Automated Recognition of Retinopathy of Prematurity with Deep Neural Networks

Yifan Wang¹, Yuanyuan Chen¹*

¹ Machine Intelligence Laboratory, College of Computer Science, Sichuan University, Chengdu, Sichuan, 610065, P.R.China

*Corresponding author’s e-mail: chenyuanyuan@scu.edu.cn

Abstract. Retinopathy of Prematurity (ROP) is a blinding disease, which primarily occurs on premature infants whose birth weights is less than 1250 grams or gestation is less than 31 weeks. ROP has become the leading cause of preventable childhood blindness throughout the world. Nowadays, more and more researchers start attempting to develop auto or semi-auto methods based on digital image analysis to diagnose ROP. However, factors like high measurement errors or redundant analysis phrases make traditional analysis methods difficult to assist diagnose ROP perfectly. In this paper, we develop an automated system to analyse premature infants’ retinal images using deep neural networks. We primarily try to solve two problems. (1) the existence of ROP, normal or ROP; (2) the severity of ROP, mild-ROP or severe-ROP. Deep neural networks take the advantages of strong representation ability and enable to nonlinear mapping, attain high accuracies and great generalization performances on retinal fundus image datasets.

1. Introduction

Retinopathy of Prematurity (ROP) is a proliferative retinal vascular disease that occurs when abnormal blood vessels grow and spread throughout the retina. Premature infants separate themselves from the maternal environment too early, so the edges of the retina may not get enough oxygen and nutrients. Doctors believe that the periphery of the retina then sends out signals to other areas of the retina for nourishment. As a result, new abnormal vessels begin to grow. These new blood vessels are fragile, weak and can bleed, leading to retinal scarring. When these scars shrink, they pull on the retina, causing it to detach from the back of the eyes.

Nowadays, ROP has become a potentially blinding eye disorder that primarily affects premature infants weighing about 2.75 pounds (1250 grams) or less and being born before 31 weeks of gestation. The shorter gestation of baby, the more likely that baby is to develop ROP.

According to ROP classification guideline [4], 5 stages are used to describe the severity degree of ROP. Stage 1 has a thin but definite structure demarcation line, with recognisable abnormal branching or arcading of vessels. The appearance of Stage 2 is that the ridges arise from the demarcation line and have height and width. Stage 3 demonstrates that neovascularization extends from the ridge into the vitreous. Stage 4 occurs partial retinal detachment. And Stage 5 occurs total Retinal detachment.

The diagnosis of ROP is based on the retinal fundus images from premature infants. A simple screening test and timely treatment by an ophthalmologist can prevent from blindness [8]. The early diagnose of ROP helps to reduce the vision loss of premature infants. However, ROP screening is facing two challenges. Firstly, clinical assessment and grading is difficult, the reason is that the quality of images such as type of lens, focus of image, size of optic nerve, pigmentation and other disease characteristics may influence final clinical assessment badly [1].
Due to lack of gold standard, high inter-variability between ophthalmologists causing ROP diagnosis is subjective and prone to low reliability [5]. Secondly, in developing countries like China, India, the large number of population has a conflict with limited medical resources. However, with the rise of birth rate of premature infants, more and more babies cannot gain enough treatment. On the other hand, labor-intensive work is prone to errors occurring.

2. Related Work
Nowadays, due to the excellent results convolutional neural network (CNN) has achieved, more and more researchers start to focus on how to apply CNN on medical image process field. Worrall et al. [12] proposed a novel CNN architecture to diagnose ROP plus disease, which is pre-trained GoogleNet with approximate Bayesian posterior over disease presence. In addition, authors train another CNN to return novel feature map visualizations of pathologies, learned directly from the data. Brown et al. [2] utilize two CNN to diagnose ROP plus disease, the first CNN is vessel segmentation network, which is used to segment retinal vessel through outputs a probability map whose size is same as input image and the value is between 0 and 1. The second CNN is classification network whose architecture adopts Inception_v1 architecture.

3. Data and Methodology

3.1 Data
Our data comes from Sichuan Provincial Peoples Hospital and Chengdu Women & Children’s Central Hospital, which contains 3350 ROP examinations from 2014 to 2017. Every examination consists of 4 to 12 retinal images, which reflects each premature infant’s fundus situation.

We adopt samples which have consistent label by different ophthalmologists to construct dataset, and discard samples do not have the same label. Noticeably, the dataset is per-image dataset rather than per-examination dataset. We manually contribute ROP samples to 2 groups, which are mild-ROP and severe-ROP respectively. Last, we divide all samples into three sets, training set, validation set and testing set respectively. Table 1 displays data distribution on every set.

Table 1. The number of samples of different datasets

|             | Normal | ROP  | Mild-ROP | Severe-ROP |
|-------------|--------|------|----------|------------|
| Training set| 6030   | 2477 | 361      | 814        |
| Validation set| 763    | 465  | 75       | 194        |
| Testing set  | 766    | 499  | 93       | 196        |
| **total**    | 7559   | 3441 | 529      | 1204       |

3.2 Classification Network

3.2.1 CNN architecture. Since 2014, Google proposed inception network architecture continually [7, 9, 10], which improved the best published result on ImageNet again and again. The successful secret of inception network is a module named “inception module” is proposed. Different from traditional convolutional neural network, Inception network contains a series of “inception modules”. The inventor of inception network extends the width of “inception module” with using different sizes of kernel to extract different spatial features, which is superior to traditional hierarchical convolutional network in performance. Figure 1(a) displays an example of inception module. He et al. [6] proposed a novel CNN architecture named “residual network”, which presents a shortcut connection. Residual network achieves the higher classification accuracy on ImageNet. The appearance of these architectures results in the great performance of CNN on image classification field. Figure 1(b) displays the detail of shortcut connection.
3.2.2 Median Frequency Balancing. Considering that dataset is imbalanced, for example the number of normal samples is much more than ROP samples. We use median frequency balancing on loss function to deal with such a problem. According to median frequency balancing, $\alpha_c$ denotes coordinate of class $c$ while training, which is formulated as:

$$\text{totalloss} = \sum_{1}^{n} \alpha_c \cdot \text{loss}(c) \quad \text{where} \quad \alpha_c = \frac{\text{medianfreq}}{\text{freq}(c)}$$

$\text{freq}(c)$ denotes the number of class $c$ divided set number, and $\text{medianfreq}$ is the median of all frequencies of classes.

3.2.3 Transfer Learning. The primary aim of “Transfer Learning” is proposed to save manual labeling costs with transferring model parameters from labeled dataset to unlabeled dataset. However, considering the difficulties to collect large scale data and the cost of labeling data, more researches gradually to focus on transferring model parameters on two domains, one domain comes from labeled large scale available data, and the other do not have enough data.

4. Experiments and Results

4.1 Normal and ROP

In the experiments of binary classification between normal and ROP. We have a try on pre-trained InceptionV2, InceptionV3 and ResNet-50 neural network. The Result were shown on Table 2. From table 2, we find ResNet-50 architecture demonstrates general best performance on normal and ROP binary classification experiment, and achieves highest accuracy and f1-score. InceptionV2 performs best precision rate on three neural networks. We think that InceptionV2 has the ability to capture more severe ROP disease feature, which is at the expense of reducing recall rate (lowest recall rate in them). However, InceptionV3 perform conversely. InceptionV3 achieves highest recall of 0.9158, but lowest precision rate of 0.8721. In the actual application scenario, doctors would prefer the system not to miss out on any of the ROP cases. Otherwise, it may cause irreversible hurt. However, doctors can tolerate several misdiagnose on negative cases. Through manual screening, they can pick up these false positive samples from positive samples. The process we think is less hard and faster than manual screening on whole of cases. Above of all, InceptionV3 is a good choice for the core of auto-recognition system.
Table 2. Performance of different networks on Normal and ROP binary classification.

| Networks          | Accuracy | Precision | Recall  | F1-Score |
|-------------------|----------|-----------|---------|----------|
| InceptionV2(pre-trained) | 0.9161   | 0.9356    | 0.8453  | 0.8882   |
| InceptionV3(pre-trained) | 0.9138   | 0.8721    | 0.9158  | 0.8935   |
| ResNet-50(pre-trained)  | 0.9272   | 0.8998    | 0.8999  | 0.8998   |

4.2 Mild-ROP and Severe-ROP

In the experiment of binary classification between mild-ROP and severe-ROP. Similarly, specific results were shown on table 3. From table 3, we find InceptionV3 achieves best performance among three neural networks on any of evaluation metrics. However, compared with Normal/ROP experiment, almost each evaluation metric is much lower. We think that two primary reasons causing such a result. One is the capacity of mild-ROP/severe-ROP dataset, which is smaller than normal/ROP dataset, and results in easy over-fitting on classification model. Alternatively, there is not enough explicit classification standard on different classes, only depends on the experience of experts.

Table 3. Performance of different networks on mild-ROP and severe-ROP binary classification

| Networks          | Accuracy | Precision | Recall  | F1-Score |
|-------------------|----------|-----------|---------|----------|
| InceptionV2(pre-trained) | 0.7326   | 0.7658    | 0.8718  | 0.8153   |
| InceptionV3(pre-trained) | 0.7847   | 0.7939    | 0.9235  | 0.8538   |
| ResNet-50(pre-trained)  | 0.7361   | 0.7913    | 0.8316  | 0.8109   |

5. Conclusion

In the future, we will try to classify the severity of ROP according to different stage and area, and detect position of lesion on premature infants’ retinal fundus images. We hope the appearance of the system can help reduce the burden of ophthalmologists and neonatologists for manual screening, and the imbalance of medical resources between developed and less developed regions.

Reference

[1] T. Aslam, B. Fleck, N. Patton, M. Trucco, and H. Azegrouz. Digital image analysis of plus disease in retinopathy of prematurity. Acta Ophthalmologica, 87(5):368–377, 2010. 1, 2
[2] J. M. Brown, J. P. Campbell, A. Beers, K. Chang, S. Ostmo, R. Chan, J. Dy, D. Erdogmus, S. Ioannidis, and J. Kalpathy Cramer. Automated diagnosis of plus disease in retinopathy of prematurity using deep convolutional neural networks. Jama Ophthalmol, 2018. 2
[3] W. M. Fierson. Screening examination of premature infants for retinopathy of prematurity. Pediatrics, 117(2):572, 2006. 1
[4] A. Garner. An international classification of retinopathy of prematurity. Ophthalmology, 92(8):987–994, 1985. 1
[5] Wallace D K, Quinn G E, Freedman S F, et al. Agreement among pediatric ophthalmologists in diagnosing plus and pre-plus disease in retinopathy of prematurity[J]. Journal of AAPOS, 2008, 12(4):352-356.
[6] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. CoRR, abs/1512.03385, 2015. 3
[7] S. Ioffe and C. Szegedy. Batch normalization: accelerating deep network training by reducing internal covariate shift. In International Conference on International Conference on Machine Learning, pages 448–456, 2015. 2
[8] P. K. Shah, V. Prabhu, S. S. Karandikar, R. Ranjan, V. Narendra, and N. Kalpana. Retinopathy of prematurity: Past, present and future. World Journal of Clinical Pediatrics, 5(1):35, 2016. 1
[9] C. Szegedy, S. Ioffe, and V. Vanhoucke. Inception-v4, inception-resnet and the impact of residual connections on learning. 2016. 2
[10] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. Rethinking the inception architecture for computer vision. In Computer Vision and Pattern Recognition, pages 2818–2826, 2016. 2.

[11] D. E. Worrall, C. M. Wilson, and G. J. Brostow. Automated Retinopathy of Prematurity Case Detection with Convolutional Neural Networks. Springer International Publishing, 2016. 2