Journal of Computing and Electronic Information Management
ISSN: 2413-1660 | Vol. 12, No. 2, 2024

Research on Retinal OCT Image Classification Based on Deep Learning

Mingxin Wang*

College of Electrical Engineering, Southwest Minzu University. Chengdu, Sichuan 610041, China
* **Corresponding author Email:** 943320673@qq.com

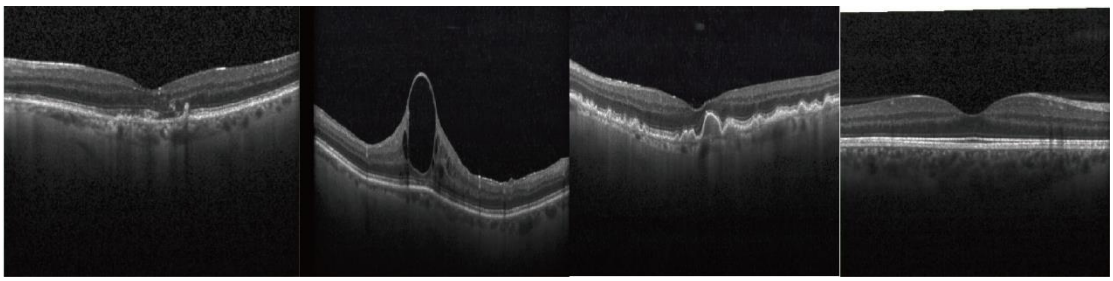
Abstract: Presently, there persists an issue in convolutional neural network (CNN)-based classification methods for retinal optical coherence tomography (OCT) images, particularly in discerning small-scale lesion areas. This dilemma results in neural network models exhibiting diminished accuracy when determining the activity of age-related macular degeneration (AMD) regarding its dry and wet stages, as well as choroidal neovascularization (CNV). Accurate identification of lesion types is paramount for ophthalmologists in devising treatment strategies. To address this challenge, a transfer learning-based EfficientNet retinal OCT image classification algorithm is proposed. Initially, retinal OCT images undergo data augmentation and preprocessing procedures. Subsequently, the pre-trained EfficientNet-B3 model is trained via transfer learning, followed by fine-tuning training using partial oversampling and class weighting techniques. The ultimate classification accuracy reaches 99.2%, signifying the model's commendable classification recognition accuracy. This underscores the clinical guidance significance of this research endeavor.

Keywords: EfficientNet; Retina; Optical coherence tomography; Image classification; Transfer learning.

1. Introduction

The structure of the fundus is complex and the types of lesions are diverse, such as choroidal neovascularization (CNV), diabetic macular edema (DME), drusen, age-related macular degeneration (AMD), and so on. Most of the fundus lesions are found in the macula, which is located in the center of the retina and is the most sensitive area of vision. The research of macular retinopathy has become an important topic [1]. Macular degeneration is one of the major causes of irreversible vision loss, and DME often causes irreversible visual impairment [2]. Optical coherence tomography (OCT) is a noninvasive, high-resolution optical imaging technique that produces cross-sectional images of objects in real time [3], and is widely used in the diagnosis of ocular diseases.

CNV lesion features are complex and varied, the lesion area boundary is blurred; DME lesion features are obvious and the lesion area is large; vitreous warts lesion area is small, the lesion boundary is blurred, the lesion features are not obvious; AMD features are complex, and the lesion features may appear differently in different lesion degrees, and the vitreous warts are one of the clinical manifestations of the early AMD.



a) CNV b) DME c) DRUSEN d) NORMAL

Fig. 1 Sample dataset

The intricate diagnosis of OCT image information requires significant manual processing and analysis by doctors, leading to extensive workload and time consumption. Deep learning models for automated analysis of OCT images can aid in streamlining the screening process for doctors. Leveraging the remarkable performance of deep learning in the field of image analysis, scholars have proposed numerous convolutional neural network-based models for classifying retinal OCT images, achieving recognition of lesions of varying quantities and types.

Fang et al. [4] investigated the recognition of three types of lesions, performing a four-class classification of CNV, DME, vitreous opacity, and normal fundus. However, the model proposed by the researchers struggled to extract features of small targets with small regions and small lesion textures, leading to poor recognition of vitreous opacity lesions with small lesion areas and ambiguous features. Ajesh et al. [5] utilized the DenseNet convolutional neural network to distinguish between AMD and normal cases. Sun et al. [6] employed a ResNet50 framework to classify retinal OCT images into AMD, DME, and normal categories. Karri et al. [7] applied fine-tuned GoogleNet using transfer learning for the classification task of DME and dry age-related macular degeneration OCT images.

Although the deep learning model algorithms used in recent years have been improved in general, the overall results are still not optimal. In this paper, a deep learning model with fine-tuned EfficientNet is used to classify and recognize

retinal OCT image data, in order to achieve the effect of accurately judging the classification of retinal OCT images.

2. Image Preprocessing

2.1. Dataset

In this paper, we use the OCT2017 retinal OCT image dataset downloaded from Kaggle, which consists of three folders: training set, validation set, and test set, and each subfolder contains NORMAL, CNV, DME, and DRUSEN retinal OCT images, totaling 84,495, and the distribution of the number of images in each folder is shown in Table 1.

Tab.1 OCT2017 dataset image distribution

	DME	CNV	DRUSEN	NORMAL
Training	11348	37205	8616	26315
Validating	8	8	8	8
Testing	242	242	242	242
total	11648	37505	8916	26515

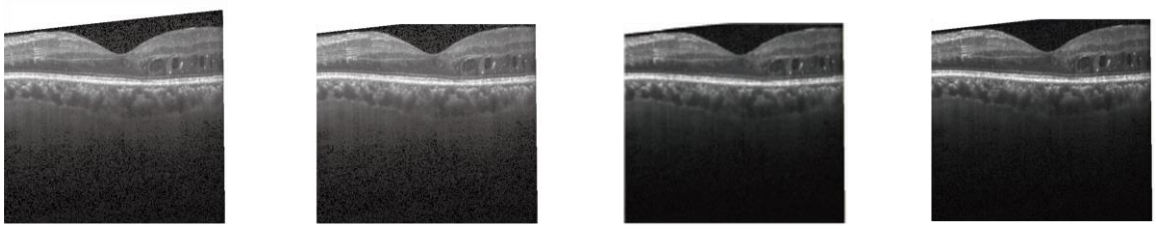
The dataset exhibits significant class imbalance, with a larger number of CNV and NORMAL labels compared to fewer instances of DME and DRUSEN labels. This unbalanced data set causes the model to over-fit the classes with a large number of categories during the training process, and then the accuracy of predicting the labels of DME or DRUSEN classes is not high during the testing process.

2.2. Image Preprocessing

Image preprocessing is an important part of the retinal classification method based on OCT images. OCT images usually produce image noise and image blurring during imaging, transmission, and storage, and these disturbing factors will directly affect the accuracy of the model classification task. In order to better perform the retinal OCT image classification task, it is necessary to pre-process the original image before inputting the network model. The specific steps are as follows.

Firstly, the operation involves removing the white borders from the images. Many redundant white borders are present in the original dataset due to improper handling during image acquisition, which can adversely affect subsequent image processing. Secondly, the resolution of the images is adjusted. Since the EfficientNet-B3 network model is most suitable for images with dimensions of (300, 300, 3), this paper adopts a bicubic interpolation algorithm to resize the images to dimensions where the shorter side is 300 pixels.

Finally, data normalization. All training data have to be normalized before they are input into the model, when we using these pre-trained models, we have to pay attention to normalize the data before putting our own data into the model, i.e., to normalize the data and distribute the data in the range of (0,1). The specific pre-processing results are shown in Figure 1.



a) Original image b) Removed white borde c) Resolution transformation d) Data normalization

Fig. 2 Image preprocessing example

2.3. Image enhancement

Deep learning models generally require a large amount of data for training, and image enhancement methods are performed on preprocessed images to expand the dataset. In this paper, we use random rotation, horizontal or vertical translation, random scaling, and random horizontal flipping to enhance the data. After enhancement, the original training set is divided into training set and validation set according to the ratio of 8:2. Considering that the number of validation set is small, the original validation set and test set are merged together as the test set.

3. Efficientnet

Based on past experience, increasing the depth of a network can yield richer, more complex features that can be effectively applied to other tasks. However, networks with excessive depth may face the issue of vanishing gradients, leading to training difficulties. Increasing the width of a network can result in obtaining finer-grained features and may also facilitate easier training. However, networks with large width and shallow depth often struggle to learn deeper features. Increasing the resolution of input images potentially allows for obtaining higher granularity feature patterns. However, for extremely high input resolutions, the gain in accuracy may diminish, and higher resolution images can also increase computational complexity.

Based on the considerations mentioned above, Tan et al. introduced a series of CNN-based backbone feature extraction networks called EfficientNet in 2019 [8]. It expands the network through a model compound scaling method, setting compound scaling coefficients to achieve a balance among resolution, width, and depth, thereby enhancing model performance.

The specific calculation of the composite scaling factor Φ is shown in Equation (1):

$$\begin{cases} Depth: d = \alpha^{\varphi} \\ Width: \omega = \beta^{\varphi} \\ Resolution: r = \gamma^{\varphi} \\ s.t. \alpha \cdot \beta^2 \cdot \gamma^2 \approx 2 \\ \alpha \geq 1, \beta \geq 1, \gamma \geq 1 \end{cases} \quad (1)$$

Where: d is the depth of the network, and its corresponding coefficient is α; w is the width, and its corresponding coefficient is β; r is the resolution, and its corresponding coefficient is γ; s.t. stands for the constraints.

Since the smallest image size in the retinal OCT image dataset is 384, in order to achieve good classification results, the Effi-cientNet-B3 model (optimal image size: 300) is selected as the backbone feature extraction network to realize the classification of retinal OCT images in this study.

Table 2 shows the network framework of EfficientNet-B3, it can be seen that the network is divided into 11 stages, the first stage is a 3×3 size convolution, including a Batch Normalization (BN) layer and the activation function Swish,

Stage2 to Stage8 are then batch stacked. MBConv structure, where #Layer indicates how many times the MBConv is re-stacked. Stage9 consists of the exact same structure as Stage1, and then outputs the result through global average pooling in Stage10 and fully connected layers in Stage11.

Tab. 2 EfficientNet-B3 architecture

Stage / i	Operator / $\widehat{F_i}$	Resolution / $\widehat{H_I} \times \widehat{W_I}$	#Channels/$\widehat{C_i}$	#Layers/$\widehat{L_i}$
1	Conv3×3+BN+Swish	300 x 300	3	1
2	MBConv1, k3×3	150 x 150	40	2
3	MBConv6, k3×3	150 x 150	144	3
4	MBConv6, k5×5	75 x 75	192	3
5	MBConv6, k3×3	38×38	288	5
6	MBConv6, k5×5	19 x 19	576	5
7	MBConv6, k5×5	19 x 19	816	6
8	MBConv6, k3×3	10×10	1392	2
9	Conv3×3+BN+Swish	10×10	1536	1
10	Gap	7×7	1536	1
11	FC+ Softmax	1×1	1536	1

3.1. MBConv structure

The MBConv is an advancement derived from the InvertedResidualBlock in the MobileNetV3 network. The key differences lie in the utilization of the Swish activation function and the addition of the SE (Squeeze-and-Excitation) module. Its structure is illustrated in Figure 3.

The MBConv structure primarily comprises a 1x1 ordinary convolution for dimensionality expansion, followed by a k×k Depthwise Conv convolution (including BN and Swish activation function). The specific value of k can be observed in the network framework of EfficientNet-B3, mainly encompassing 3x3 and 5x5 scenarios. Additionally, it includes an SE module and a 1x1 ordinary convolution for dimensionality reduction (including BN), along with a Dropout layer. The SE module, illustrated in Figure 4, consists of a global average pooling and two fully connected layers. The first fully connected layer has a number of nodes equal to one-fourth of the channels in the input MBConv feature matrix, utilizing the Swish activation function. The second fully connected layer's node count is equivalent to the number of channels in the input feature matrix of the depthwise separable convolution layer, employing the Sigmoid activation function.

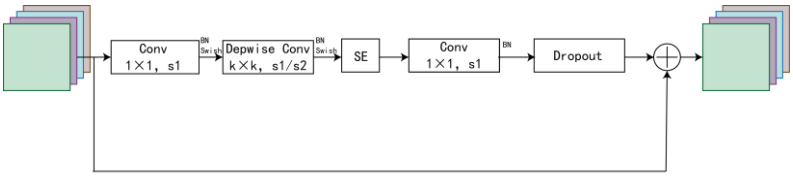


Fig.3 Diagram of the MBConv

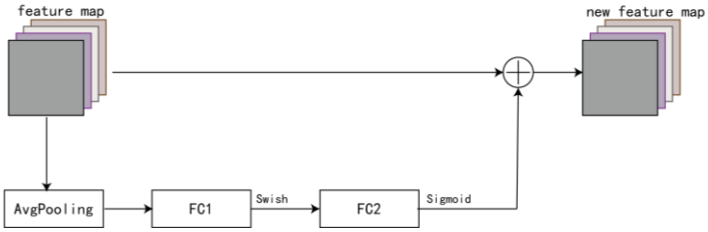


Fig.4 Diagram of the Squeeze-and-Excitation

4. Comparison of experimental results

The hardware configuration used in the experiment is: 64-bit Win10 operating system, Intel Core i7-10700KF CPU, 32G RAM, Nvidia GeForce RTX 3060 GPU, and the software mainly includes Anaconda and image processing libraries. In this paper, we use Accuracy and Confusion Matrix to evaluate the performance of the model for the image classification task. The specific expression formula is as follows (2).

$$Accurracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

In the equations, TP represents True Positives, FP represents False Positives, FN represents False Negatives, and TN represents True Negatives. These metrics are typically calculated using a confusion matrix. As shown in Table 3, each column of the confusion matrix represents the number of samples that are true for positive or negative categories, and each row represents the number of samples that are predicted to be positive or negative. The task of classifying OCT retinal diseases in this paper is a multiclassification task, in which a specific category is specified as the positive sample, for example, vitreous warts, at which time the negative samples are the remaining three categories other than vitreous warts. The number of TP, FP, FN, and TN can be obtained by the calculation of the confusion matrix, which leads to the calculation of the evaluation indexes mentioned above.

Tab.3 Confusion Matrix

True label / Predicted content	True positive class	True negative class
Predicted positive class	TP	FP
Predicted negative class	FN	TN

The model hyperparameters are configured with a batch

size of 16 and a learning rate of 0.001. Additionally, set the function to adjust the learning rate automatically. Adam optimizer and cross entropy loss function are employed. In order to compare the effect of different experimental methods on the classification readiness rate under the unified network structure, two sets of experiments are designed in this paper as follows.

Experiment 1 uses data augmentation, a class weighting approach was used to adjust the loss function, while transfer learning was performed using EfficientNet-B3 model weights pre-trained on the ImageNet dataset [9].

Experiment 2 employs oversampling method to over-sample the DME and DRUSEN by a factor of 2 respectively, using data Augmentation, class weighting techniques, and unfreezing the last convolutional and fully connected layers in the EfficientNet-B3 model, and updating the parameters of these two layers to retrain the model.

After 100 iterations of the above two experimental methods, the accuracy and LOSS values have basically converged, and the resultant accuracies on the training set and test set are shown in Table 4.

Tab. 4 Comparison of results after 100 iterations of two experiments

Method	Training Accuracy/%	Testing Accuracy/%
Experiment 1	95.8	95.1
Experiment 2	99.3	99.2

Through the experimental comparison, it is found that the results of Experiment 2 are better, which indicates that the classification accuracy of the method with partial oversampling and partial retraining of the model parameters is higher. In order to visualize the degree of matching between the real categories of the samples and the predicted categories, the confusion matrix is output according to the experimental results of the test set, and the specific results are shown in Fig. 5.

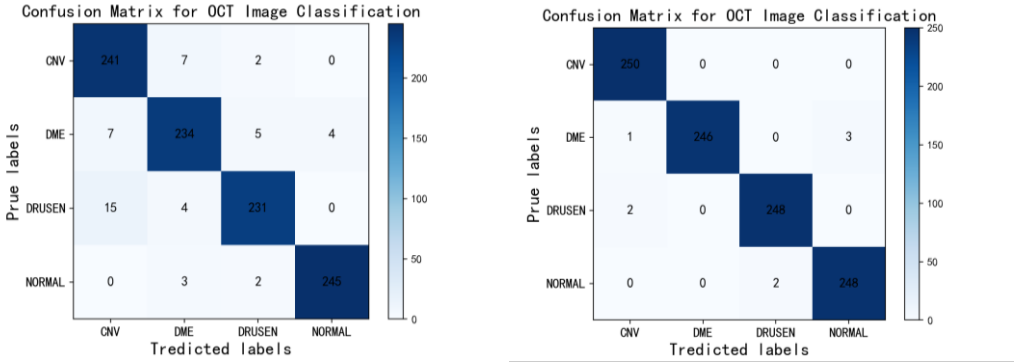


Fig. 5 Confusion matrix of experimental results

5. Conclusion

In this paper, we propose an EfficientNet classification model based on trasfer learning and perform fine-tuning training to make the classification of retinal OCT more robust and effective, which solves the issue of limited medical imaging datasets. Based on the classification results of the model, preliminary clinical diagnosis can be made, which greatly reduces the pressure of ophthalmologists in reading films and improves the diagnostic efficiency on the one hand, and on the other hand, the results of the model-based training are relatively objective, which excludes the subjective factors in the diagnosis to a certain extent and improves the accuracy of the diagnosis. Later, the model is validated by other datasets in order to adjust the parameters of the model to improve the generalization of the model, and at the same time, combined with the classification results of the corresponding retinal color fundus images, the multimodal comprehensive judgment is made in order to improve the accuracy of diagnosis.

Acknowledgments

This work was financially supported by the Southwest Minzu University Graduate Innovative Research Project No. (YB2023256) fund.

References

[1] FERRIS III F L, WILKINSON C P, BIRD A, et al. Clinical classification of age-related macular degeneration [J]. Ophthalmology, 2013, 120(4): 844-851.

[2] ZHANG Hao-rui, GUI Xiao, ZHAO Na, et al. Research progress on pharmacotherapy for diabetic macular edema [J]. Chinese Journal of Ophthalmology and Otorhinolaryngology, 2021, 21(3): 226-229.

[3] HUANG D, SWANSON E A, LIN C P, et al. Optical coherence tomography [J]. Science, 1991, 254(5035): 1178-1181,

[4] LIU X, BAI Y, CAO J, et al. Joint disease classification and lesion segmentation via one-stage attention-based convolutional neural network in OCT images [J]. Biomedical Signal Processing and Control, 2022, 71: 103087.

[5] Ajes h F, Abraha m A. Detection and classification of age-related macular degeneration using integration of densenet169andconvolutional neural network[C]. Innovations in Bio-Inspired Computing and Applications: Proceedings of the 12th International Conference on Innovations in Bio-Inspired Computing Proceedings of the 12th International Conference on Innovations in Bio-Inspired Computing and Applications (IBICA 2021). Cham: Springer International Publishing, 2022: 226-238.

[6] Sun Y, Zhang H, Yao X. Automatic diagnosis of macular diseases from OCT volume based on its two-dimensional feature map and convolutional neural network with attention mechanism[J]. Journal of Biomedical Optics, 2020, 25(9): 096004-096004.

[7] Karri S P K, Chakraborty D, Chatterjee J. Transfer learning based classification of optical coherence tomography images with diabetic macular edema and dry age-related macular degeneration[J]. Biomedical Optics Express, 2017, 8(2): 579-592.

[8] TAN M, LE Q V. Efficientnet: rethinking model scaling for convolutional neural networks [C]. International Conference on Machine Learning, 2019:6105-6114.

[9] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C]. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.