Long-Tailed Multi-label Retinal Diseases Recognition via Relational Learning and Knowledge Distillation

Qian Zhou, Hua Zou*, Zhongyuan Wang
Wuhan University, Wuhan, China
zouhua@whu.edu.cn

1. INTRODUCTION

Over the past decades, ocular diseases have become one of the main causes of blindness. It is of great importance to detect and treat ocular diseases timely. However, one patient may suffer from more than one retinal disease and these diseases follow a long-tailed distribution, making accurate diagnosis difficult. Here, we propose a new framework that utilizes the relations among diseases for long-tailed multi-label retinal diseases classification. The relations are considered from three main aspects (i.e., multi-task relation, feature relation, and pathological region relation).

2. METHOD

2.1 Framework

The proposed framework contains two stages, named multi-task pretraining and multi-label fine-tuning. When pre-training, we take the DR lesion segmentation and DR severity grading as the downstream tasks and design two branches respectively. The two parallel branches share the same bottom network, followed by another subnet, i.e., 'Subnet-S' for segmentation and 'Subnet-C' for grading.

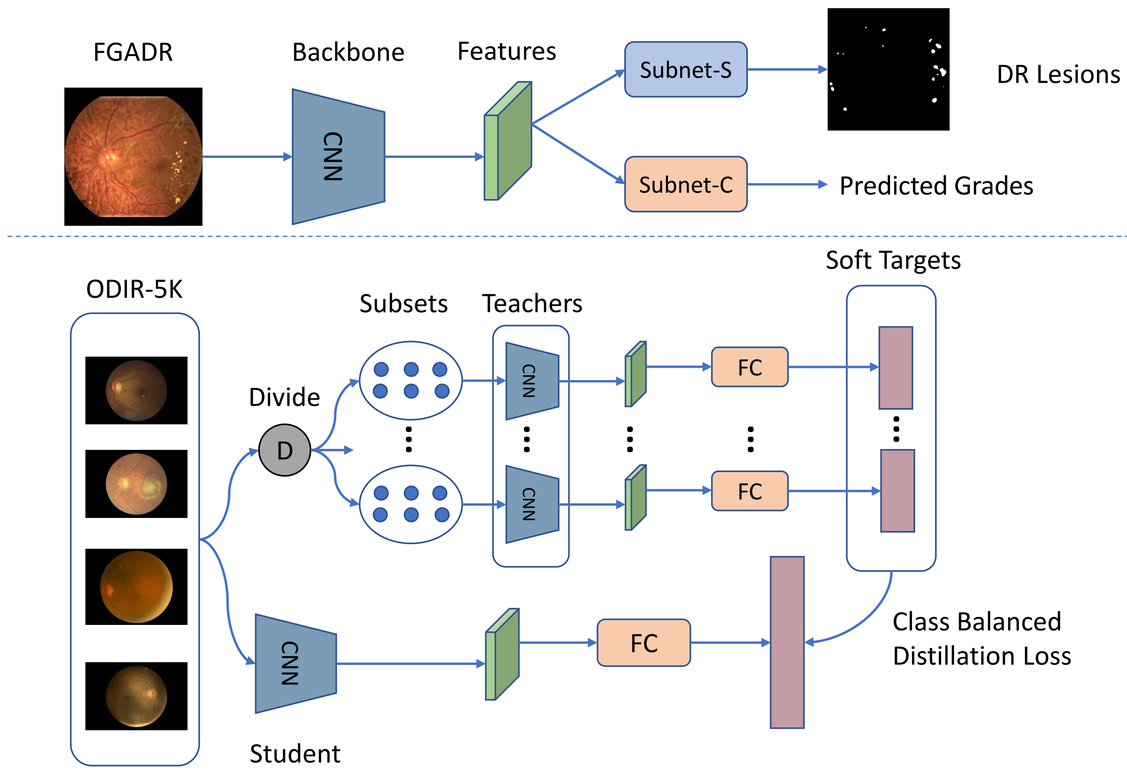

Figure1. Pipeline of the proposed framework. Above the line is the multi-task pre-trained model, below the line is the target model for retinal diseases recognition.

2.2 Multi-task Pre-training

We take the DR segmentation and severity grading as the pre-training tasks for three main reasons.

1. We can obtain enough well-labeled public datasets about DR.
2. Many ocular diseases are related to hyperglycemia, which is the main cause of DR.
3. The multitask pre-training can exploit the correlations to transfer prior knowledge from DR to multi-label retinal diseases recognition.

2.2 Region-based Attention

To improve the localization ability of important regions, the trainable convolutional layer is combined with the non-parameter pooling layer.

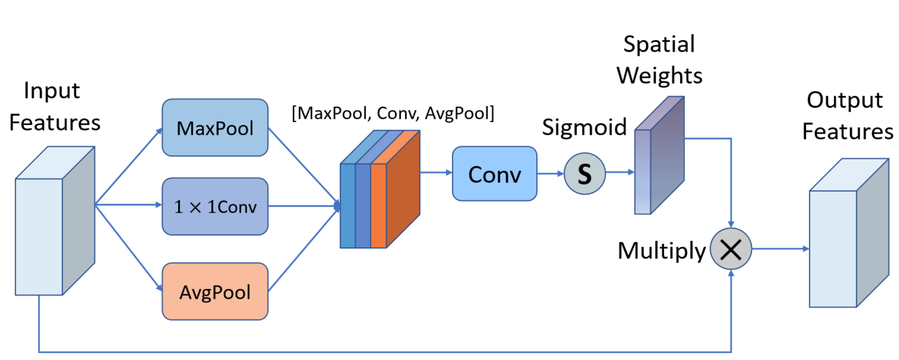

Figure 2. The overview of proposed region-based attention module. As illustrated, we concatenate both pooling outputs and convolution outputs and forward them to a convolution layer with sigmoid activation function to generate spatial weights.

2.3 Relational Subsets Generation

To reduce the multi-label samples and alleviate the improve the class imbalance, the original long-tailed dataset is divided into several relational subsets according to the semantic relations among diseases.

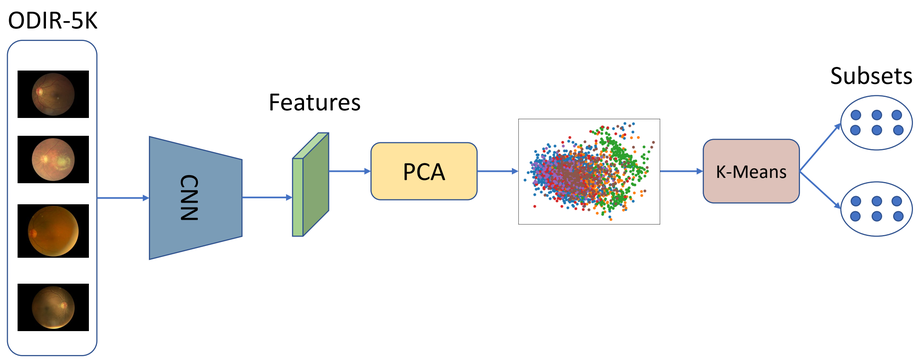

Figure 3. Diagram of relational subsets generation. We only perform the pipeline on six labels (i.e., D, G, C, AMD, H, M).

2.4 Class-Balanced Knowledge Distillation

To distill the teacher models into a uniform student model, a class-balanced distillation loss is proposed to avoid multi-label overfitting. w_i^j is a class-balanced weight for each sample.

$$L_{KD} = \sum_{j}^{C} \sum_{i}^{N_i} KL\left(\hat{q}_i^j \parallel q_i^j\right) \cdot w_i^j = \sum_{j}^{C} \sum_{i}^{N_i} \hat{q}_i^j \log \frac{\hat{q}_i^j}{q_i^j} \cdot w_i^j$$

3. EVALUATION

3.1 Quantitative Performance

The proposed method is compared with other approaches on ODIR-5K dataset.

Backbone	Methods	Kappa	F1	AUC
ResNet-50 [15]	SCFKD [5]	0.635 ± 0.009	0.911 ± 0.007	0.927 ± 0.014
	IB Sampling	0.553 ± 0.010	0.974 ± 0.009	0.887 ± 0.012
	CB Sampling	0.601 ± 0.013	0.886 ± 0.011	0.920 ± 0.015
	Focal Loss [17]	0.625 ± 0.011	0.895 ± 0.009	0.930 ± 0.012
	RSKD [16]	0.660 ± 0.014	0.920 ± 0.013	0.935 ± 0.014
	DB Loss [18]	0.673 ± 0.008	0.930 ± 0.007	0.940 ± 0.009
	Ours	$\mathbf{0.712 \pm 0.011}$	$\mathbf{0.935 \pm 0.013}$	$\mathbf{0.944 \pm 0.012}$
DenseNet-121 [21]	CCT-Net [11]	0.749 ± 0.007	0.952 ± 0.011	0.960 ± 0.013
	Ours	0.744 ± 0.010	0.960 ± 0.013	0.964 ± 0.010
	Ours + Focal Loss [17]	$\mathbf{0.767 \pm 0.012}$	$\mathbf{0.965 \pm 0.008}$	$\mathbf{0.970 \pm 0.014}$

3.2 Quantitative Performance

Grad-CAM is used to visualize the results to perform interpretability analysis. With the proposed method, model can pay more attention to the lesions in the images to improve the accuracy.

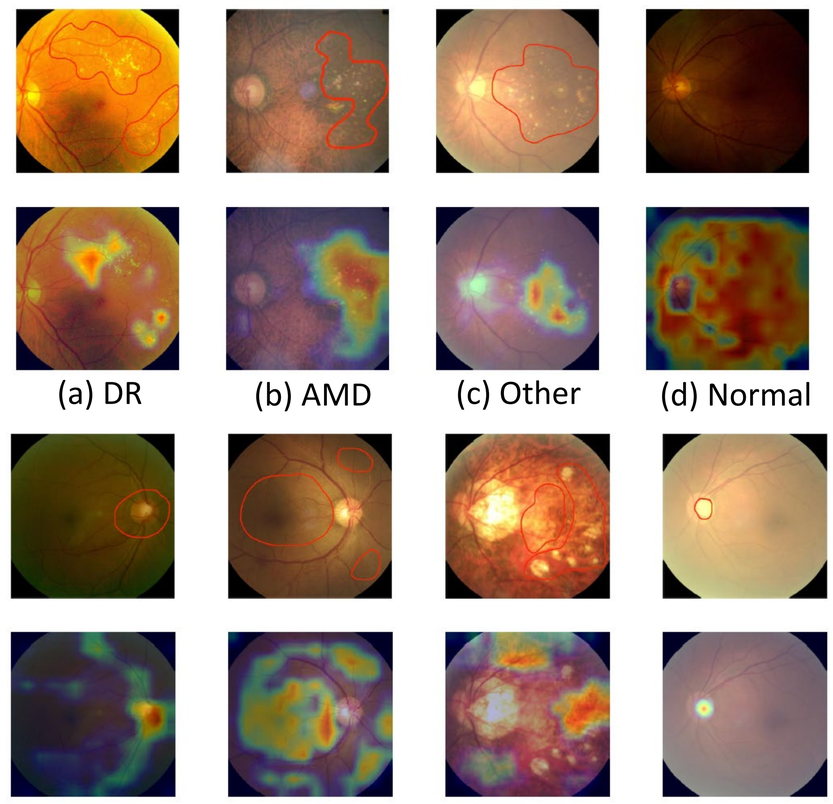

Figure 4. Class activation maps (CAM) of different diseases.

4. CONCLUSION

The presented new framework leverages the relations among diseases from three main aspects. Especially, we propose a simple but effective algorithm to divide the long-tailed dataset into relational subsets. This algorithm can be easily expanded into other fields without strong prior knowledge.

5. Acknowledgements

This work is supported by Science and Technology Cooperation Project of The Xinjiang Production and Construction Corps (No. 2019BC008, 2017DB004).