: Image Processing, edited by Ivana Išgum,
Bennett A. Landman, Proc. of SPIE Vol. 11596, 115961S · © 2021
SPIE · CCC code: 1605-7422/21/$21 · doi: 10.1117/12.2580809

1. INTRODUCTION

Retinopathy of prematurity (ROP) is a vascular proliferative disease, which is one of the most dangerous and serious ocular complications in premature infants[1]. According to [2], abnormal retinas of prematurity mainly includes five stages of ROP, a type of ancillary illness called plus disease and three zones. Specially, the international classification of ROP (ICROP) defines three zones according to the location of the symptom in ROP with each centered on the optic disc, which are detailed in Table 1. Figure 1 shows the examples of three zones of ROP.

Many previous studies based on automated or semi-automated methods for ROP diagnosis are mainly focused on plus disease, which is defined by the abnormality of vessels. For example, Mao et al. used the deep learning network to segment retinal vessels and optic discs and diagnoses plus disease based on the automatic quantitative analysis of vascular curvature, width and other pathological characteristics[3]. In recent years, several researchers used DNNs based on transfer learning for ROP screening and severity identification. For example, our previous work adopted a ResNet18 pre-trained on ImageNet with attention mechanism to automated screening of ROP[4] and Wang et al used a pre-trained Inception-V2 to recognize the existence and severity of ROP[5]. Zhao et al. used ResNet50 pre-trained on Microsoft COCO dataset to automatically draw the boundary of zone I on the fundus image as a diagnostic aid[6]. However, as far as we know, the study of identifying three zones of ROP has not been reported, which is important for the evaluation of ROP severity. In this paper, we propose a simple and novel deep learning based method for automated zones recognition of ROP. We adopt transfer learning, attention mechanism and deep supervision strategy, which can speed up the convergence of the model, help model optimize high-level feature and make full use of the shallow feature information, respectively.

TABLE 1. Definitions of zone I to III of ROP

Zone	Definitions
zone I	A circular area with a radius of twice the distance from the center of the optic disc to the fovea of the macula.
zone II	A annular area with a radius of the distance from the optic disc to the nasal serrated margin except zone I.
zone III	The remaining crescent shaped areas outside zone I and II

2. METHODS

In this section, the proposed method is described as four parts: structure of the proposed deep network， structure of DenseNet121, the proposed attention block and deep supervised strategy.

*Corresponding author: E-mail: xjchen@suda.edu.cn

(a) (b) (c)

Figure 1. Examples of Zone I, Zone II and Zone III. (a) Zone I. (b) Zone II. (c) Zone III.

2.1 Structure of the proposed deep network

The architecture of our proposed ROP recognition network is shown in Figure 3, which is based on DenseNet121. On the basis of DenseNet121 network, the SACAB attention block is designed and embedded into the middle layer of convolutional neural network. In addition, deep supervision strategy is introduced by adding auxiliary classifiers after some intermediate convolution layers.

2.2 Structure of DenseNet121

DenseNet121 is basic network in our study, which consists of convolutional layers, max-pooling layer, avg-pooling layer, and fully connected layer. In order to adapt the DenseNet121[7] pre-trained on ImageNet dataset with 1000 categories to our three classification, we adjust the output of fully connected layer to 3 categories.

2.3 The proposed attention block

Considering that zoning is related to the content and location of lesions, it is similar to channel attention and spatial attention in attention mechanism. Inspired by Dual Attention Network (DANet)[8], we propose a new attention block named Spatial and Channel Attention Block (SACAB), which combines channel and spatial attention mechanism and can complement each other (shown in Figure 2). There are three main differences between our SACAB and DANet. First, for the spatial attention module, we directly multiply the original feature map and the learned similarity function to obtain the spatial attention matrix. Second, for the channel attention module, we use SE-block[9] to replace the channel attention module part of DANet. Finally, we fuse the channel attention feature map and spatial attention feature map by average operation instead of addition operation. In this way, the network can enhance important information and suppress irrelevant information. In Figure 2, r and c represent compression ratio and channel numbers, respectively. In our experiments, the compression ratio r is set to 16.

2.4 Deep supervision strategy

Previous studies have shown that deep supervised learning is helpful to train deep network and can improve its classification accuracy[10,11]. On one hand, for small training data and relatively shallow network, deep supervision can provide powerful "normalization" for classification. On the other hand, for large-scale training data and deeper network, deep supervision can make the network converge better and improve the classification performance. In order to further improve the accuracy of ROP zone recognition, deep supervision strategy is adopted by increasing auxiliary classifier after some of the intermediate convolutional layers in our proposed network. As can be seen from Figure 3, we add two auxiliary classification as supervision branches, which allows more shallow information to be used.

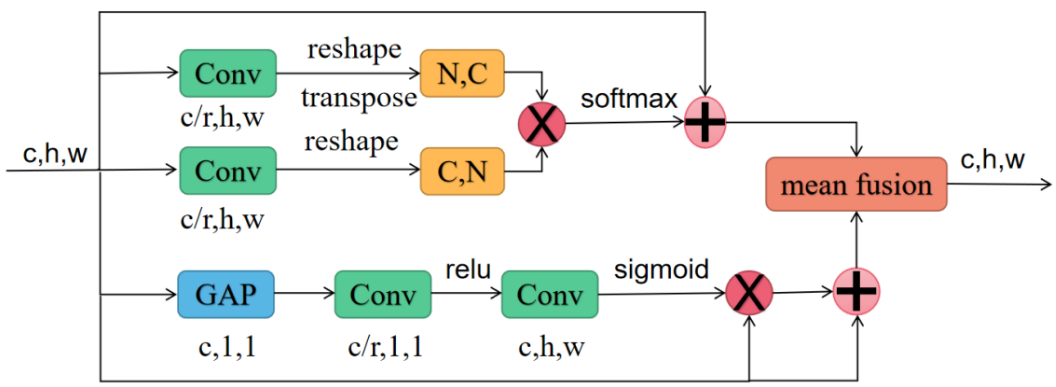

Figure 2. Spatial and Channel Attention Block (SACAB).

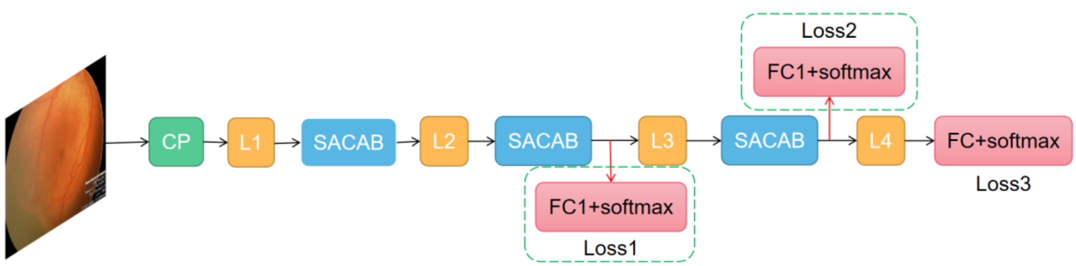

Figure 3. The architecture of our model based on DenseNet121. The additional supervision loss branches are in the dashed green boxes. ‘CP’ represents convolution and pooling operation and ‘L’ represents multiple stacked dense connection modules.

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 were labeled as zone I, zone II or zone III by two ophthalmologists including a chief physician and an attending physician.

662 (82 zone I, 299 zone II, 281 zone III) fundus images from 150 examinations are used to train and evaluate the performance of model. In order to reduce the computational cost, all fundus images are downsampled to 256 ×256 × 3 by using bilinear interpolation. To prevent over-fitting and enhance the generalization ability of the model, online data augmentation has been performed, including random rotation 30°, horizontal flipping and vertical flipping.

3.2 Parameter settings

The proposed network is based on DenseNet121with ImageNet pre-trained weights. The implementation of the proposed network is based on the public platform PyTorch and a NVIDIA Tesla K40 GPU with 12GB memory. The model is trained with back-propagation algorithm by minimizing the loss function, which is an average of the overall (final-layer) loss and two companion losses associated with two intermediate layers. The loss function is defined as follow:

$$\mathrm{L} = \frac{1}{3}\sum_{i=1}^{3} L_i \tag{1}$$

Where,

$$L_i = -\frac{1}{M}\sum_{j=1}^{M} y_j^T \ln(a_j^L) \tag{2}$$

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. Both initial learning rate and weight decay are set to 0.0001. The batch size and epoch are set to 16 and 40, respectively. During training, all networks are trained with identical optimization schemes and the best model is saved on validation set with 5-fold cross validation.

3.3 Results

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.

In order to evaluate the effectiveness of SACAB and deep supervision strategy in the pre-trained DenseNet121, a series of ablation studies are conducted. We call the fine-tuned DenseNet121 as 'Backbone'. As shown in Table 2, compared to Backbone, Backbone with SACAB (Backbone+SACAB) improves the accuracy, weighted F1 score and kappa by 2.11%, 2.13% and 2.22%, respectively. Specially, compared to Backbone with DANet (Backbone+DANet), Backbone with SACAB (Backbone+SACAB) improves the accuracy, weighted F1 score and kappa by 0.61%, 0.61%, and 0.58%, respectively. Table 2 also shows the effectiveness of deep supervision strategy (Backbone+DS), which improves two of the three quantitative indicators. The proposed network (Backbone+All) performs well in terms of all quantitative metrics, which demonstrates the effectiveness of the proposed method in the zone recognition of ROP.

Table 2. Zoning Results.

Models	ACC	W_F1	kappa	Parameters (M)
Backbone	0.8566±0.0268	0.8564±0.0265	0.8383±0.0383	**6.956931**
Backbone+DANet	0.8716±0.0197	0.8716±0.0195	0.8547±0.026	7.345071
Backbone+SACAB	0.8777±0.0114	0.8777±0.0111	0.8605±0.0244	7.388022
Backbone+DS	0.8641±0.0222	0.8643±0.0222	0.8379±0.0118	6.959241
Backbone+All	**0.8852±0.0194**	**0.8850±0.0194**	**0.8699±0.0264**	7.390332

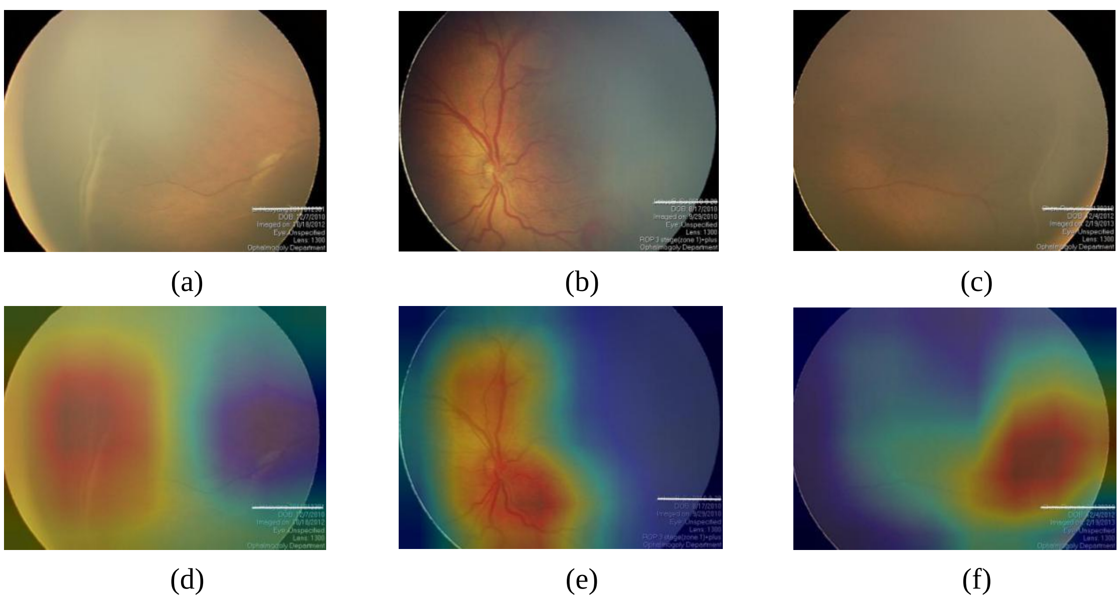

Figure 4. Heat maps of class activation. (a), (b) and (c) Original ROP fundus images. (d), (e) and (f) Corresponding heat maps. The red regions in (d), (e) and (f) represent the primary focus of the network.

4. CONCLUSIONS

In this paper, we propose a DenseNet121 based network for ROP zone recognition, which is pre-trained on ImageNet. Firstly, we introduce the pre-trained DenseNet121 as backbone for the automated zone recognition of ROP. Secondly, we propose a new attention block named Spatial and Channel Attention Block (SACAB) inspired by Dual Attention Network (DANet)[8], which combines channel and spatial attention mechanism and can complement each other. In this way, the network can focus on the key information of the ROP zone in our task and output rich semantic feature map. Finally, we also introduce deep supervision learning to further improve the classification accuracy, which can make full use of shallow information. The experimental results demonstrate the effectiveness and feasibility of the proposed method. The proposed method provides a promising technology which can assist pediatric ophthalmologists in ROP diagnosis. In the near future, we focus on the further study of ROP and AP-ROP (aggressive posterior retinopathy of prematurity) staging.

5. ACKNOWLEDGEMENTS

This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFA0701700, in part by the National Nature Science Foundation of China under Grant 61622114, and in part by the National Basic Research Program of China under Grant 2014CB748600.

6. REFERENCE

1. Zarei, Mohammad, et al. "Prevalence and Risk Factors of Retinopathy of Prematurity in Iran." *Journal of ophthalmic & vision research* 14(3) (2019): 291-298.
2. International Committee for the Classification of Retinopathy of Prematurity. "The international classification of retinopathy of prematurity revisited." *Archives of ophthalmology (Chicago, Ill.: 1960)* 123(7) (2005): 991-999.
3. Mao, Jianbo, et al. "Automated diagnosis and quantitative analysis of plus disease in retinopathy of prematurity based on deep convolutional neural networks." *Acta Ophthalmologica* 98(3) (2020): e339-e345.
4. Peng, Yuanyuan, et al. "Automated retinopathy of prematurity screening using deep neural network with attention mechanism." *Medical Imaging 2020: Image Processing*. Vol. 11313. International Society for Optics and Photonics, (2020): 1131321-1-7.
5. Hu, Junjie, et al. "Automated analysis for retinopathy of prematurity by deep neural networks." *IEEE transactions on medical imaging* 38 (1) (2018): 269-279.
6. Zhao, Jinfeng, et al. "A deep learning framework for identifying zone I in RetCam images." *IEEE Access* 7 (2019): 103530-103537.
7. Huang, Gao, et al. "Densely connected convolutional networks." *Proceedings of the IEEE conference on computer vision and pattern recognition*. (2017): 4700-4708.
8. Fu, Jun, et al. "Dual attention network for scene segmentation." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. (2019): 3146-3154.
9. Hu J, Shen L, Sun G. "Squeeze-and-excitation networks." Proceedings of the IEEE conference on computer vision and pattern recognition, (2018): 7132-7141.
10. Wang, Liwei, Chen-Yu Lee, Zhuowen Tu, and Svetlana Lazebnik. "Training deeper convolutional networks with deep supervision." *arXiv preprint arXiv:1505.02496* (2015).

11. Lee, Chen-Yu, Saining Xie, Patrick Gallagher, Zhengyou Zhang, and Zhuowen Tu. "Deeply-supervised nets." In *Artificial intelligence and statistics*, (2015): 562-570.