긴꼬리 분포의 광간섭 단층촬영 데이터세트에 대한 다중 레이블 이미지 분류

Phuoc-Nguyen Bui[1], 정경희[1], Duc-Tai Le[2], 추현승[3]
[1]성균관대학교 수퍼인텔리전스학과
[2]성균관대학교 컴퓨터정보학과
[3]성균관대학교 전기컴퓨터공학부
email: {phuocnguyen, datakira, ldtai, choo}@skku.edu

Multi-Label Image Classification on Long-tailed Optical Coherence Tomography Dataset

Phuoc-Nguyen Bui[1], Kyunghee Jung[1], Duc-Tai Le[2], Hyunseung Choo[3]
[1]Dept of Superintelligence, Sungkyunkwan University
[2]College of Computing and Informatics, Sungkyunkwan University
[3]Dept of Electrical and Computer Engineering, Sungkyunkwan University

Abstract

In recent years, retinal disorders have become a serious health concern. Retinal disorders develop slowly and without obvious signs. To avoid vision deterioration, early detection and treatment are critical. Optical coherence tomography (OCT) is a non-invasive and non-contact medical imaging technique used to acquire informative and high-resolution image of retinal area and underlying layers. Disease signs are difficult to detect because OCT images have many areas which are not related to any disease. In this paper, we present a deep learning-based method to perform multi-label classification on a long-tailed OCT dataset. Our method first extracts the region of interest and then performs the classification task. We achieve 98% accuracy, 92% sensitivity, and 99% specificity on our private OCT dataset. Using the heatmap generated from trained convolutional neural network, our method is more robust and explainable than previous approaches because it focuses on areas that contain disease signs.

1. Introduction

Retinal disorders have become a global problem with at least 2.2 billion people have a near or distance vision impairment. To prevent the vision loss, it is important to detect and monitor the development of retinal diseases. Since the development of optical coherence tomography (OCT) in 1991, it has become the most commonly used imaging modality in ophthalmology with 4.39 million, 4.93 million, and 5.35 million OCTs performed in 2012, 2013, and 2014 respectively in the US Medicare population. A high-resolution OCT image contains information about retinal area and underlying layers. Therefore, it is widely used in clinic center to diagnosis eyes disorders.

Deep learning (DL) has dominated the field of computer vision since the introduction of AlexNet [1] in 2012. With a strong ability to extract high-level features automatically, DL-based methods show the stability under high variation of the medical data. The applications of DL in OCT image classification have been explored in [2-4]. These studies only focus on binary or multi-class classification task, which means that the DL models produce one class for an OCT image. However, an OCT image may contain signs of multiple retinal diseases, in which previous approaches cannot be applied. Therefore, multi-label classification on OCT dataset remains a practical and challenging task.

In medical field, acquiring high classification

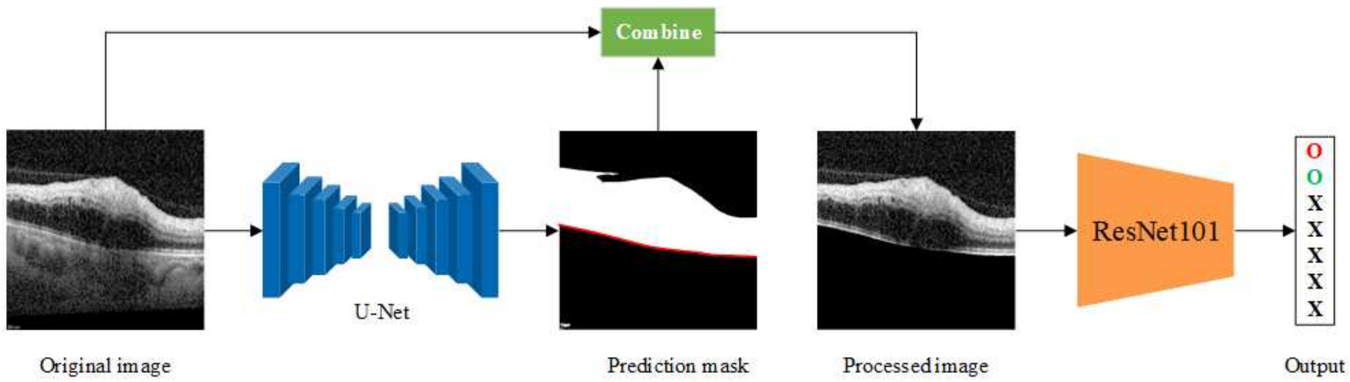

Figure 1. Details of the proposed method. Each color of O symbol in output represents one class.

performance (i.e., accuracy, specificity, and sensitivity) may be insufficient. A reliable DL-based computer-aided diagnosis system in clinic centers must reach to a certain level of robustness with explainable decisions. Particularly, in OCT image classification, such a DL-based system produces its output based on related areas, which contain disease signs, instead of irrelevant regions. However, a high-resolution OCT image contains information about retinal area and underlying layers, which may confuse DL model. Subsequently, an explainable DL-based classification solution for OCT image remains a challenging problem.

In this paper, we present a deep learning-based method for multi-label image classification on a long-tailed OCT dataset. The main contributions are described as follows:

1. We propose a simple and effective method that removes irrelevant areas and extract regions of interest in OCT image for training DL model.
2. The proposed method not only achieves high classification performance in terms of accuracy, specificity, and sensitivity but also produces explainable predictions.

2. Method

2.1. Region of interest extraction module

The structure of a standard OCT image consists of two parts: the upper part is the retinal area and the bottom one contains the underlying layers. Diseases signs of retinal disorders only appear in the upper part. We observe that training a convolutional neural network (CNN) with the original OCT image may lead to the mismatch between CNN's output and heatmap. To address this problem, a segmentation model is utilized to separated the upper and the bottom parts. As described in Figure 1, prediction mask generated from U-Net [5] is used to extract the region of interest (RoI) by combining with the original OCT image. Particularly, the detected boundary between two parts of an OCT image is used to remove the irrelevant area.

2.2. Classification module

As mentioned previously, the applications of DL in OCT image classification limit in binary or multi-class tasks. However, an OCT image contains disease signs from multiple retinal disorders. For example, an OCT image in our dataset can be classified into seven classes, in which each image can be labeled with more than one class. Particularly, these seven classes are age-related macular degeneration (AMD), epiretinal membrane (ERM), macular edema (ME), macular hole (MH), myopic foveoschisis (MF), retinal detachment (RD), and Normal (NM).

As illustrated in Figure 1, we use a variant of ResNet [6] architecture (i.e., ResNet-101) as CNN backbone to perform the multi-label classification task. Different from a plain network, mainly based on the philosophy of VGG nets [7], ResNet inserts a shortcut connection, which leads to better performance compared with previous networks. We use the multi-label loss as objective function to train ResNet-101.

3. Experiments

3.1 Dataset and implementation details

In this work, we use a private dataset collected from a hospital. Personal information of the patient is removed before any studies are conducted. This dataset includes 18,433 OCT images in seven classes as described in Table 1. We implement the proposed method based on the PyTorch framework with the pre-trained model on the ImageNet dataset.

Class	# of images
AMD	3940
ERM	3581
ME	2756
MH	19
MF	103
RD	34
NM	8000

Table 1. Details of the OCT dataset

3.2. Performance evaluation

We summarize the experimental results in Table 2. ResNet-101 is trained with original OCT images (Before RoI extraction) and processed OCT ones (After RoI extraction). The proposed method improves Sensitivity from 0.91 to 0.92 and Accuracy from 0.97 to 0.98 while maintaining Specificity at 0.99.

Class	Before RoI extraction			After RoI extraction		
	Acc	Sen	Spe	Acc	Sen	Spe
AMD	0.98	0.97	0.98	0.98	0.96	0.99
ERM	0.93	0.84	0.97	0.94	0.90	0.96
ME	0.96	0.90	0.98	0.96	0.94	0.97
MH	1.00	0.89	1.00	1.00	0.89	1.00
MF	1.00	0.87	1.00	1.00	0.97	1.00
RD	1.00	0.91	1.00	1.00	0.78	1.00
NM	1.00	1.00	1.00	1.00	1.00	1.00
Mean	0.97	0.91	0.99	0.98	0.92	0.99

Table 2. Experimental results on OCT dataset.

To study the robustness of the proposed method, we generate heatmap or class activation map of trained ResNet-101. As shown in Figure 3, heatmap of ResNet-101 trained with original OCT images shows that the output is inconsistent. In other words, the focused area of the heatmap is matched with input. In terms of ResNet-101 trained with processed input, generated heatmaps focus on areas where diseases signs appear in both cases.

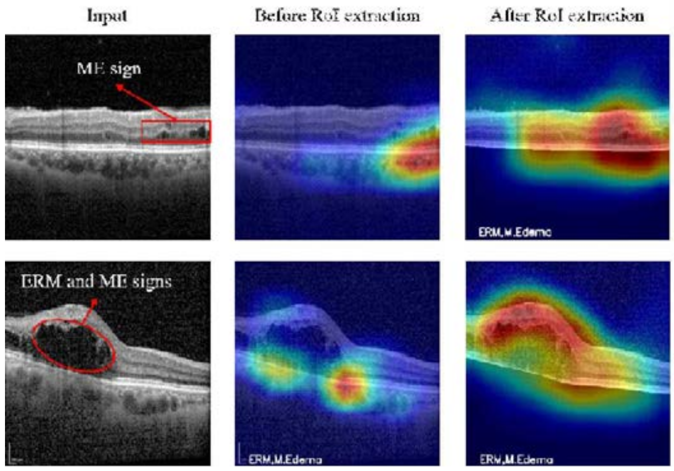

Figure 2. Heatmaps of ResNet-101 for two cases of input: original and processed.

References

[1] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Communications of the ACM 60.6, pp. 84-90, 2017.

[2] Wang, Depeng, and Liejun Wang. "On OCT image classification via deep learning." IEEE Photonics Journal 11.5, pp. 1-14, 2019.

[3] Awais, Muhammad, et al. "Classification of sd-oct images using a deep learning approach." IEEE International Conference on Signal and Image Processing Applications (ICSIPA), 2017.

[4] Wang, Jing, et al. "Deep learning for quality assessment of retinal OCT images." Biomedical Optics Express 10.12, pp. 6057-6072, 2019.

[5] Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." International Conference on Medical Image Computing and Computer-Assisted Intervention, 2015.

[6] He, Kaiming, et al. "Deep residual learning." Image Recognition 7, 2015.

[7] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint:1409.1556, 2014.