International Core Journal of Engineering Volume 11 Issue 6, 2025
ISSN: 2414-1895 DOI: 10.6919/ICJE.202506_11(6).0010

Pathological Myopia Diagnosis based on ResNet SR-init Structured Pruning and Self-learning Distillation

Huishuang Jia[a], Jiaze Li[b], Yijiao Sun[c], Jianing Qiu[d], Yuntian Du[e], and Xuebin Chen[f,*]

North China University of Science and Technology, Tangshan 063210, China

[a]13473036217@163.com, [b]1412894870@qq.com, [c]878723312@qq.com, [d]3452706530@qq.com, [e]dyt2585526816@163.com, [f,*]chxb@qq.com

Abstract

This paper proposes a ResNet-based SR-init structured pruning and self-learning distillation system for the diagnosis of pathological myopia. By comparing the performance of VGGNet, GoogLeNet, and ResNet-50 on the iChallenge-PM fundus image dataset, the ResNet-50 model with the best performance was selected based on key indicators such as accuracy and loss value. The SR-init structured pruning technique is then employed to prune filters with minimal performance impact, thereby reducing the model's complexity and computational requirements without significantly affecting its performance, and finally, self-distillation technology is used to further optimize the model. In this process, the model uses its own predicted outputs on the training data as new training labels. These outputs serve as soft labels to retrain the model using the cross-entropy loss function, which aids the model in deeply understanding the complex characteristics of pathological myopia. To prevent overfitting and ensure optimal performance on unseen data, an early stopping strategy was introduced. The final model results performed well, significantly improving the automatic diagnosis efficiency of pathological myopia and reducing the misdiagnosis rate. This study is applicable to resource-constrained medical environments and is of great significance for promoting intelligence and automation in the medical industry.

Keywords

ResNet-50; SR-init structured Pruning; Self-learning Distillation; System for the Diagnosis of Pathological Myopia; Early Stopping Strategy.

1. Introduction

As a serious ophthalmic disease, pathological myopia has seen a sharp rise in prevalence worldwide with the progress of globalization and the increase in screen time. This trend is particularly evident among adolescents, which not only affects their learning and quality of life, but may also lead to serious retinal complications in the long run, such as macular degeneration and retinal detachment, which in turn increases the risk of blindness in patients. According to the World Health Organization, pathological myopia has become one of the main causes of blindness, posing a major challenge to the public health system.

The traditional diagnosis of pathological myopia relies on complex instrumental examinations and the clinical experience of doctors. The diagnostic process is not only time-consuming and labor-intensive, but the results are limited by the professional level and experience of doctors, and there is a certain degree of subjectivity. In addition, for resource-constrained areas, there is a lack of advanced medical equipment and a shortage of professional doctors. These factors have seriously hindered the timely and accurate diagnosis of pathological myopia.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202506_11(6).0010

With the development of artificial intelligence technology, especially the widespread application of deep learning in the field of medical image processing, the automatic diagnosis of pathological myopia using deep learning models has become an important means to improve diagnostic efficiency and accuracy. Deep learning technology can identify the characteristic signs of pathological myopia by learning a large amount of medical image data, thereby automatically analyzing and diagnosing the disease. However, existing deep learning models usually require a lot of computing resources and have high hardware requirements, which to some extent limits their application in actual clinical settings.

2. Literature Review

Pathological myopia is a complex eye disease that not only involves extreme myopia, but is also accompanied by potential damage to the retina and other key eye structures, posing a serious threat to the patient's vision. Traditional diagnostic methods are highly dependent on intuitive evaluation by ophthalmologists, a process that is not only time-consuming and labor-intensive, but also inefficient. In view of this, in order to improve the speed and accuracy of diagnosis of pathological myopia and thereby achieve more efficient and accurate medical services, the use of deep learning technology for automatic diagnosis of pathological myopia has become a research hotspot in the international academic community today.

2.1 Research on Deep Model Architecture Optimization.

Li Yongan's team [1] proposed the Rep VGG-DAPPM architecture, which improved the feature extraction efficiency based on VGG through structural reparameterization technology; Wan Cheng's team [2] used the ResNeXt-50 network to implement multi-branch residual learning, and its lightweight design significantly reduced the parameter scale; Yoo's team [5] innovatively combined ResNet50 with OCT image features and achieved a prediction accuracy of 0.92 in the field of optical coherence tomography.

The self-distillation framework proposed by Zhang et al. [7] enhances the model representation capability through internal knowledge transfer; Zhou et al. [12] used evolutionary algorithms to optimize the network initialization strategy so that the pruned network maintains 91.3% of the original accuracy; the structural pruning scheme developed by Fang team [13] for the diffusion model can reduce the computational effort by 30%.

2.2 Model Lightweighting and Efficiency Optimization

Zhang Yu's team [3] innovatively used an expected value pruning strategy to accurately identify redundant convolution kernels, reducing the amount of computation by 50% while retaining 98% accuracy; Ismail et al. [10] optimized layer pruning through a genetic algorithm, enabling MobileNet-V2 to achieve 98.48% accuracy in COVID-19 detection; Chen Yahao's team [3] used Laplacian filtering enhancement to reduce network depth requirements.

Preprocessing optimization technology Namra team [4] integrated multi-scale filtering preprocessing into the CNN model, reducing the verification loss to 0.1457; the Blend AutoAugment method developed by Im et al. [8] improved the classification accuracy by 3.2% through linear blending enhancement; the Laplacian feature enhancement scheme of Chen Yahao team [3] significantly improved the ResNet fundus feature extraction capability.

The knowledge distillation system of Alkhulaifi's team [6] compressed the model size by 70% with only a 2.1% loss in accuracy through a teacher-student framework; the benchmark platform built by Xia et al. [9] supports cross-model transfer learning and achieves an average transfer efficiency of 0.93 on the fundus dataset.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202506_11(6).0010

2.3 Research on Cross-Modality and Generalization Ability

Ismail's team[10] established a multimodal diagnostic system by integrating CT and ECG data, with an accuracy rate of 99.65%. The Optos image analysis system developed by Shi et al.[4] achieved 85% cross-device generalization capability.

The large-scale fundus benchmark dataset constructed by Xia et al. [9] contains 28 types of lesions, promoting cross-disease domain adaptation research; the evolutionary pruning scheme of Zhou team [12] maintains a stable accuracy of over 90% on 5 medical imaging datasets.

Through a systematic literature review, this study found that the current research focus shows three major trends: model architecture design evolves from a single scale to multi-granularity feature fusion; model optimization pays more attention to the balance between interpretability and resource efficiency; cross-modal learning becomes the key path to break through the bottleneck of generalization in the field. However, as pointed out by various literatures, the risk of overfitting based on specific data sets [1,3,4,10], real-time constraints [3,5], and heterogeneous data compatibility [9,12] are still the core challenges that need to be solved.

3. Data Source and Processing

3.1 Data Source

This study uses the ichallenge-PM dataset, which is a medical dataset about pathologic myopia (PM) provided in the iChallenge competition jointly organized by Baidu Brain and Zhongshan Eye Center of Sun Yat-sen University. The dataset contains 400 training set images and 400 validation set images.

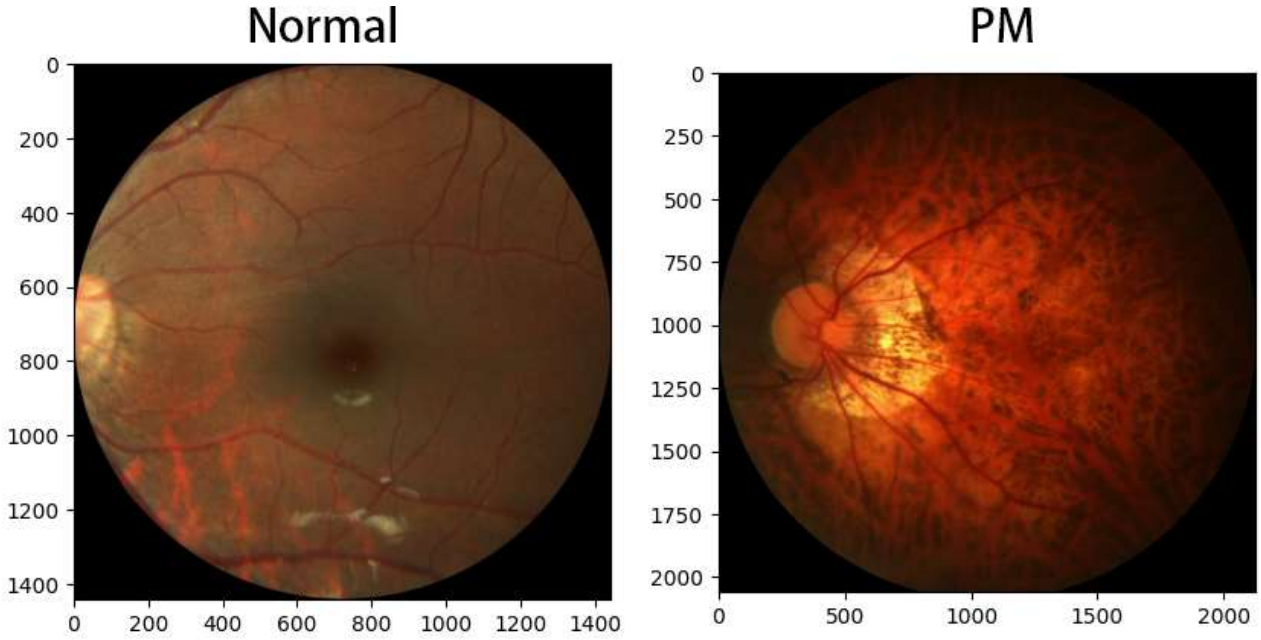

Figure 1. Data display

3.2 Data Preprocessing

First, the AutoAugment algorithm is used to search for an image augmentation scheme suitable for a specific dataset in the search space of a series of image augmentation sub-strategies. The dataset is expanded to help the model learn the ability to observe lesions from different angles, thereby improving its generalization.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202506_11(6).0010

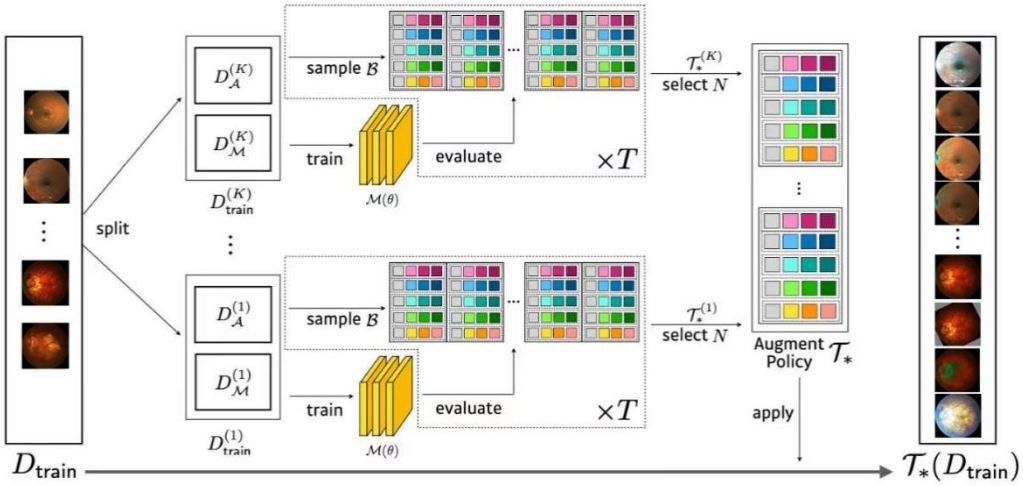

Figure 2. AutoAugment Algorithm Process

Then, the same size is applied to the images for normalization, and all are resized to 224x224. The scale value, mean, variance, and order of all images are normalized.

4. Model Selection

This paper selects three widely recognized deep learning models: VGGNet-16, GoogLeNet, and ResNet-50, to compare performance on the iChallenge-PM fundus image dataset. These models each have their own unique structure and advantages, suitable for different image processing needs. VGGNet-16 is known for its simple and uniform network structure, which is very suitable for image feature extraction and classification tasks. GoogLeNet uses an innovative Inception module, which effectively captures the multi-scale features of the image by processing convolution kernels of different sizes in parallel, thereby improving the recognition ability of the model. ResNet-50 solves the gradient vanishing problem in deep networks by introducing a residual learning framework, allowing the network to maintain stable performance while increasing depth.

To ensure the fairness and scientificity of the experiment, all models are trained under a unified parameter configuration. The learning rate of the model is uniformly set to 0.001 to ensure that all models are optimized at the same learning rate. The batch size is set to 64, which is the result of a balance between efficiency and resource consumption. Adam is selected as the optimizer because its adaptive learning rate adjustment mechanism is very suitable for deep learning tasks. At the same time, L2 regularization is used to reduce overfitting.

Finally, in order to prevent overtraining and optimize the use of computing resources, this paper adopts an early stopping strategy, with the accuracy of the model on the validation set as the monitoring indicator. Setting the patience parameter to 5 means that if the improvement in accuracy is less than 0.1% for 5 consecutive training cycles, the early stopping mechanism will be triggered. Once the early stopping is triggered, the model state will be restored to the optimal state of the monitoring indicator.

This paper conducts preliminary training on the model, statistically calculates the accuracy and loss indicators of the model, and obtains the final indicator values of the three models based on the early stopping strategy, completing the preliminary performance comparison of the three models:

ISSN: 2414-1895 DOI: 10.6919/ICJE.202506_11(6).0010

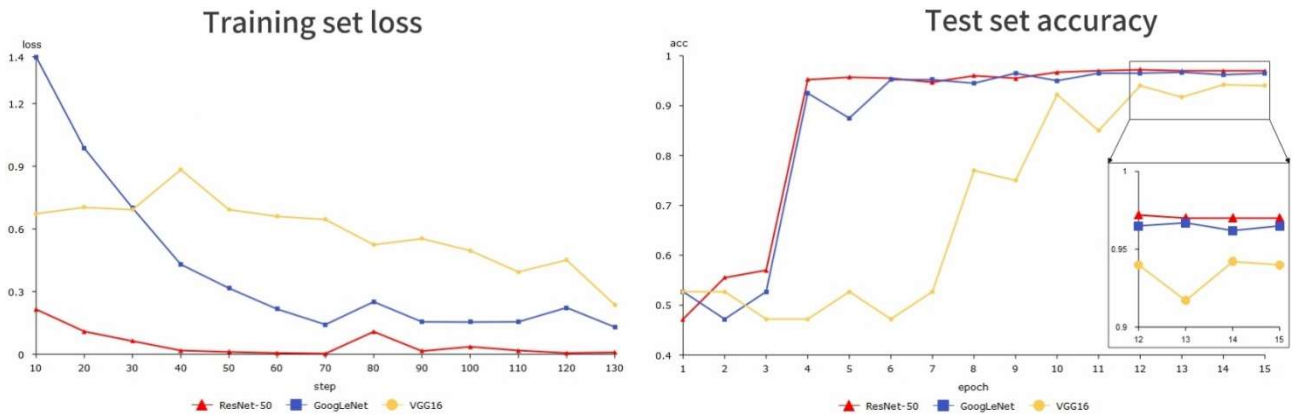

Figure 3. Comparison of three models

Finally, the ResNet-50 model with better performance was selected as the baseline model for subsequent experiments. The basic structure of this deep learning model is shown below:

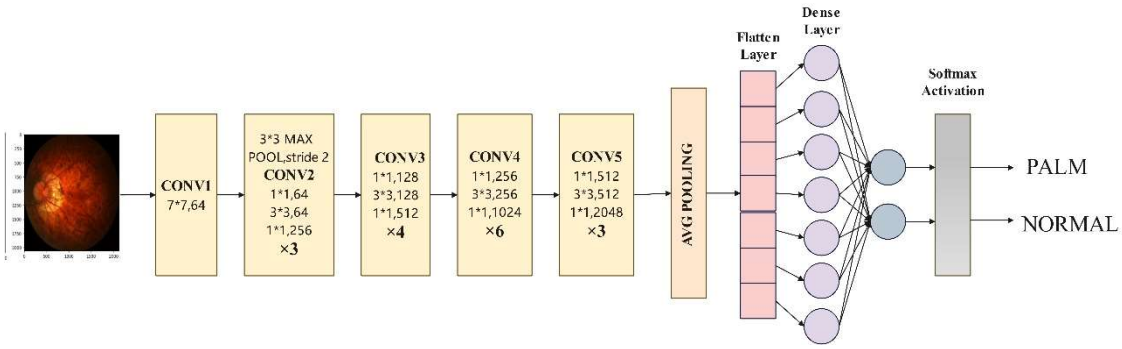

Figure 4. ResNet model structure

5. SR-init Structured Pruning

SR-init structured pruning reduces model complexity by evaluating the importance of filters in each convolutional layer of the network and deciding whether to retain these filters based on a set threshold. The core of this method is to evaluate the importance of these parameters by randomly reinitializing the layer parameters. By comparing the output differences of the model before and after reinitialization, the impact of each filter on the model performance can be determined.

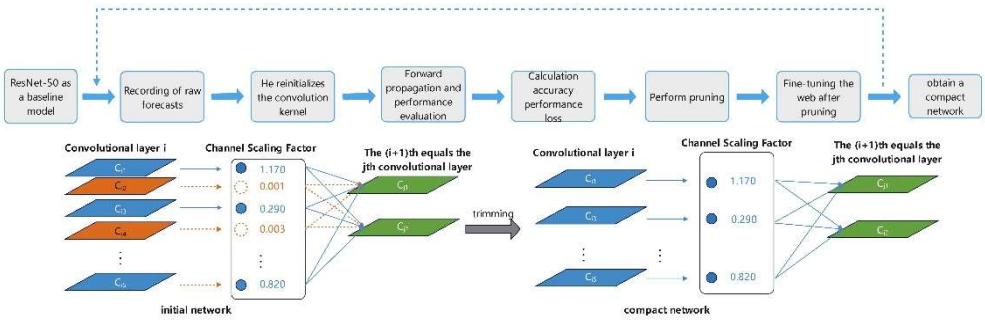

Figure 5. SR-init pruning process

First, this paper uses the optimal model ResNet-50 selected in the previous step as the baseline model, and takes the prediction results of the ResNet-50 model as the original prediction results. Then, the

ISSN: 2414-1895 DOI: 10.6919/ICJE.202506_11(6).0010

convolution kernels in the convolution layer are selectively reinitialized by He. This method is a weight initialization method specially designed for the ReLU activation function. The weight initialization formula of the filter is as follows:

$$W \sim \mathcal{N}(0, \sqrt{\frac{2}{n_l}}) \tag{1}$$

In the formula,W represents the weight of the model, n_l is the number of neurons in the previous layer, The weights are drawn from a normal distribution with mean 0 and variance $\mathrm{Var}(W) = \frac{2}{n_{|-1}}$.

Use the weights after He reinitialization to propagate the model forward and record the accuracy of the reinitialized model. Calculate the difference between the reconstructed network output and the original network output. The specific performance loss formula is:

$$\mathrm{drop}_i = |\ \mathrm{Accuracy}_{\mathrm{pruned}} - \mathrm{Accuracy}_{\mathrm{baseline}}\ | \tag{2}$$

In the formula, $\mathrm{Accuracy}_{\mathrm{pruned}}$ represents the accuracy of the model on the validation set after a certain layer is reinitialized, while $\mathrm{Accuracy}_{\mathrm{baseline}}$ represents the accuracy of the original model. This difference measures the impact of the reinitialization of each layer's weights on the model performance. If the value of drop_i is less than the predetermined performance loss threshold θ, it is considered that pruning the layer will not significantly affect the overall performance of the model, and pruning is performed.

The layers of the randomly initialized model are used to replace the corresponding layers of the baseline model one by one[11], and the accuracy change of the model is calculated, as shown in Figure 7. The blue dotted line in the figure is the accuracy of the baseline model, and the layers passed by the orange line are the layers that can be removed. The blue dotted line in the figure is the accuracy of the baseline model, and the layers passed by the orange line are the layers that can be removed.

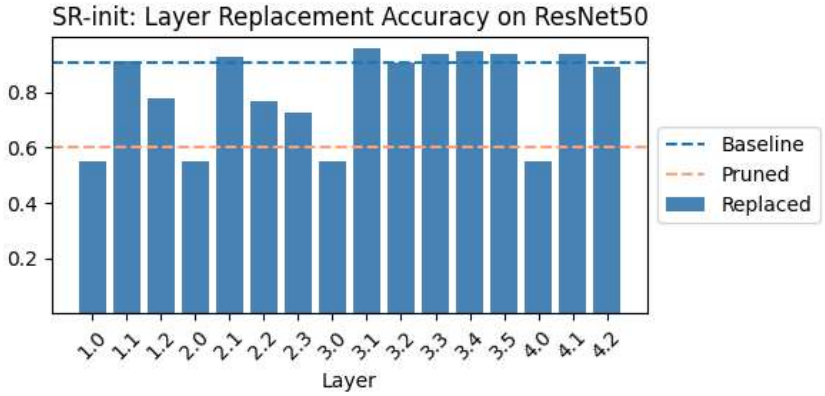

Figure 6. Accuracy changes after replacing layers in the ResNet50 model

The reasons for choosing these layers for pruning are explained in detail below. As shown in Figure 8, four pruning methods were tried, namely

1) Only pruning 9 layers near the baseline accuracy (indicated by the orange dotted line, Pruned_0)

2) On the basis of 1, additional pruning of layers 1.2 (indicated by the green dotted line, Pruned_1)

3) On the basis of 2, additional pruning of layers 2.1 (indicated by the light blue dotted line, Pruned_2)

ISSN: 2414-1895 DOI: 10.6919/ICJE.202506_11(6).0010

4) On the basis of 3, additional pruning of layers 2.2 (indicated by the purple dotted line, Pruned_3)

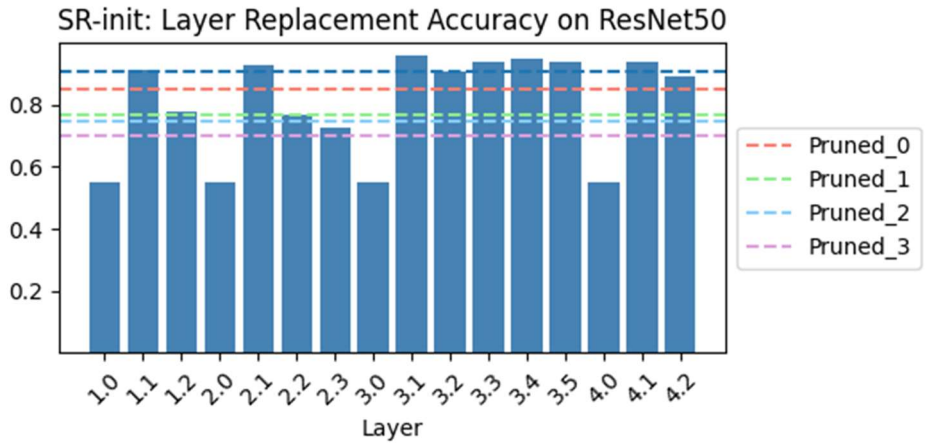

Figure 7. Comparison of four pruning methods

The pruned model was restored using self-distillation technology, and its accuracy, precision, recall and other indicators were tested on the test set. The results are shown in Figures 8.

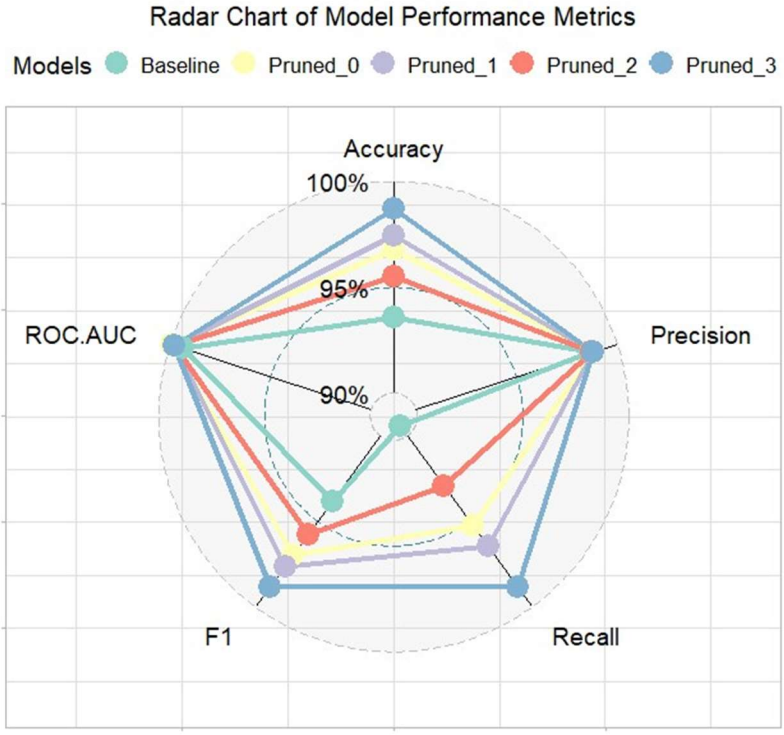

Figure 8. Model performance indicators

As can be seen from the figure, the five performance indicators of the Pruned_3 pruning method are higher than those of the other three pruning methods and the baseline model. Therefore, we chose Pruned_3 as the final pruning method. The model parameters and FLOPs of the computational overhead under various pruning strategies are calculated, and the results are shown in the figure9:

ISSN: 2414-1895 DOI: 10.6919/ICJE.202506_11(6).0010

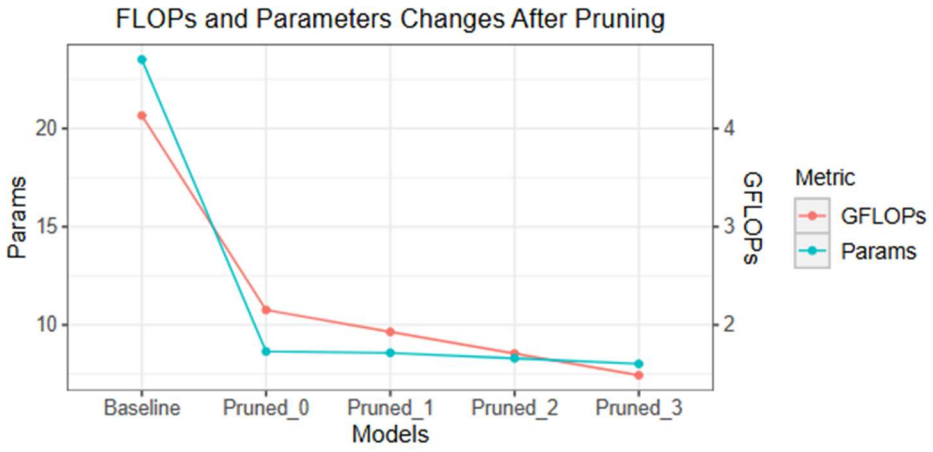

Figure 9. Reduction of model FLOPs and parameters

After pruning, the number of model parameters and computational overhead FLOPs have decreased, as shown in Figure 9. The number of parameters in the baseline model is about 23.5M, which is reduced to about 8M after pruning, a reduction of 65.9%; the FLOPs of the baseline model is about 4.1G, which is reduced to about 1.5G after pruning, a reduction of 64%.

6. Self-distillation Technology

In order to better restore the pruned model, this paper adopts self-distillation technology and uses the model's own output as a "soft label" to optimize the performance of the neural network.

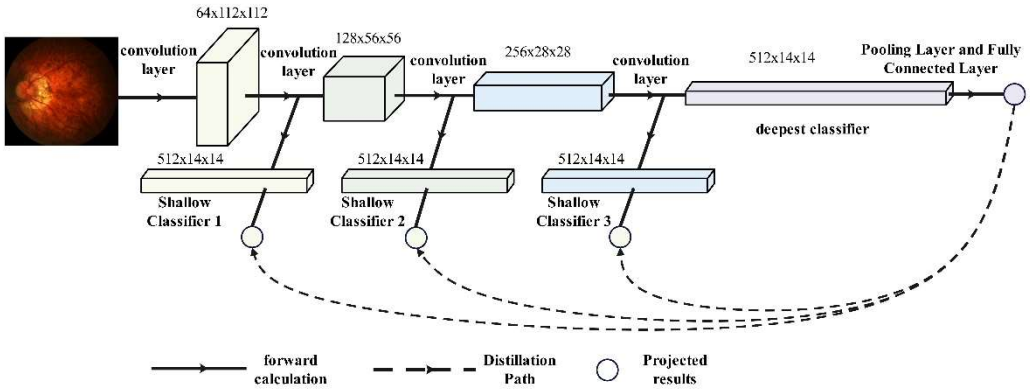

Figure 10. Self-distillation process

In the self-distillation process, the model first forward propagates the training data and outputs the predicted probability of each category. These probabilities are temperature-adjusted to generate soft labels:

$$q_i = \frac{\exp(z_i/T)}{\sum_j \exp(z_j/T)} \tag{3}$$

Where (z_i) is the original output $logits$ of the model for the i category, and T is the temperature parameter used to control the smoothness of the probability distribution. Then the loss function is configured. First, the model performs forward propagation again, and new outputs will be generated

ISSN: 2414-1895 DOI: 10.6919/ICJE.202506_11(6).0010

using the same batch of data. These outputs are used to calculate the loss with the soft label. The formula for the distillation loss is as follows:

$$L_{distill} = -\sum_i q_i \log p_i \quad (4)$$

Wherep_i is the current output probability of the model at a given temperature T, i.e., the new $logits$, which is the result obtained after softmax transformation. Here q_i is the soft label calculated previously.

The classification loss is the standard cross entropy loss calculated between the model predictions and the original hard labels:

$$L\_class = -\sum(y_i \log p_i) \quad (5)$$

In the formula,y_i is the actual class label. After obtaining the distillation loss and classification loss, the total loss function can be obtained:

$$L = (1-\alpha)L_class + \alpha T^2 L_distill \quad (6)$$

Where α is the distillation loss weight, set to 0.5, and T is the temperature parameter, set to 2. The T^2 factor is used to appropriately scale the distillation loss to ensure that the gradient contributed by the distillation loss is balanced relative to the gradient of the classification loss. Finally, during the self-distillation training process, the performance of the model on the validation set is continuously monitored. An early stopping strategy is implemented to prevent overfitting. If the validation set accuracy of the model does not increase by more than the preset 1% in 5 consecutive training cycles, the training is stopped.

Distillation temperature is a key hyperparameter in knowledge distillation, which is used to control the smoothness of the probability distribution output by the teacher model. The significance of optimizing the distillation temperature is to balance the effect of knowledge transfer and the learning ability of the student model.

In this paper, we use the Bayesian optimization method to optimize the distillation temperature in the range of (0, 20). The corresponding situation of the distillation temperature and the validation set loss is shown in Figure 11 below. When the distillation temperature is 8.54, the validation set loss is the lowest, about 0.05.

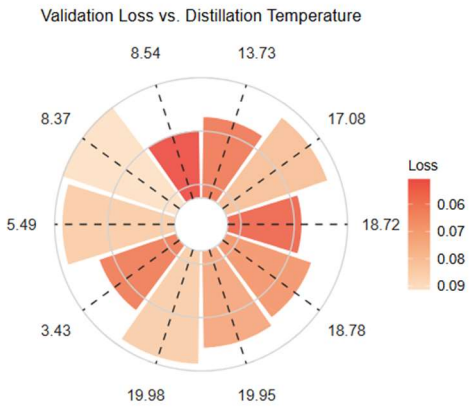

Figure 11. Distillation temperature and validation set loss

ISSN: 2414-1895 DOI: 10.6919/ICJE.202506_11(6).0010

7. Conclusion

This study proposes a pathological myopia diagnosis system based on ResNet's SR-init structured pruning and self-learning distillation, aiming to improve diagnostic efficiency, reduce misdiagnosis rate, and be suitable for resource-constrained medical environments. Through experiments on the iChallenge-PM fundus image dataset, this study verified the superior performance of the ResNet-50 model, and significantly reduced the model's parameter volume and computational overhead through SR-init pruning technology, with a 65.9% reduction in parameters and a 64% reduction in FLOPs. Self-distillation technology was further introduced, and the model was optimized using soft labels to enhance its ability to learn the complex features of pathological myopia.

In addition, this study used the AutoAugment algorithm for data enhancement, and combined with the early stopping strategy to effectively prevent overfitting and ensure the generalization performance of the model on unseen data. The experimental results show that the optimized model performs well in key indicators such as accuracy, precision, and recall, providing an efficient and reliable solution for the automatic diagnosis of pathological myopia.

Future research can further explore cross-modal learning, the applicability of different pruning strategies, and verification of larger datasets to improve the compatibility and generalization ability of the model. The results of this study provide an important reference for medical intelligence and automation, and have broad clinical application prospects.

Acknowledgments

The successful completion of this study is inseparable from the support and help of many parties. First of all, I would like to thank Baidu Brain and Sun Yat-sen University Zhongshan Eye Center for jointly providing the iChallenge-PM dataset, which provides a high-quality data basis for the experiment. In addition, I would like to thank every member of the research team for their time and energy in model design, experimental verification and result analysis.

In addition, I would like to thank peer experts for their valuable suggestions during the review process, which greatly improved the quality and rigor of the paper. Finally, I would like to thank my unit for its support in scientific research equipment and resources, which provided a good working environment for the research.

Supported by the 2024 University-level College Students' Innovation and Entrepreneurship Training Program Project (X2024010).

References

[1] Li Yongan. Research and implementation of pathological myopia recognition and lesion segmentation technology based on deep learning [D]. Guilin University of Technology, 2024. DOI: 10.27050/d.cnki.gglgc.2023.000733.

[2] Wan Cheng, Chen Baibing, Shen Jianxin, et al. Diagnosis method of high myopia based on artificial intelligence ResNeXt [J]. Practical Geriatrics, 2022, 36(03): 280-283.

[3] Zhang Yu, Wu Hai, Lin Fanchao, et al. Pruning technology of deep learning model in image recognition [J]. Journal of Nanjing University of Science and Technology, 2023, 47(05): 699-707. DOI: 10.14177/j.cnki.32-1397n.2023.47.05.018.

[4] Chen Yahao, Zhang Dong. Research on classification algorithm of color fundus images based on ResNet[J]. Computer Applications and Software, 2023, 40(08): 250-254+320. Shi Zhengjin, Wang Tianyu, Huang Zheng, Xie Feng, Song Guoli. A method for the automatic detection of myopia in Optos fundus images based on deep learning[J]. International journal for numerical methods in biomedical engineering, 2021, 37(6): e3460.1-e3460.15

[5] Yoo Tae Keun, Ryu Ik Hee, Kim Jin KukLee In Sik. Deep learning for predicting uncor-rected refractive error using posterior segment optical coherence tomography images[J]. Eye,2022,36(10):1959-1965.

[6] Abdolmaged Alkhulaifi, Fahad Alsahli, Irfan Ahmad. Knowledge distillation in deep learning and its applications[J]. PeerJ Computer Science,2021,-1(-1):-1.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202506_11(6).0010

[7] Zhang L, Bao C, Ma K. Self-distillation: Towards efficient and compact neural net-works[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 44(8): 4388-4403.

[8] Im J, Kasahara J Y L, Maruyama H, et al. Blend AutoAugment: Automatic Data Aug-mentation for Image Classification Using Linear Blending[J]. IEEE Access, 2024.

[9] Xia X, Li Y, Xiao G, et al. Benchmarking deep models on retinal fundus disease diagnosis-sis and a large-scale dataset[J]. Signal Processing: Image Communication, 2024: 117151.

[10] Ismail W N, Alsalamah H A, Mohamed E A. Genetic-efficient fine-tuning with layer pruning on multimodal Covid-19 medical imaging[J]. Neural Computing and Applica-tions, 2024, 36(6): 3215-3237.

[11] Doshi D, He T, Gromov A. Critical Initialization of Wide and Deep Neural Networks using Partial Jacobians: General Theory and Applications[J]. Advances in Neural Infor-mation Processing Systems, 2024, 36.

[12] Zhou R, Hu T. Evolving Better Initializations For Neural Networks With Prun-ing[C]//Proceedings of the Companion Conference on Genetic and Evolutionary Com-putation. 2023: 703-706.

[13] Fang G, Ma X, Wang X. Structural pruning for diffusion models[J]. Advances in neural information processing systems, 2024, 36.