European Review for Medical and Pharmacological Sciences 2022; 26: 8660-8674

A few-shot learning-based eye diseases screening method

Z.-K. HAN[1], H. XING[1], B. YANG[1], C.-Y. HONG[2]

[1]Zhejiang University, Hangzhou, Zhejiang, China
[2]Department of Ophthalmology, Center for Rehabilitation Medicine, Zhejiang Provincial People's Hospital (Affiliated People's Hospital, Hangzhou Medical College), Hangzhou, Zhejiang, China

Abstract. – OBJECTIVE: This study aims to construct a brand-new ophthalmic disease screening task and establish a practically valuable ophthalmic disease screening model in the case of insufficient data.

MATERIALS AND METHODS: The main methods are as follows: firstly, we mixed data from different sources (these data may come from different cameras, including different fundus diseases) to get a new dataset. Based on this dataset, we conducted subsequent experiments on fundus multi-disease screening. However, in the past public datasets, each dataset often only corresponded to the screening diagnosis of one disease. Secondly, we proposed a method to simulate the characteristics of different fundus cameras by using a method based on style transfer, and to augment the training data, so that the model could learn the features of ophthalmic diseases in a more comprehensive way. Finally, a robust disease screening model based on few-shot learning was constructed on the combined dataset, and compared with benchmark algorithms.

RESULTS: We focused on the study of eye disease screening methods based on the metric-based few-shot learning model, data augmentation methods, and focus on key technologies such as data augmentation based on style transfer. Experiments have shown that our method can significantly improve the generalization ability of the disease screening model.

CONCLUSIONS: By introducing few-shot learning theory and data augmentation based on style transfer into ophthalmic disease screening, the generalization ability of the model is greatly improved, and it has certain practical value.

Key Words:
Fundus images processing, Few-shot learning, Eye diseases screening, Computer vision.

Introduction

Vision is responsible for most of the information we absorb from our five combined senses. Blindness is the most serious disability, and its social concern is second only to cardiovascular diseases and tumors. Visual impairment has a significant impact on both personal life and social development. According to the Second National Survey of Disabled Persons in China, the annual number of sick people in China exceeds 5 billion, and the visually impaired population reaches 124 million.

Irreversible blindness caused by diabetic retinopathy, age-related macular degeneration, glaucoma, etc., has become the main cause of blindness in China at this stage. Most of these diseases need early detection and reasonable and timely intervention.

The current basic situation of blindness in China is worrying, and grass-roots blindness prevention work and primary eye care have become an important foundation for blindness prevention and blindness treatment projects. Much effort has been put on preventing the occurrence of eye diseases. However, there are only about 30,000 ophthalmologists in China, of which 10,000 doctors can treat fundus diseases.

Another major advantage of AI-assisted medical care is its high accessibility[1,2]. In recent years, AI diagnosis technology based on fundus color photography has become the target of many researchers. The convenience and efficiency of AI technology make it have great application prospects in ophthalmic disease screening, diagnosis and treatment[3,4]. The development of AI technology can optimize the allocation of medical resources, and more importantly, it can reduce the burden of medical costs and medical insurance funds, and effectively promote blindness prevention and treatment. We can use the current information technology conditions to collect a large amount of grassroots "fundus photo" data at the grassroots level, upload it to the cloud, and then analyze, identify, and classify it through experts or AI. To realize this vision, we can start with fundus photo data collection and AI analysis. It is

Corresponding Author: Bin Yang; e-mail: seven7@zju.edu.cn

relatively simple to realize this kind of condition similar to "general screening of eye diseases". First of all, we need a fundus camera that is portable, inexpensive, and can be easily taken by anyone with simple training. In addition, we need to design a special screening algorithm.

Few-Shot Learning

Few-shot learning is an application of meta-learning in supervised learning. Meta-learning is to let the model master the ability to learn, also known as "learning to learn". The core problem of few-shot learning is to allow machine learning models to acquire human-level fast learning capabilities.

The training set of few-shot learning consists of many classes, and there are multiple samples in each class. During the training phase, N categories will be randomly selected from the training set, with K samples in each category (a total of *N*K* data), and a task will be constructed as the input of the model's support set. A batch of samples is extracted from the remaining data in the model as the prediction object (batch set) of the model. That is, the model is required to learn how to distinguish the N categories from *N*K* data, such a task is called the N-way K-shot problem.

In recent years, few-shot learning has achieved a lot of results in computer vision. The mainstream few-shot learning methods can be roughly divided into three categories.

1. Model Based method: through the design of the model structure, the parameters are quickly updated on a small number of samples, and the mapping function between the input x and the predicted p-value is directly established.
2. Metric Based method: by measuring the distance between the sample and the sample in the support set, it is classified with the help of the idea of the nearest neighbor.
3. Optimization Based method: ordinary gradient descent methods are difficult to fit in few shot scenarios, so few shot classification can be accomplished by adjusting the optimization method.

Here we mainly use the metric-based method. Mainstream deep learning network classification models, such as Resnet and VGG, usually require a large amount of data for training to achieve better results, and when the data is insufficient, they tend to overfit. Instead, if we model the distance distribution between samples in the framework of meta-learning, so that similar samples are close, and heterogeneous samples are far away, the performance of model will be improved. Siamese Network[5] is a pioneering work on the metric-based method. The core idea of this method is to select two images from the data set as a set of sample pairs input each time, and the case where the labels of the images are consistent is recorded as a positive pair and the case where they are inconsistent is recorded as a negative pair. After the sampling is completed, the two images x0, x1 are input into a convolutional neural network f(x), we can regard f(x) as the image Convert x to a function of feature vector y, use f(x) to extract the Embedding y0, y1 of the two, and compare them to calculate the similarity between the two. The parameters of f(x) are shared between different tasks. The Siamese network is not a network structure, but uses a unique Contrastive Loss Function on the basis of other networks, the network structure for extracting Embedding is called Embedding Network, which makes the Siamese network structure adaptable.

Contrastive loss was originally proposed by Hadsell et al[6], which is mainly used in dimension reduction algorithms to ensure the similarity of features between samples after dimension reduction.

In 2015, Schroff et al[7] improved the Siamese network and proposed Triplet loss. The core idea is to select three samples (Anchor, Positive, Negative) from the dataset each time as a triplet for training. The training goal is to make Anchor and Positive as close as possible, and Anchor and Negative as far away as possible. Based on the idea of Triplet loss, Hoffer et al[8] proposed a Triplet Network, which has better experimental results than Siamese network on CIFAR, MNIST and other datasets. The core idea of Siamese network and triplet network is very simple, and they have achieved great results in practical applications. However, these two methods still have some shortcomings, mainly:

1. The number of possible pairs/triplets grows quadratically/cubically with the number of examples. It is infeasible to process them all.
2. We generate pairs/triplets randomly. As the training continues, more and more pairs/triplets are easy to deal with (their loss value is very small or even 0), preventing the network from training. We need to provide the network with hard examples.
3. Each image that is fed to the network is used only for computation of contrastive/triplet loss

for only one pair/triplet. The computation is somewhat wasted; once the embedding is computed, it could be reused for many pairs/triplets.

In order to deal with these problems effectively, Hermans et al[9] made a series of improvements to Triplet Network, and proposed Online Pair/Triplet Selector, which can better filter out intractable data (Hard Example). At the same time, by comparing the Embedding of the pictures, the efficiency is greatly improved.

Compared with the Siamese network, the Prototype Network[10] is based on the following idea: each category has a prototype expression, and the prototype of this category is the mean value of the support set in the embedding space. Then, the classification problem becomes the nearest neighbor problem in the Embedding space.

The network structures introduced above use a fixed measurement method for the final distance measurement, such as cosine distance, Euclidean distance, etc. All learning processes under this model structure occur in the Embedding stage of the sample.

Data Augmentation Based on Style Transfer

Data augmentation is a technology that variates the original data by artificial methods to obtain new training data, and enhances the size and quality of the training dataset, thereby obtaining a better model. It can effectively improve the robustness of the model and prevent overfitting. In the medical field where it is difficult to obtain a large amount of data, the importance of data enhancement technology has become more prominent.

Image data enhancement can be divided into two categories according to whether a neural network model is used. The first category is basic image processing methods, such as geometric transformation, translation, flipping, cropping, rotation, color space transformation, and noise injection[11], image mixing[12], etc. His deep learning model was originally the object of data augmentation services, but with the rapid development of deep learning in image processing, researchers quickly proposed deep learning-based methods for data augmentation, for example, Generative Adversarial Network (GAN), and style transfer.

Style is an abstract concept in the field of art. The style transfer method based on deep learning, originally proposed by Leon Gatys[13] in 2015, is one of the most important research hotspots in computer vision based on deep learning. Gatys believes that the artistic style of an image is the combination of its basic shapes and colors. The content information (shape) is represented by the feature map output by each layer of the neural network, and the style information is represented by the Gram matrix output by each layer of the neural network. Given a content image imgcontent and a style image imgstyle, the goal of model optimization is to make the output result close to imgcontent in content and imgstyle close. The loss function of the model is set as the weighted sum of the content loss losscontent and the style loss lossstyle:

$$loss = w_{content} * w_{content} + w_{style} * w_{style}$$

Where $w_{content}$, w_{style} represent the weight of content loss and style loss. The content loss can be measured by common loss functions in computer vision, such as MSE, while the style loss needs to be obtained through the Gram matrix. In this scenario, the meaning of the Gram matrix is the correlation between each feature in the feature maps output by different filters. The method proposed by Gatys opens up a new field. Subsequent researchers have carried out a lot of work on improving rendering speed and model reusability. According to how much style a model represents, the current mainstream style transfer methods can be divided into three categories:

1. Per-Style-Per-Model[14,15].
2. Multiple-Style-Per-Model[16-18].
3. Arbitrary-Style-Per-Model[19-21].

The practice of data augmentation based on style transfer has achieved success. For example, in autonomous driving, style transfer can be used to change the characteristics of day, night, weather, etc. of the original training data. At the same time, the camera used by the driverless car play the role as eyes which may be defaced or damaged. Style transfer can convert the data captured by the normal camera into the data in this extreme case. In addition, there are also cases of converting real-time rendered images in computer games to simulate real-world data[22].

DR

In the following chapters, we will introduce the current situation of AI ophthalmic diseases diagnosis. Diabetic retinopathy (DR) is the leading cause of blindness in adults. The classification of DR has always been a hot spot in ophthalmic AI research. As early as 1975, the STARE project[23] collated 400 fundus photos, 101 of them were from DR patients. Kaggle held two competitions around the classification of DR lesion grades in 2015 and 2019, attracting thousands of teams, and released the EyePACS dataset (also known as the Kaggle dataset) and the APTOS- 2019 dataset.

In recent years, the rapid development of deep learning has greatly improved the possibility of DR diagnosis with AI. Some scholars[24-26] applied classification models such as AlexNet and VGG networks to the DR classification task, and achieved very good results, which demonstrates that deep learning can effectively screen for diabetic retinopathy.

Most of the above research results are based on the ImageNet champion model of the year, or other excellent image classification, and conduct experiments on their private datasets or re-labeled public datasets. After these studies, the research became more diverse[27-29].

Glaucoma

Glaucoma is a chronic neurodegenerative disease and one of the leading causes of irreversible but preventable blindness in the world. Now, the high-quality public dataset for glaucoma is the Baidu-GON dataset, and the iChallenge-GON competition[30] is held around this dataset.

Early researchers viewed glaucoma diagnosis as a classification task, including binary classification based on the fundus photos' visual characteristics. Some scholars[31-34] use a relatively simple network for glaucoma classification, which achieves better results on small datasets, but sacrifices generalization ability. There are also some scholars[35-38] who use deeper models.

Image-level classification cannot fully utilize information on images. Another widely used approach is to limit the analysis to the optic disc region. This region is the region most affected by glaucoma, and this approach yields better performance than learning from full images, such as the method proposed by Chen et al[31]. The difficulty of this method is that the features of the optic cup and the optic disc are so close that the boundaries between the two are not evident in many fundus photos. How to train a model with generalization ability has become the biggest challenge in computer vision-based glaucoma screening.

AMD

Age-related macular degeneration (AMD), is a degenerative disease of the macula. It occurs mainly in people over the age of 45, and the incidence in the elderly is even higher than that of DR.

The etiology of AMD is not fully understood and may be related to a variety of factors, including genetics, chronic photo damaging effects, and nutrition. AMD is divided into dry AMD and wet AMD. Dry AMD (also called non-exudative AMD) does not develop new blood vessels. It is characterized by progressive atrophy of the retinal pigment epithelium (RPE). Wet AMD (also known as neovascular or exudative AMD) is characterized by neovascularization, exudation, hemorrhage, and scarring that, if left untreated, can eventually lead to irreversible damage to photoreceptors and rapid vision loss.

Early diagnosis of AMD is also critical for treatment and prognosis. Typical signs of AMD include drusen, oozing, bleeding, and more.

Yim et al[39] introduced a new AI system to predict the risk of total blindness in AMD patients in a paper published in Nature Medicine. Grassmann et al[40] used AlexNet, InceptionV3, ResNet and other models for experiments, and tested on the AREDS dataset; the sensitivity and specificity were 84.2% and 94.3% respectively.

Pathological Myopia

Pathological myopia (PALM) causes irreversible visual damage to patients. However, due to the lack of fundus photo datasets specifically for pathological myopia for a long time, there are few AI diagnosis research results for PALM. The method for diagnosing pathological myopia based on computer vision is similar to the previous diseases.

Materials and Methods

This paragraph presents our approach to initial screening questions for common fundus diseases, as well as the datasets and methods used. We first constructed a dataset containing various eye diseases, and then, augmented the data with a style transfer-based approach to make it consistent in style for subsequent training of the screening model. Finally, we constructed a multi-eye disease screening model based on few-shot learning to realize the screening of different eye diseases.

Data Collection

According to our needs, we are required to conduct a preliminary screening and diagnosis to determine whether there is a certain eye disease. To accomplish this task, we constructed a new dataset using the collected SYSU datasets, IChallenge-AMD, IChallenge-GON, IChallengePALM datasets with these four common ophthalmic diseases. The dataset labels are divided into six categories: healthy, AMD, glaucoma, PALM, NPDR (Non-Proliferative DR), and PDR (Proliferative DR). The dataset includes all patients and healthy people data in SYSU, IChallenge-AMD, IChallenge-GON, and PALM patient data in IChallenge-PALM.

Data Augmentation Based on Style Transfer

After investigating the public datasets, we found that the fundus photos taken by different models of fundus cameras are very different, such as the color of the optic disc, blood vessels, lesion area, the brightness of the photo, the color of the background part, and even the same type of fundus. Cameras, under the use of different users, will also have great differences due to operations, such as focusing. As it can be seen from Figures 1, 2, and 3, the fundus pictures from different datasets are generally consistent in shooting angle and shooting content. The biggest difference is that due to the different models of fundus cameras, it will have a greater impact on color and other aspects. Since the source of our training data is relatively single, during the training process, the features learned by the model may be the features of the fundus camera, not the features of the disease. To mitigate the impact of this problem, we employed a special data augmentation method for the problems brought about by the fundus camera model, drawing on the idea of style transfer, and trained a neural network model for a specific model of fundus camera. We converted the photos taken by a certain fundus photo to the style captured by other models of fundus cameras.

The effects of these two methods are not satisfactory. We decided to try the Multiple-Style-Per-Model method represented by the method proposed by Dumoulin et al[17]. Due to space limitations, more pictures were not arranged for comparison. As shown in Figure 6,

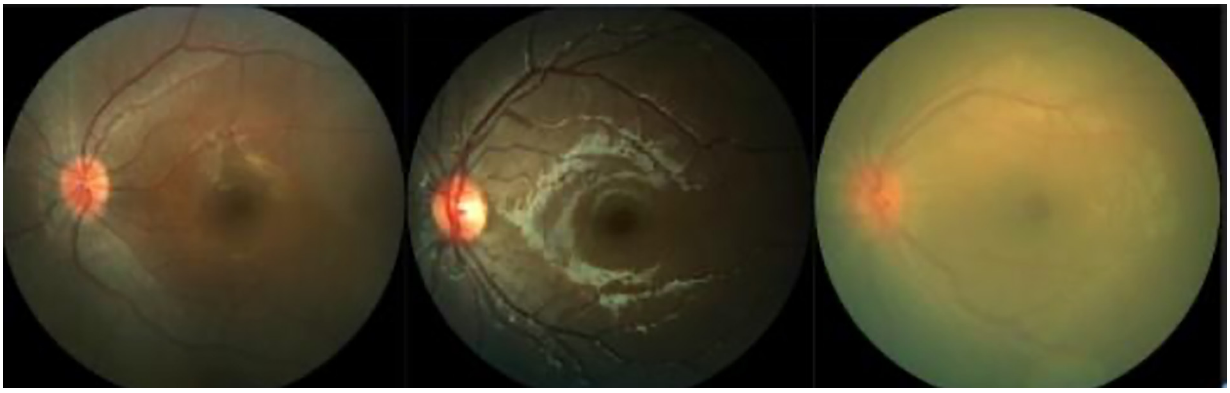

Figure 1. Baidu AMD Zeiss Visucam 500 healthy people (group 2).

Figure 2. SYSU dataset Topcon TRC 50DX, Japan healthy people (group 1).

Figure 3. Baidu-Glaucoma-Canon-CR-2.

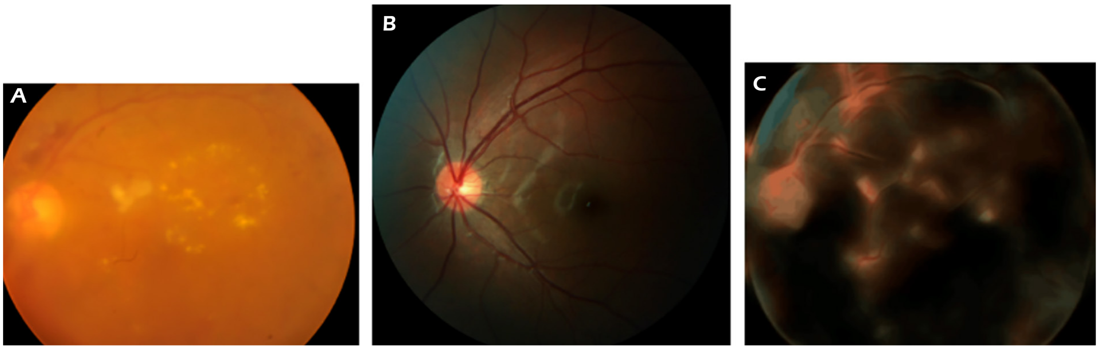

Figure 4. Universal Style Transfer style transfer effect on fundus images. **A,** Content image. **B,** Style image . **C,** Output result.

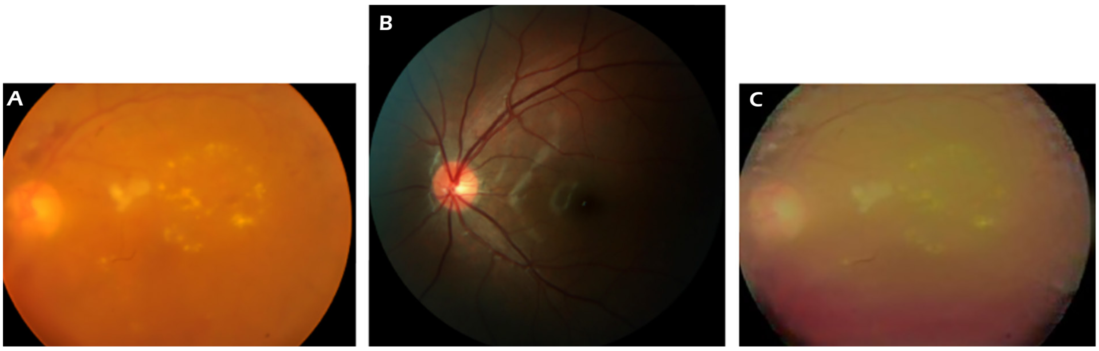

Figure 5. The effect of Gatys classic style transfer on the fundus image. **A,** Content image. **B,** Style image . **C,** Output result.

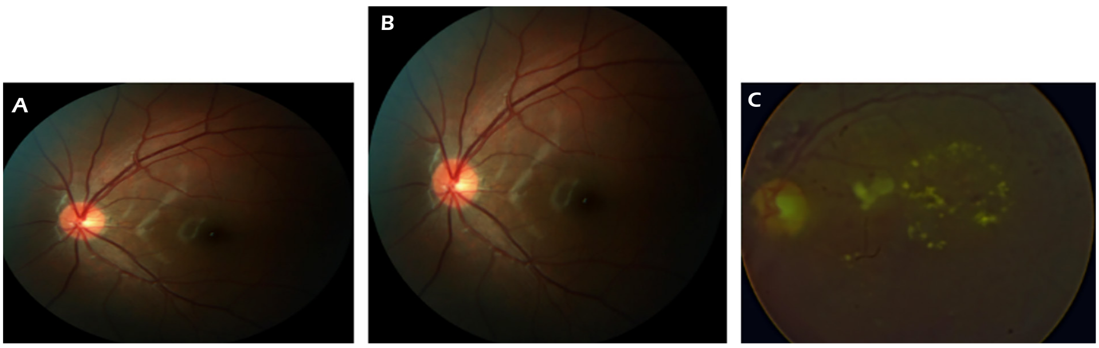

Figure 6. The effect of Dumoulin style transfer on fundus images. **A,** Content image. **B,** Style image . **C,** Output result.

despite its shortcomings, the Dumoulin method is the most suitable for our scene, successfully learning the content of the input image while preserving the content of the input image. While having a certain degree of controllability, the training speed is also acceptable. We can train special models for different types of fundus cameras, and convert photos taken by other fundus cameras to achieve the purpose of data enhancement. The data augmentation method based on style transfer is shown in Figure 7. Specific steps are as follows:

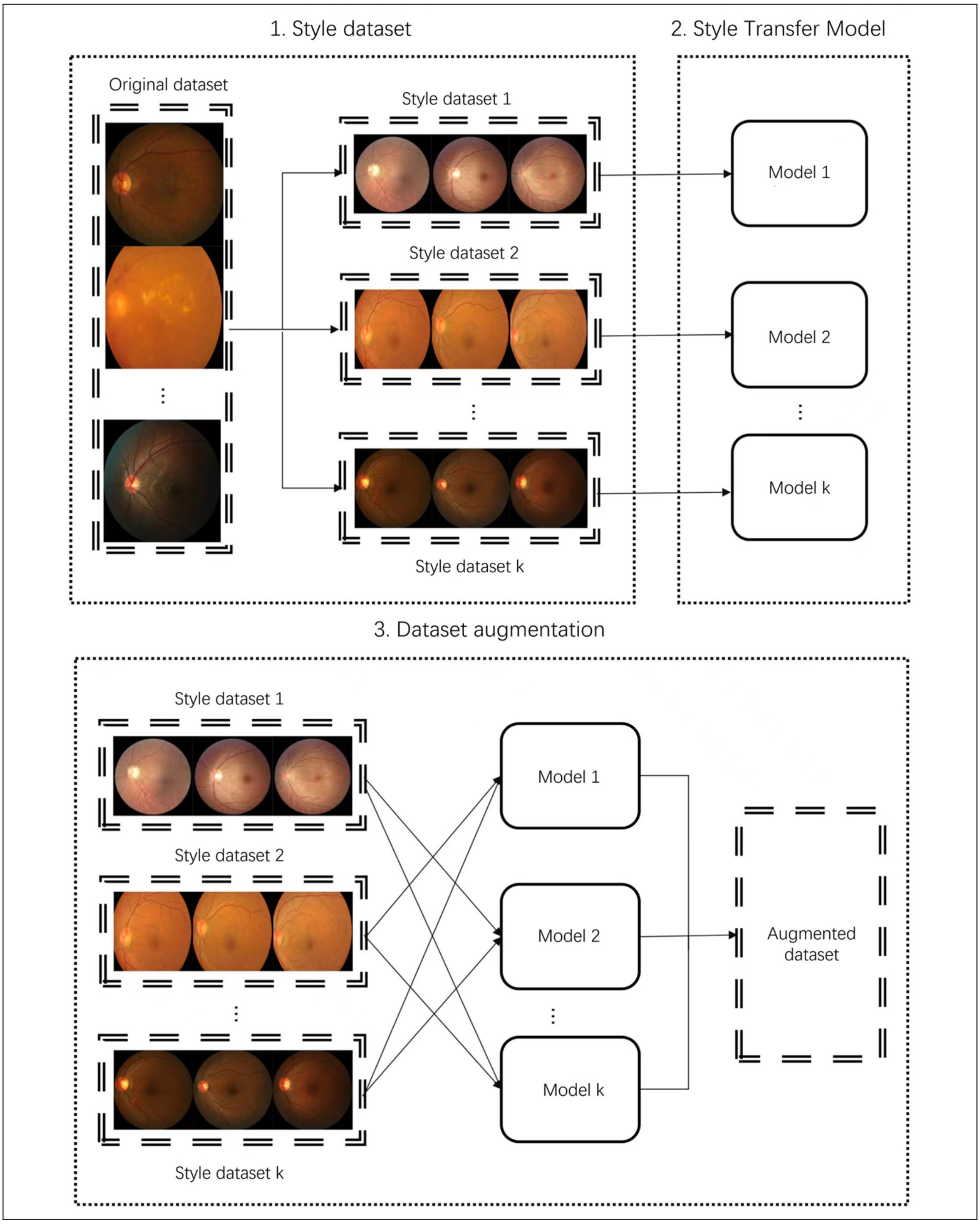

Figure 7. Data augmentation method based on style transfer.

1. According to the fundus camera model and other factors in the data, the original data set is divided into personal style datasets;
2. Train different style transfer models for different style datasets;
3. Each fundus photo in the dataset has its own style dataset. For the fundus photos under a certain style dataset, k-1 style transfer models with different styles are used for processing. The generated fundus photos were added to the augmented dataset.

In this way, the number of samples in the data set can be greatly expanded, and the influence of factors, such as fundus camera model and brightness on the data set can be reduced, thereby effectively improving the model performance and generalization ability.

Multi-Eye Disease Screening Model Based on Few-Shot Learning

In this scenario, the data distribution is extremely unbalanced, and the number of samples is extremely rare. In this case, Softmax classifier with cross-entropy loss will be severely overfitted. Even in some tasks with ample data, with hundreds of samples per class, neural networks can still overfit. Deep neural networks have millions of parameters to fit the training data, so they can learn a huge function space to fit various complex distributions. However, this also places great demands on the quality and quantity of training data, which has become a major obstacle in many tasks. To address this issue, we used a Siamese network to screen fundus photographs for ophthalmic diseases from the perspective of few-shot learning. Compared with methods such as Triplet Network, Prototype Network, and Relational Network, Siamese network is simple and easy to implement. At the same time, it requires less data to converge. Due to the simplicity of the Siamese network, it is more helpful for us to understand the problems in the data. We first use the Siamese network to conduct experiments, and then gradually improves it. The training diagram of Siamese network is shown in Figure 8. The specific steps are as follows:

1. Simultaneously extract two training data x_1, x_2 to form a sample pair;
2. Input the sample pair into the model to get their respective feature vectors V_1, V_2;
3. Compare the labels of the two, calculate the loss by comparing the loss function, and update the model parameters.

Evaluation Method

If we ignore the differences in detail and treat the network model as a black-box model, then, we can denote the output of the model as f(x), which accepts an image x as an input. For ordinary classification models and Siamese networks, their model output is the predicted image

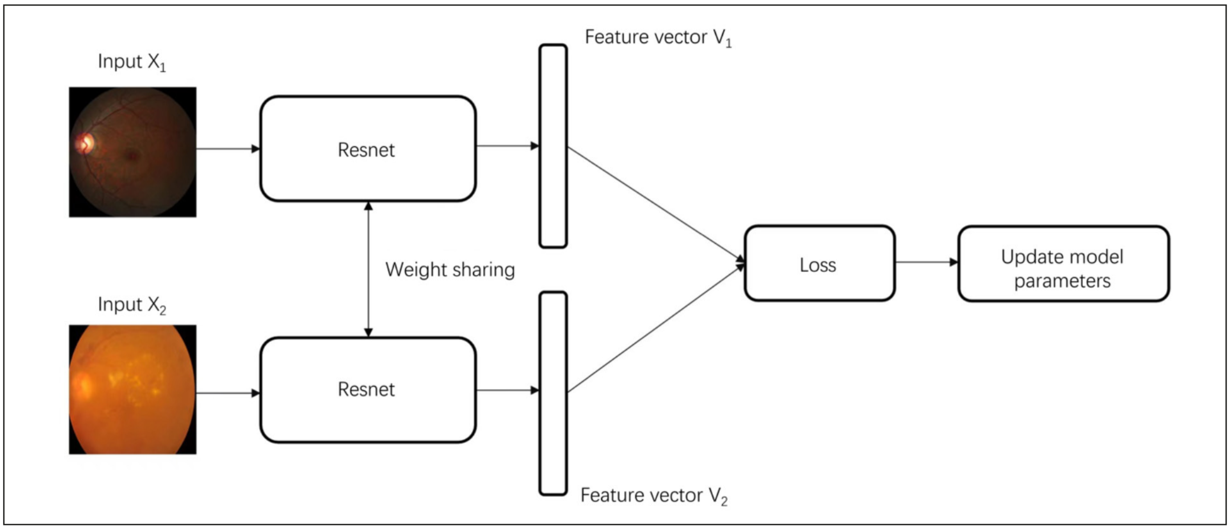

Figure 8. Raining diagram of Siamese network.

Table I. Dataset distribution statistics.

Labels	AMD	Glaucoma	Healthy people	NDPR	PALM	PDR	Total
Quantity	89	40	1302	462	213	58	2164

category y-pred. For this type of model, we only need to compare y-pred with the correct label y of the image. For the prototype network model, the output is also the type of image y-pred, and an evaluation method similar to the ordinary classification model can also be used. For the prototype network model, the output is also the type of the image y-pred, and an evaluation method similar to that of the ordinary classification model can also be used.

The Siamese network model accepts two images x_1, x_2 as input at the same time, and the output $f(x_1, x_2)$ is the distance between x_1 and x_2, rather than the category of a picture, so the evaluation method of the Siamese network model is quite different from that of the ordinary classification once[5,7].

The evaluation method[5] is used to extract several samples from each category in the test set, and these samples are used as the support set according to the original category. When the data of a query is input, the query data is compared with the data of each category in the support set, and finally the category with the highest similarity is selected as the result output. We will finally conduct a comprehensive evaluation of the model from three aspects: accuracy, recall, and f1 score.

Results

Training of Style Transfer Model

Based on the four data sets of SYSU, IChallenge-PM, IChallenge-GON, and IChallenge-AMD, this paper merges to form a new data set, which is divided into the following six categories according to disease types. Health, NPDR, PDR, AMD, glaucoma, PALM, the distribution of the dataset are shown in Table I.

The data set is divided according to the ratio of 4:1, with 1731 training data and 433 test data. In addition, we count the models of fundus cameras used in different datasets, as shown in Table II.

Considering that human error may lead to different styles of image captured by the same fundus camera, we can divide a dataset into several styles. In real conditions, such influence is unavoidable[41]. On this basis, we select several groups of representative images as style images. After training, we obtain the four styles as Table III.

Based on the four styles of data obtained, we conducted subsequent experiments with screening models and compared with methods without augmented data.

Table II. Fundus camera model.

Dataset name	Fundus camera model
SYSU	Topcon TRC-50DX45
iChallenge-GON	Training set Zeiss Visucam 500 test set Canon CR-2
iChallenge-AMD	Zeiss Visucam 500 Canon CR-2
iChallenge-PALM	Unknown

Table III. Style transfer model trained.

Model name	Data source	Style image sample size
Canon_CR2_0	Baidu-AMD	30
Zeiss_Visucam_500_1	Baidu-AMD	90
Topcon_TRC-50DX_1	SYSU	3
Topcon_TRC-50DX_2	SYSU	3

Table IV. Resnet-50 model test results.

	Precision	Recall	F1 value	Sample size
0 AMD	0.000	0.000	0.000	16
1 Glaucoma	0.000	0.000	0.000	8
2 Healthy	0.864	0.952	0.906	248
3 NPDR	0.784	0.842	0.812	95
4 PALM	0.865	0.900	0.882	50
5 PDR	0.000	0.000	0.000	15
Accuracy		0.836		432
Mean	0.419	0.449	0.433	432
Weighted average	0.769	0.836	0.801	432

Training of Screening Model

In order to verify the effect of the data augmentation method based on style transfer and the eye disease screening model based on few-shot learning proposed in this paper, this section conducts comparative experiments on Resnet and Siamese-Resnet on unenhanced and enhanced data, respectively.

In order to compare the difference between the usual classification network and the Siamese network, we first conduct experiments using the Resnet-50 model and Siamese-Resnet-50.

The experimental results of Resnet with unenhanced data are shown in Table IV. In order to show the distribution of the prediction results more intuitively and compare with the subsequent experimental results, we draw the output results of the final fully connected layer of Resnet-50 on a two-dimensional plane, as shown in Figure 9. As it can be seen from Table IV, the model only has high recognition ability for healthy people, non-proliferative diabetic retinal complications (NPDR), and pathological myopia (PALM) with sufficient sample size. As it can be seen from Figure 9, there is a large intersection between various categories, indicating that the model has not fully learned the characteristics of various categories, and that the Resnet-50 model has greater limitations in the task of ophthalmic disease decoration.

Table V and Figure 10 show the Siamese-Resnet performance on unenhanced data. For the types with insufficient samples, the classification effect of the modular Siamese-Resnet has been improved to a certain extent. The intersection between healthy people and patients was reduced, and PDR and NPDR within DR patients were also better differentiated. However, there is still a large overlap between healthy people and patients.

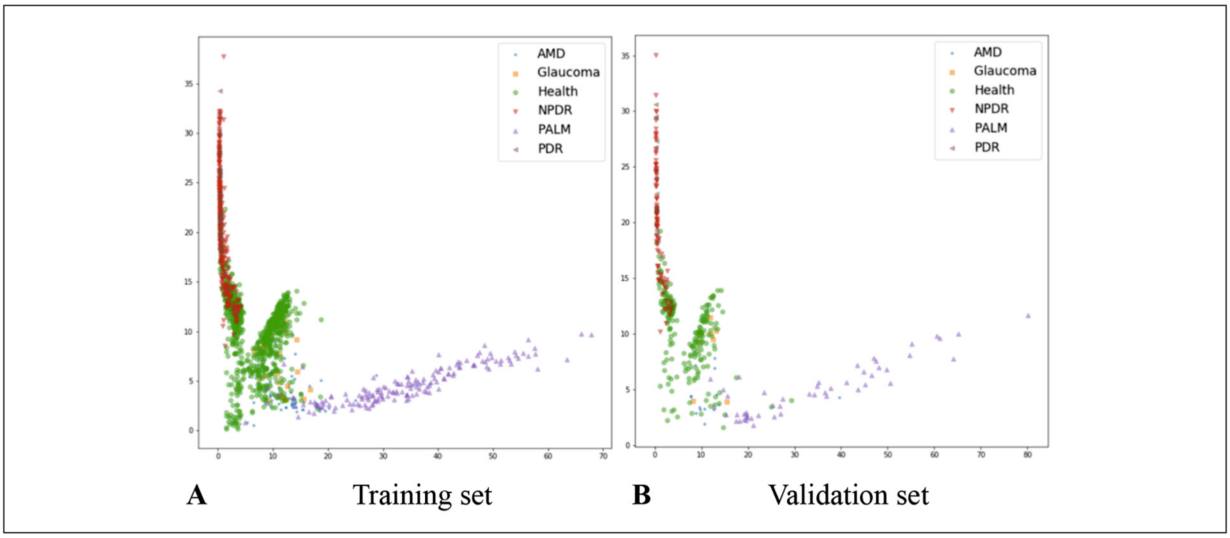

Figure 9. Classification results of ResNet-50 after 40 rounds of training. **A,** Training set. **B,** Validation set.

Table V. Siamese-Resnet-50 model test results.

	Precision	Recall	F1 value	Sample size
0 AMD	0.063	0.312	0.105	v16
1 Glaucoma	0.084	0.550	0.146	8
2 Healthy	0.830	0.416	0.555	248
3 NPDR	0.514	0.437	0.472	95
4 PALM	0.885	0.751	0.813	50
5 PDR	0.148	0.520	0.230	15
Accuracy		0.458		432
Mean	0.421	0.498	0.387	432
Weighted average	0.701	0.462	0.531	432

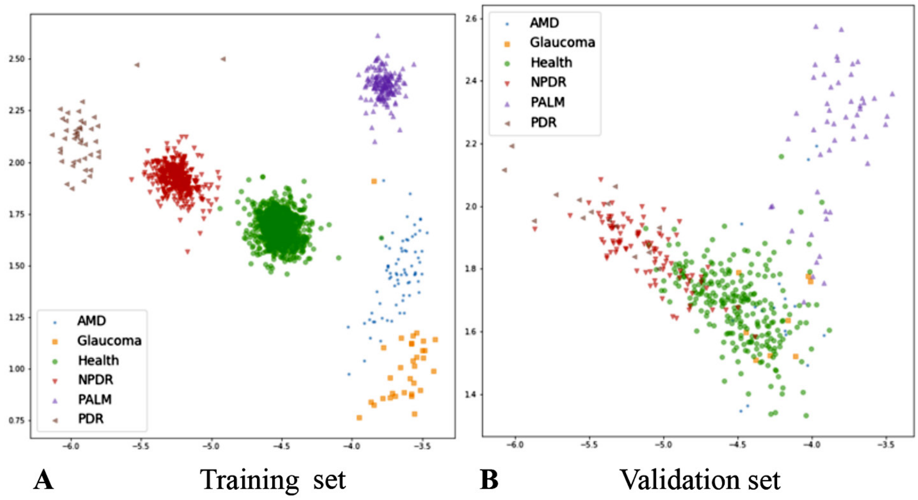

Figure 10. Classification results of Siamese-Resnet-50 after 40 rounds of training. **A,** Training set. **B,** Validation set.

After that, we performed follow-up experiments by augmenting the data with the style transfer method described in the previous section. The experimental results are shown in Table VI and Table VII, and Figure 11 and Figure 12. Compared with the effect before data enhancement (Table V), after Siamese-Resnet uses style transfer enhancement, the misclassification of the model is also greatly reduced, and the effect is greatly improved. By comparing Table IV, it can be seen that Siamese-Resnet-50 is much better for PDR identification.

Table VI. Resnet-50 model with augmented data test results.

	Precision	Recall	F1 value	Sample size
0 AMD	0.300	0.188	0.231	16
1 Glaucoma	0.333	0.125	0.182	8
2 Healthy	0.883	0.940	0.911	248
3 NPDR	0.784	0.842	0.812	95
4 PALM	0.887	0.940	0.913	50
5 PDR	0.000	0.000	0.00	15
Accuracy		0.843		432
Mean	0.531	0.506	0.508	432
Weighted average	0.799	0.843	0.819	432

Table VII. Siamese-Resnet-50 model with augmented data test results.

	Precision	Recall	F1 value	Sample size
0 AMD	0.271	0.600	0.374	16
1 Glaucoma	0.071	0.250	0.110	8
2 Healthy	0.947	0.784	0.857	248
3 NPDR	0.846	0.783	0.814	95
4 PALM	0.953	0.930	0.941	50
5 PDR	0.374	0.660	0.477	15
Accuracy		0.779		432
Mean	0.577	0.668	0.596	432
Weighted average	0.864	0.779	0.813	432

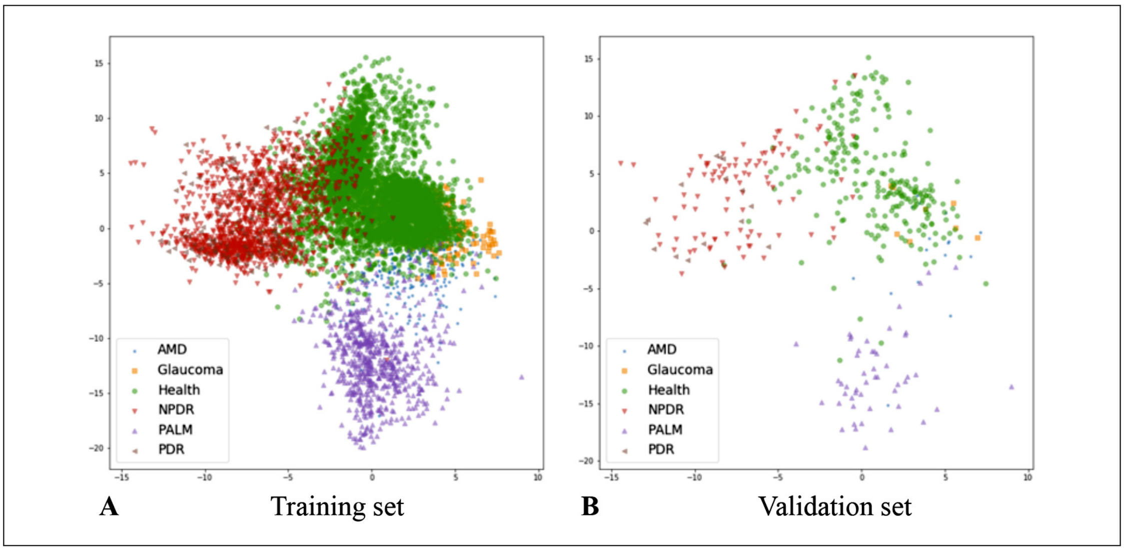

Figure 11. Classification results of ResNet-50 with augmented data after 40 rounds of training. **A,** Training set. **B,** Validation set.

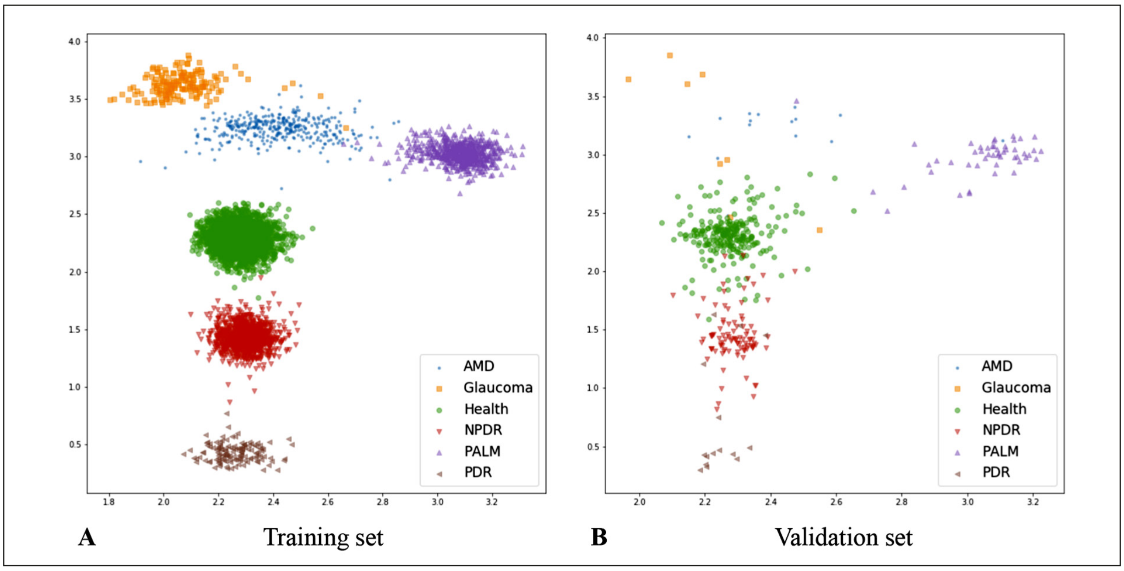

Figure 12. Classification results of Siamese-Resnet-50 with augmented data after 40 rounds of training. **A,** Training set. **B,** Validation set.

It can be seen from Figure 11 that the cluster centers of different classes and the decision boundary between classes are very clear, and the distribution is ideal. The experimental results show that Siamese-Resnet-50 can be well qualified for the screening and classification task of ophthalmic diseases.

Discussion

In the data enhancement method based on style transfer proposed in this paper, the style image set still needs to be manually selected, but for larger data sets, manual selection is not feasible. If we can propose an unsupervised clustering method to automatically gather similar images together, it will be more conducive to the selection of style images.

The method based on style migration is still expanded on the basis of the existing data. For classes with extremely small data, the effect is still not evident. If we can further study the method based on GAN and generate various types of fundus photos in a controlled way, it will bring more far-reaching impact.

Furthermore, fundus photos are not the only basis for diagnosis of ophthalmic diseases, and multimodal data such as medical records, should be integrated into the diagnosis method in our future research.

Conclusions

This paper starts with the current situation of difficult prevention and treatment, sorts out the efforts of AI in the prevention and treatment of ophthalmic diseases in recent years and summarizes the necessity and significance of developing a deep learning-based ophthalmic disease prevention method. Based on the research direction of this paper, the introduction chapter focuses on the medical knowledge related to ophthalmology and the data augmentation methods that may be used to solve the data insufficiency. In addition, some pre-research and related work related to AI in ophthalmology are also introduced. Based on the work of many predecessors, this paper proposes a model that can learn the characteristics of different types of fundus cameras through style transfer, thereby improving model performance and generalization ability and alleviating the data shortage. This paper introduces how to use a metric-based few-shot learning model, and analyzes the feasibility and necessity of using a metric-based few-shot learning model for ophthalmic diseases. Based on the above ideas, this paper trains a robust ophthalmic disease classification model on a dataset composed of common ophthalmic disease image datasets, and compares it with the benchmark algorithm. The experimental results show the effect of the data augmentation method based on style transfer proposed in this paper and demonstrate the superiority of our algorithm. In addition, a brand-new task screening and classification of multiple ophthalmic diseases was constructed, and a variety of metric-based few-shot learning models were tested into this task, which achieved good results.

Conflict of Interest

The Authors declare that they have no conflict of interests.

Ethics Approval

Not required.

Funding

This paper is supported by Provincial key research and development plan of Zhejiang science and technology program (No. 2021C03103).

Availability of Data and Material

We conduct experiments on public datasets which are available on the official website.

Informed Consent

Not applicable.

Authors' Contribution

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Zhike Han, Hao Xing, Bin Yang and Chaoyang Hong. The first draft of the manuscript was written by Zhike Han, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

ORCID ID

Zhike Han: 0000-0002-4947-2426.

References

1) Castro A F, Silva-Turnes J C, Gonzalez F. Evaluation of retinal digital images by a general practitioner. Telemed J E-health 2007; 13: 287-292.
2) Pongsakonpruttikul N, Angthong C, Kittichai V, Chuwongin S, Puengpipattrakul P, Thongpat P, Boonsang S, Tongloy T. Artificial intelligence assistance in radiographic detection and classification of knee osteoarthritis and its severity: a cross-sectional diagnostic study. Eur Rev Med Pharmacol Sci 2022; 1549-1558.
3) Alabdulwahhab KM, Sami W, Mehmood T, Meo SA, Alasbali TA, Alwadani FA. Automated detection of diabetic retinopathy using machine learning classifiers. Eur Rev Med Pharmacol Sci 2021; 583-590.
4) Alsharif MH, Alsharif YH, Yahya K, Alomari OA, Albreem MA, Jahid A. Deep learning applications to combat the dissemination of COVID-19 disease: A review. Eur Rev Med Pharmacol Sci 2020; 24: 11455-11460.
5) Koch G, Zemel R, Salakhutdinov R. Siamese neural networks for one-shot image recognition. ICML Deep Learning Workshop 2015.
6) Hadsell R, Chopra S, LeCun Y. Dimensionality reduction by learning an invariant mapping. 2006 IEEE Comput. Soc. Conference on Comput. Vision and Pattern Recognition IEEE 2006; 2: 1735-1742.
7] Schroff F, Kalenichenko D, Philbin J. Facenet: A unified embedding for face recognition and clustering. Proceedings of the IEEE Conference on Comput. Vision and Pattern Recognition 2015: 815-823.
8) Hoffer E, Ailon N. Deep metric learning using triplet network. Intl Workshop on Similarity-based Pattern Recognition Springer Cham 2015; 84-92.
9) Hermans A, Beyer L, Leibe B. In defense of the triplet loss for person re-identification. arXiv preprint arXiv 1703.07737, 2017.
10) Snell J, Swersky K, Zemel R. Prototypical networks for few-shot learning. Advances in Neural Info Processing Systems 2017; 30.
11) F. J. Moreno-Barea, F. Strazzera, J. M. Jerez, D. Urda, L. Franco. Forward noise adjustment scheme for data augmentation. 2018 IEEE Symposium Series on Comput. Intelligence IEEE 2018; 728-734.
12) Inoue H. Data augmentation by pairing samples for images classification. arXiv preprint arXiv: 1801.02929, 2018.
13) Gatys L A, Ecker A S, Bethge M. A Neural Algorithm of Artistic Style. arXiv preprint arXiv: 1508.06576, 2015.
14) Johnson J, Alahi A, Fei-Fei L. Perceptual losses for real-time style transfer and super-resolution. European Conference on Comput Vision Springer Cham 2016; 694-711.
15) Li C, Wand M. Precomputed real-time texture synthesis with markovian generative adversarial networks. European Conference on Comput Vision Springer Cham 2016; 702-716.
16) Chen DD, Yuan L, Jing Liao J, Nenghai Yu NH, Hua G. Stylebank: An explicit representation for neural image style transfer. Proceedings of the IEEE Conference on Comput Vision and Pattern Recognition 2017; 1897-1906.
17) Dumoulin V, Shlens J, Kudlur M. A learned representation for artistic style. arXiv preprint arXiv: 1610.07629, 2016.
18) Zhang H, Dana K. Multi-style generative network for real-time transfer. Proceedings of the European Conference on Comput Vision Workshops 2018.
19) Jing Y, Liu X, Ding Y, Wang X, Ding E, Song M, Wen S. Dynamic instance normalization for arbitrary style transfer. Proceedings of the AAAI Conference on Artificial Intelligence 2020: 34; 4369-4376.
20) Li Y, Fang C, Yang J, Wang Z, Lu X, Yang MH. Universal style transfer via feature transforms. Advances in Neural Inform Processing Systems 2017; 30.
21) Shen F, Yan S, Zeng G. Neural style transfer via meta networks. Proceedings of the IEEE Conference on Comput Vision and Pattern Recognition 2018; 8061-8069.
22) Shorten C, Khoshgoftaar TM. A survey on image data augmentation for deep learning. Journal of Big Data 2019; 6: 1-48.
23) Hoover A D, Kouznetsova V, Goldbaum M. Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. IEEE T Med Imaging 2000; 19: 203-210.
24) Abràmoff MD, Lou Y, Erginay A, Clarida W, Amelon R, Folk JC, Niemeijer M. Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. Invest Ophth Vis Sci 2016; 57: 5200-5206.
25) Abràmoff MD, Lavin PT, Birch M, Shah N, Folk JC. Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. NPJ Digit Med 2018; 1: 39.
26) Ting DSW, Cheung CY, Lim G, Tan GSW, Quang ND, Gan A, Hamzah H, Garcia-Franco R, San Yeo IY, Lee SY, Wong EYM, Sabanayagam C, Baskaran M, Ibrahim F, Tan NC, Finkelstein EA, Lamoureux EL, Wong IY, Bressler NM, Sivaprasad S, Varma R, Jonas JB, He MG, Cheng CY, Cheung GCM, Aung T, Hsu W, Lee ML, Wong TY. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. JAMA 2017; 318: 2211-2223.
27) Zeng X, Chen H, Y Luo Y, Ye W. Automated diabetic retinopathy detection based on binocular siamese-like convolutional neural network. IEEE Access 2019; 7: 30744-30753.

28) Jain A, Jalui A, Jasani J, Lahoti Y, Karani R. Deep learning for detection and severity classification of diabetic retinopathy. 2019 1st Intl Conference on Innovations in Inform and Commun Tech IEEE 2019; 1-6.

29) Gwenolé Q, Katia C, Yassine Bi, Béatrice C, Mathieu L. Deep image mining for diabetic retinopathy screening. Med Image Anal 2017; 39: 178-193.

30) Khojasteh P, Aliahmad B, Kumar D K. Fundus images analysis using deep features for detection of exudates, hemorrhages and microaneurysms. BMC Ophthalmol 2018; 18: 1-13.

31) Chen X, Xu Y, DW Kee Wong, TY Wong, J Liu. Glaucoma detection based on deep convolutional neural network. 2015 37th Annual Intl. Conference of the IEEE Eng in Med and Boil Soc IEEE 2015; 715-718.

32) Chen X, Xu Y, Yan S, Wong DWK, Wong TY, Liu J. Automatic feature learning for glaucoma detection based on deep learning. Intl Conference on Med image Comput and Comput-assisted Intervention Springer Cham 2015; 669-677.

33) U Raghavendra, Hamido Fujita, Sulatha V Bhandary, Anjan Gudigar, Jen Hong Tan, U Rajendra Acharya. Deep convolution neural network for accurate diagnosis of glaucoma using digital fundus images. Inform Sci 2018; 441: 41-49.

34) Li A, Wang Y, Cheng J, Liu J. Combining multiple deep features for glaucoma classification. 2018 IEEE Intl. Conference on Acoustics, Speech and Signal Processing IEEE 2018; 985-989.

35) Christopher M, Belghith A, Bowd C, Proudfoot James A, Goldbaum Michael H, Weinreb Robert N, Girkin Christopher A, Liebmann Jeffrey M, Zangwill Linda M. Performance of deep learning architectures and transfer learning for detecting glaucomatous optic neuropathy in fundus photographs. Sci Rep 2018; 8: 1-13.

36) Li ZX, He YF, Stuart K, Meng W, Robert T. Chang, He MG. Efficacy of a deep learning system for detecting glaucomatous optic neuropathy based on color fundus photographs. Ophthalmology 2018, 125: 1199-1206.

37) José I O, Elena P, Mariana del F, Matthew B B. Convolutional neural network transfer for automated glaucoma identification. 12th intl. symposium on med. Inform Processing and analysis SPIE 2017; 10160: 241-250.

38) Gómez-Valverde JJ, Antón A, Fatti G, Liefers B, Herranz A, Santos A, Sánchez CI, Ledesma-Carbayo MJ. Automatic glaucoma classification using color fundus images based on convolutional neural networks and transfer learning. Biomed Opt Express 2019; 10: 892-913.

39) Jason Y, Reena C Terry S, Jim W, Annette O, Christopher K, Harry A, Marko L, Josef H, Katrin Fr, Gabriella M, Clemens M, Marc W, Jonathan D, Cian HR, Peng TK, Alan a, Dominic K, Demis H, Mustafa S, Trevor B, Joseph RL, Pearse AK, Jeffrey DF. Predicting conversion to wet age-related macular degeneration using deep learning. Nat Med 2020; 26: 892-899.

40) Felix G, Judith M, Caroline B, Sebastian H, Martina E. Z, Birgit L, Annette P, Iris MH, Christoph P, Bernhard HF, Weber. A deep learning algorithm for prediction of age-related eye disease study severity scale for age-related macular degeneration from color fundus photography. Ophthalmology 2018; 125: 1410-1420.

41) Emma B, Elizabeth B, Fred H, Anna L, Lauren W, Paisan Ruamviboonsuk, Laura M. Vardoulakis. A human-centered evaluation of a deep learning system deployed in clinics for the detection of diabetic retinopathy. Proceedings of the 2020 CHI Conference on Human Factors in Comput Systems 2020; 1-12.