Paper:

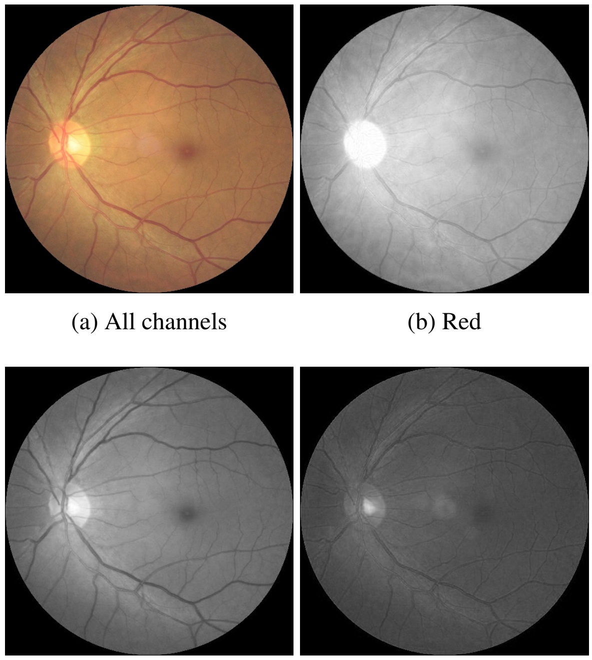
Exploring Effective Channels in Fundus Images for Convolutional Neural Networks

Shoya Kusunose, Yuki Shinomiya, and Yukinobu Hoshino

Kochi Unversity of Technology, 185 Miyanokuchi, Tosayamada, Kami City, Kochi 782-8502, JAPAN
E-mail: {245054x@gs., shinomiya.yuki@, hoshino.yukinobu@}kochi-tech.ac.jp

Abstract. This paper focuses on useful channel information for automated diagnosis using fundus images, especially for disease classification using convolutional neural networks (CNNs). In general, only the green channel of the image is used for the analysis of fundus images, and the green channel image is processed in various ways to analyze. The reason for this is that it is difficult to capture features such as blood vessels and optic nerve papillae from other channels. However, CNNs have the potential to acquire features that are difficult for humans to capture because they acquire features in the image by learning. We trained CNNs on fundus images for each combination of channels and compared their accuracy. As a result, we found that the appropriate channel varies depending on the disease, and the green channel is mainly accurate. In addition, the results of learning the appropriate ratio of channels using depth-wise convolution showed that green and red channels were enhanced, and from these results, we considered that the green and red channel are useful in the fundus classification.

Keywords: Convolutional Neural Networks, Fundus, Image Processing



(a) All channels (b) Red

(c) Green (d) Blue

Fig. 1. : Images of each channel in the fundus image. (b) is saturated around optic disks and (d) is overall dark, so these have low contrast. On the other hand, (c) has high contrast than the others.

1. Introduction

The fundus examination is an important diagnosis for detecting a variety of diseases. In addition to eye diseases such as glaucoma and cataract, the fundus examination can also detect abnormalities in the body including diabetes and hypertension. Manual fundus examinations by physicians use fundus images captured in three RGB channels, and diagnosis is based on capillaries and optic disks in the images. However, the contrast of the captured images is low and capillaries are difficult to see[1]. Manual diagnosis is also expensive to process a large number of images when many features are present in an image. In order to solve this problem, automatic diagnosis and annotation using computer-aided image analysis have been proposed[2–4]. In these methods, only the green channel of the fundus image is used to obtain an accurate contrast between the blood vessels and the background. Image processing of the fundus image has been better to use only green channel because the blue channel has little contrast and the red channel causes saturation[5, 6]. **Fig. 1** shows each image of the channel.

On the other hand, these methods are designed for limited tasks, and there are few methods to diagnose various diseases at once. In this study, we focused on the classification of fundus images by convolutional neural networks (CNN) in order to divide them into multiple diseases simultaneously. The CNN is a method specialized for image classification tasks, which learn feature extractions and classifications simultaneously. The CNN has many convolutional filters in the network model, and the combination of these filters provides high accuracy and multi-class classification[7].

In this paper, we focused on the channel information required for fundus classification by CNN. The green channel is widely used in the analysis of fundus images, and its usefulness in CNN is discussed. In addition, since CNN

Table 1. : Number of samples in each class. The ODIR include labels that is Normal (N), Diabetes (D), Glaucoma (G), Cataract (C), Age related Macular Degeneration (A), Hypertension (H), Pathological Myopia (P), Other diseases/abnormalities (O). ODIR dataset has imbalance of samples between each class. Normal class and Diabetes class have 1000 samples or more. On the other hand, other classes have fewer samples extremely than the former classes.

The number of samples per disease							
(N)	(D)	(G)	(C)	(A)	(H)	(P)	(O)
2873	1608	284	293	266	128	232	708

is an extension of image processing, channels with good accuracy in CNN have a possibility to promote further improvement in accuracy in other methods.

2. Proposal Approach

We propose the approach that explores the best channel combination of fundus image classification by CNN. We trained CNN models using single or multi channels information of fundus images as inputs, and compared their performance to find the most effective channels for fundus analysis using CNN. We explore the most effective channels for fundus analysis using CNN by comparing the performance of different combinations of channels. In the experiment section, we evaluate all combinations of channels not only a single channel considering the performance improvement by combining multiple channels. Finally, the channel combinations that explored were red, green, blue, red + green, green + blue, red + blue, and red + green + blue.

We also propose a model architecture that uses a depth-wise convolutional layer with 1×1 kernels behind the input layer to acquire the channel selection ratio by learning. The depth-wise convolutional layer is a layer with the same number of convolutional filters as the number of input channels of the layer, and each input channel is convoluted with a corresponded filter[9, 10]. The 1×1 convolutional filters without downsampling multiply each element without overlap[11], which is the same as scalar multiplication for channels. In other words, the depth-wise convolution of 1×1 kernels is a function of constant multiplication of each channel, and the multiplication factor can be learned simultaneously with the weights of other CNN.

We compared the model accuracy for different combinations of channels with the model accuracy trained on the proposed architecture. We also checked the weights acquired by the depth-wise convolutional layer in the model and discussed the useful channels.

Table 2. : The accuracy of the model trained by each selected channel images. The model that performed the best accuracy when only the green channel was used.

channel(s)	mean $\pm$ sd [%]
red	45.75 ± 2.33
green	$\mathbf{49.39 \pm 2.75}$
blue	42.82 ± 2.34
red + green	45.90 ± 1.64
green + blue	48.67 ± 1.55
red + blue	48.79 ± 2.27
red + green + blue	47.90 ± 1.62

3. Experiment

We performed 8-class classification of fundus images using CNN and compare these results. We used uniquely labeled images from the Ocular Disease Recognition (ODIR) dataset[8], which is a physician-labeled image dataset of 5000 patients. The ODIR include labels that is Normal (N), Diabetes (D), Glaucoma (G), Cataract (C), Age related Macular Degeneration (A), Hypertension (H), Pathological Myopia (P), Other diseases/abnormalities (O). **Table 1** shows numbers of samples in each class included the dataset.

This data was trained by the CNN model that are ResNet18 with pre-activations [12, 13]. The training epoch size is 100 epochs, and momentum is 0.9 on Nesterov momentum[14]. Learning rate declined from 0.1 with a cosine decay, and learning rate on the first 5 epochs increased to 0.1 from 0 to warm up the model. Input images of the model was resided to 224×224 pixels, and the pixel values were standardised. Furthermore, all trained models in the experiment were evaluated by 3 fold cross validation.

3.1. Exploration of Effective Channels

Models were trained with some channel combinations which are red, green, blue, red + green, green + blue, red + blue, and red + green + blue. **Table 2** shows the accuracy of the models trained on each combination of channels.

The accuracy of the model trained by only blue channel is the worst, and the one by only green channel is the best. For this 8-class classification, learning using only the green channel was effective in directly improving the accuracy. However, this comparison does not allow us to compare the accuracy for each class; ODIR is an imbalance dataset, and since this experiment is a general learning approach of CNN, the distribution of the recall for each class has a possibility to have imbalance.

Table 3 shows each class recall values of the models trained on each combination of channels. In the case of a single channel, class (N), (G), (H), (P), and (O) show the best recall when the input was the green channel. Then, class (C) and (A) show the best when the input was the red one, and class (D) had the best when the input was the

Table 3. : Each class recall value of the models trained on each combination of channels. Scores of bold font show the best value in each class. Almost of the best values are performed by models trained by images including the green channel.

channel(s)	mean [%]							
	(N)	(D)	(G)	(C)	(A)	(H)	(P)	(O)
red	64.74	32.96	10.56	**63.47**	21.07	3.89	69.80	14.83
green	**70.27**	31.9	23.57	58.70	16.92	4.65	74.50	17.51
blue	59.87	38.43	9.15	36.52	6.02	3.10	58.59	15.54
red + green	67.07	35.95	**24.99**	62.45	**21.42**	6.20	71.52	17.09
green + blue	68.01	**39.37**	21.49	50.50	18.05	**6.98**	**75.37**	**18.22**
red + blue	63.24	37.62	14.80	51.53	19.55	3.10	66.77	13.84
red + green + blue	66.66	37.00	20.08	54.25	17.30	3.88	69.34	17.51

blue one. In the case of mixed channels, the best recall was achieved when the green channel was mixed in for all classes.

The result of all cases led us to believe that the green channel was effective in most of the classes. On the other hand, the best recall was achieved when red was used in class (C). And in the class (A), the best recall was achieved when red + green was used, and the recall was higher when only red was used than when only green was used. From these results, the red channel is more useful than the green channel for class (C) and class (A).

Cataract is a disease in which the entire fundus image turns white. And age-related macular degeneration can be confirmed by the deformation of the black shadow in the center of the fundus image. These features are different from those of vessels and optic nerve papillae, and we considered that the features are not saturated in the red channel. Therefore, we considered the classification of cataract and age-related macular degeneration is more accurate than that using the green channel because there is no degradation in the red channel.

3.2. Exploration by depth-wise Convolution

We trained the model which is ResNet18-Preact inserted depth-wise convolutional layer after the input, and the accuracy is 47.26 ± 1.21. This accuracy is not much different from the results in the previous section, so the same level of learning should have taken place. The weights of the first depth-wise convolutional layer are shown in the **Table 4**. The weighting factors $w_r, w_g, w_b \in \mathbb{R}$ of the depth-wise convolution layer corresponds to each color channel to emphasize $|w| \geq 1$ or weakened $|w| < 1$ because the input was standardized. The weight of the green channel is larger than the other channels, indicating that the green was the most emphasized channel. Although the weight of the red channel is smaller than the one of the green, it was emphasized the input because it is more than one absolute value. We considered that the green channel is the most useful, however it can be combined with the red channel to obtain additional potential.

Table 4. : Weights of the first depth-wise convolutional layer. The weight applied to the green channel is the largest in the weights, and the weight enhance the green channel of input images of the model.

weights of channels		
red	green	blue
1.13	1.42	0.65

4. Conclusion

This paper focused on the useful channel information for disease classification using CNNs. We trained the CNN model for each combination of channels and compared their accuracy. We also used depth-wise convolution in the first layer of the model to learn the proportion of appropriate channels. As a result, we found that the appropriate channels were different for each disease, and we considered that the green and red channels were mainly useful.

References:

[1] Y. Hatanaka, T. Inoue, S. Okumura, C. Muramatsu and H. Fujita, "Automated microaneurysm detection method based on double-ring filter and feature analysis in retinal fundus images," Proc. of 2012 25th IEEE International Symposium on Computer-Based Medical Systems (CBMS), pp. 1–4, 2012

[2] D. Usher, M. Dumskyj, M. Himaga, T. H. Williamson, S. Nussey and J. Boyce, "Automated detection of diabetic retinopathy in digital retinal images: a tool for diabetic retinopathy screening," Diabetic Medicine, vol. 21, no. 1, pp. 84–90, 2004

[3] H. Nagayoshi, Y. Hamatsu, H. Sako, M. Himaga and S. KATO, "Detection of Fundus Lesions Using Classifier Selection," IEICE Transactions on Infor mation and Syste ms, vol. E92.D, no. 5, pp. 1168–1176, 2009

[4] T. Walter, P. Massin, A.Erginay R. Ordonez, C. Jeulin, and J. Klein, "Automatic detection of microaneurys ms in color fundus images," Medical Image Analysis, vol. 11, no. 6, pp. 555–566, 2007

[5] A. D. Hoover, V. Kouznetsova and M. Goldbaum, "Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response," IEEE Transactions on Medical Imaging, vol. 19, no. 3, pp. 203–210, 2000

[6] J. Staal, M. D. Abramoff, M. Niemeijer, M. A. Viergever and B. van Ginneken, "Ridge-based vessel segmentation in color images of the retina," IEEE Transactions on Medical Imaging, vol. 23, no. 4, pp. 501–509, 2004

[7] Y. Lecun, L. Bottou, Y. Bengio and P. Haffner, "Gradient-based

learning applied to document recognition," Proc. of the IEEE, vol. 86, no. 11, pp. 2278–2324, 1998

[8] "Ocular Disease Recognition — Kaggle," https://www.kaggle.com/andrewmvd/ocular-disease-recognition-odir5k, 14 Aug 2021

[9] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov and L. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4510–4520, 2018

[10] A. Howard, M. Sandler, G. Chu, L. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan, Q. V. Le and H. Adam, "Searching for MobileNetV3," Proc. of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 1314–1324, 2019

[11] M. Lin, Q. Chen and S. Yan, "Network In Network," Proc. of 2nd International Conference on Learning Representations, 2014

[12] K. He, X. Zhang, S. Ren, J. Sun, "Deep residual learning for image recognition," Proc. of the IEEE conference on computer vision and pattern recognition, pp. 770–778, 2016

[13] K. He, X. Zhang, S. Ren, J. Sun, "Identity mappings in deep residual networks," Proc. of European conference on computer vision, pp. 630–645, 2016

[14] Yu. E. Nesterov, "A method of solving a convex programming problem with convergence rate $O\left(\frac{1}{k^2}\right)$," Dokl. Akad. Nauk SSSR, vol. 269, pp. 543–547, 1983