IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 21 September 2023, accepted 10 October 2023, date of publication 13 October 2023, date of current version 20 October 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3324401

RESEARCH ARTICLE

Two-Stage Cross-Domain Ocular Disease Recognition With Data Augmentation

QIONG WANG, ZHILIN GUO, JUN YAO, AND NAN YAN
Engineering and Technical College, Chengdu University of Technology, Leshan 614000, China
Corresponding author: Jun Yao (40516473@qq.com)

ABSTRACT Ophthalmic diseases afflict many people, and can even lead to irreversible blindness. Therefore, the search for effective early diagnosis methods has attracted the attention of many researchers and clinicians. At present, although there are some ways for the early screening of ophthalmic diseases, the early screening of fundus images based on deep learning is generally favored by the medical community due to its non-contact characteristic, non-invasive characteristic and high recognition accuracy. However, the generalization performance of a common model and cross-domain identification is usually weak due to different collection equipment, race, and patient conditions. Although the existing fundus image recognition technology has achieved some results, the effect is still in the cross-domain problem and is not satisfactory. This paper proposes a cross-domain retinal image recognition framework based on data augmentation and deep neural networks. Firstly, the ResNeXt101 model pretrained on the ImageNet dataset is selected as the base framework. The one-stage model is then trained in the source domain using this framework. Secondly, the model obtained from the first stage is further fine-tuned in the target domain to obtain the two-stage final model. During this process, various data augmentation techniques and focal loss are employed to improve the recognition performance. Experimental results demonstrate that by incorporating common data augmentation techniques and focal loss, the proposed framework achieved the following performance metrics in cross-domain experiments from train-site to on-site: a kappa score of 0.845, an F1 score of 0.923, and an AUC (Area Under the Curve) value of 0.974. In conclusion, the proposed method effectively addresses the issue of poor generalization in cross-domain early retinal screening and provides insights and directions for future related work.

INDEX TERMS Cross-domain, data enhancement, focal loss, fundus image, ocular disease, two-stage.

I. INTRODUCTION

Nowadays, Ophthalmic diseases have increasingly affected people's lives and tortured the psychology and spirit of patients and their relatives [4]. For example, the typical symptoms of glaucoma are visual field defect, eye pain, and nausea and vomiting in some patients. Some ophthalmic diseases can also cause irreversible blindness, such as macular degeneration. Therefore, it is essential to find an effective solution for early diagnosis. Although there were some ways to diagnose ophthalmic diseases, such as optical coherence tomography (OCT) images [5] and some clinical methods [6], [7] etc., early fundus screening is an economical and faster method to prevent blindness caused by ophthalmic diseases [8].

Recently, deep learning has made great progress in the field of computer vision, it has superior feature extraction ability when processing image data, especially in Euclidean space [11]. Furthermore, it is also gradually being applied in the field of medical images, such as lung cancer diagnosis [1], brain disease diagnosis [2], orthopedic disease screening [3], etc. Due to its powerful performance, fundus images based on deep learning are becoming increasingly popular in the clinical diagnosis of ophthalmic diseases.

Although the existing fundus image recognition based on depth learning has achieved good results, due to insufficient training data, lack of labels and other problems, the generalization performance is not ideal, which affects its application

The associate editor coordinating the review of this manuscript and approving it for publication was Inês Domingues.

in clinical practice, and the training of highly recognized neural network models requires a large-scale original data set for supervised learning. When the training set and test set are in different hospitals, the recognition effect is significantly reduced

In this paper, a framework based on fundus image screening and deep learning is proposed to improve the effect of cross-domain ophthalmic disease diagnosis.

First, we select a classic convolutional neural network (CNN), ResNeXt101 [12], as the recognition backbone that pretrained on the ImageNet dataset [15], [17]. Compared with other CNN networks, such as Inception [10], [19], VGG [18], ResNet [20], [21], etc., the ResNeXt101 network can get more features from fundus images to achieve higher accuracy and generalization performance.

In addition, fundus images are taken from different devices and environments due to the fact that the model parameters obtained by training a group of samples from a certain domain are only of high accuracy and generalization for this sample domain. Therefore, to reduce the gap between different domains, In the first stage, we utilize a larger number of samples from the source domain to train a model with good robustness. In the second stage, we fine-tune the model using a subset of samples from the target domain.

Finally, the general cross-entropy loss function is very effective for those datasets with balanced sample numbers. However, we often encounter an imbalance in the number of samples of different categories, which also leads to poor generalization performance of the final model and a larger gap between domains. Therefore, we apply focal loss [9] and data augmentation techniques to overcome this problem.

The main contributions of the proposed framework can be summarized as follows:

1) ResNeXt [12] model is chosen as the CNN backbone in our framework for better feature extraction;

2) In the first stage, a convolutional neural network is trained using a larger number of samples from the source domain;

3) In the second stage, transfer learning is applied to train the final model using a subset of samples from the target domain;

4) Some data augmentation methods are adopted to improve model generalization performance. And the focal loss function is applied during training to solve the class imbalance issue.

II. RELATED WORK

A. DATASET

Many systemic diseases, such as hypertension, arteriosclerosis, and diabetes, will produce changes in small blood vessels in various body parts. Because such small vessels can only be seen directly on the retina throughout the body, it is possible to check whether these diseases have caused vascular diseases through fundus screening [36]. Based on the results of this examination, we can refer to the course of disease research and treatment.

Fundus screening is a basic requirement for general practitioners or non-professional ophthalmologists. It can help general practitioners estimate the condition of systemic diseases and carry out specialized ophthalmic treatment as early as possible to prevent the aggravation of the disease and the risk of blindness [37].

The difference in the shape and position of the patient's eye structure has caused great trouble for clinicians in correctly diagnosing the disease [13], and this is also labor-intensive work. Clinicians find it difficult to quickly and accurately diagnose diseases through naked eye observation. With the help of fundus image screening based on deep learning, it is able to help them better complete their work [14].

Currently, the clinical effects of fundus image screening mainly include the following indicators: kappa is an indicator used for consistency tests, and it can also be used to measure the effect of classification. F1 score can be regarded as a weighted average of model accuracy and recall rate, and Area Under roc Curve(AUC) is a standard which used to measure the quality of recognition models.

Optical CoherenceTomography (OCT) is a new non-invasive imaging diagnostic technology. Currently, it is mainly used for image recognition, such as segmentation, detection and classification. He et al. proposed a method to section the surface and pathological changes of the retinal OCT layer [23], and Asgari et al. proposed a method using a decoder for each target class and solve the Drusen segmentation as a multi-task problem [24]. In order to investigate the symmetry between the eyes to better detect early ocular diseases, Marzieh Mokhtari et al. calculated the local Cup Disk Ratio (CDR) by combining fundus images and the OCT B-scan method [25]. A system containing neural network algorithms and data enhancement methods for the multi-category and multi label classification is proposed by Mehta et al. The system can classify the OCT images of four common retinopathy [26].

B. CROSS DOMAIN

Domain adaptation method can be regarded as a sub-direction of transfer learning. No matter what kind of domain adaptation technology, its goal is to reduce the distance between the target domain and the source domain. In recent years, popular domain adaptation methods have become the focus, such as Domain Adversarial Neural Networks (DANN) [27]. Tzeng et al. proposed an Adversarial Discriminative Domain Adaptation (ADDA) which is used to align the feature distribution between two different domains [34]. In addition, there are other methods to achieve alignment in pixel space through image conversion. PixelDA proposed a method to achieve cross-domain alignment by learning one-to-many mapping to synthesize images in the target domain [28]. A circular consistent adversarial domain adaptation (CyCADA) method is proposed by Hoffman, which converts images in the source domain into images in the target domain, and then combines the converted images with the target domain images for

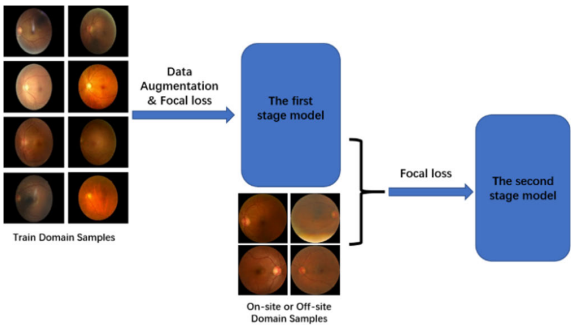

FIGURE 1. Two-stage training process of the proposed framework.

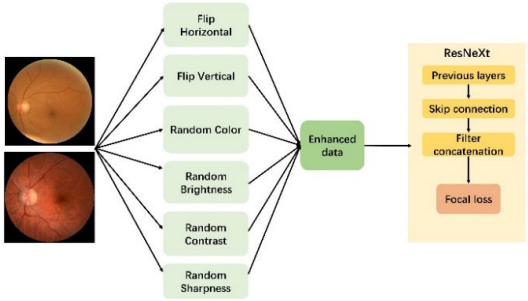

FIGURE 2. The training process of the proposed framework.

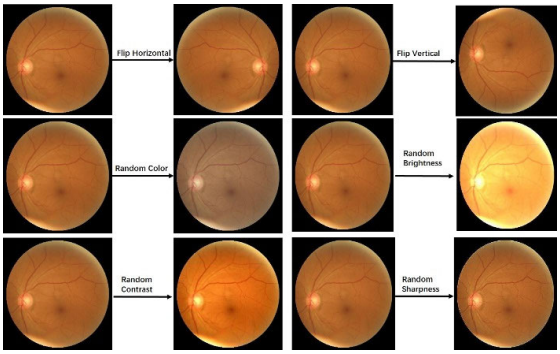

FIGURE 3. Data augmentation methods used in this paper.

training, So as to narrow the gap between different domains [29]. Tsai et al. used adversarial learning to achieve domain adaptation on semantic segmentation [30].

Although the datasets have increased, there is still a shortage of samples for the deep learning framework that requires a large scale of data. To overcome this problem, we can use some cross-domain methods, such as data enhancement, Generative Adversarial Networks (GAN) and other means.

Data enhancement is to perform some operations, such as sharpening and flipping on the image based on the original data, which can effectively expand the size of the dataset. The GAN aligns samples from different domains, narrows the gap between different domains, and indirectly and effectively uses limited datasets [35].

Typical medical applications include cross domain synthesis of medical images, including CT to PET, MR to CT, CT to MR, and T1, T2, FLAIR, etc., in MRI.

Cohen et al. proposed a novel system based on FCN and GAN networks, which generates virtual PET images from CT scanning, thus lowering expensive PET scanning costs [31].

Wolterink et al. used GAN to convert 2D brain MR image slices into 2D brain CT image slices, reducing the error in synthetic CT images caused by the dislocation between pairs of images [32].

Salman et al. proposes a method of synthesizing multi-contrast MRI images using conditional GAN, which realizes the mutual conversion and synthesis of T1 and T2 in MRI, and uses the adversarial loss function to maintain the middle and high frequency details of the image [33].

III. METHOD

In this paper, we propose a framework based on a convolutional neural network for fundus screening, as shown in FIGURE 1. and FIGURE 2.

The FIGURE 1. illustrates the two-stage training process. Firstly, the train dataset with a larger number of samples is utilized to conduct the first-stage model training. Subsequently, the target domain data with a smaller number of samples is employed to fine-tune the model trained in the first stage.

In FIGURE 2. The images on the left are the fundus images we have obtained. They are from the OIA-ODIR dataset. This dataset is one of the subsets of the OIA dataset, which contains 10,000 fundus images. The sampling population covers all age groups, more than 96% of whom are 30 to 80 years old. This data is mainly for multiple eye diseases. Then we enhanced the input data, including flip horizontal, flip vertical, random color, random brightness, random constraint, and random sharpness. and the enhanced data are sent to the ResNeXt network for training with the cross entropy loss function is replaced to focal loss function, which finally achieved good results. We discuss the framework in detail in the following subsections.

A. DATA AUGMENTATION

As shown in FIGURE 3, we have enhanced the data to be sent over the network. For instance, the original images are horizontally flipped, vertically flipped, subjected to random color modifications, random brightness adjustments, random contrast changes, and random sharpening.

B. ResNeXt

ResNeXt network uses the bottleneck structure in each branch. First, use 1×1 convolution to reduce the dimension and reduce the number of channels in the feature map. Then, use grouped convolutions to extract features. Finally, use 1×1 convolution to increase the dimension and restore the number of channels in the feature map. Each block in ResNeXt can be shown in FIGURE 4.

Each block in ResNeXt can be represented as follows

$$y = x + \sum_{i=1}^{C} T_i(x) \tag{1}$$

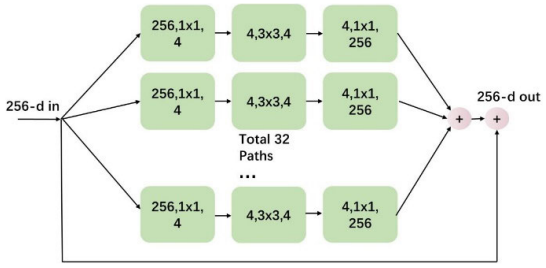

FIGURE 4. A ResNeXt block with cardinality equal to 32. Each layer displayed represents the input channel, filter size and output channel respectively.

In formula (1), C is the number of branches in the block, T_i is the subnet of each branch, and x is the shortcut connection.

ResNeXt involves the following related work: 1) ResNeXt uses multi-branch subnets for feature fusion, 2) uses packet convolution to control network parameters and floating point computation, unlike common compression methods that cost model accuracy, 3) ResNeXt uses multi branch packet convolution and other operations to further improve the model's expression ability while controlling the model parameters and floating point calculations, and 4) uses multi branch subnetworks for feature fusion. ResNeXt is different from the integrated learning method in that each branch is identical [12]

C. FOCAL LOSS

Usually, the cross entropy loss function is used to train the neural network, but this is more suitable for the case of balanced sample categories. When the number of sample categories varies greatly, the negative samples account for a large proportion of the total loss due to their large number, which makes the optimization direction of the model deviate from our expectations. Focal loss function is modified based on the standard cross entropy loss. The function can reduce the weight of easily classified samples and make the model focus on more difficult samples during training.

The specific form of focal loss is as follows:

$$L_{fl} = \begin{cases} -\left(1-\hat{p}\right)^{\gamma} \log\left(\hat{p}\right) & (if\ y=1) \\ -\hat{p}^{\gamma} \log\left(1-\hat{p}\right) & (if\ y=0) \end{cases} \tag{2}$$

There y is label, which corresponds to 0, 1 in the binary classification. where

$$p_t = \begin{cases} \hat{p} & (if\ y=1) \\ 1-\hat{p} & otherwise \end{cases} \tag{3}$$

Focal loss expression will unify into one expression:

$$L_{fl} = -\left(1-p_t\right)^{\gamma} \log\left(p_t\right) \tag{4}$$

In this expression, p_t represents the distance from category y. The closer p_t is to 1, the closer the sample is to ground truth. p_t also reflects the difficulty of classification. The larger p_t is, the higher the confidence level of classification is, and the easier the samples are to distinguish; The smaller p_t is, the lower the confidence level of classification, the more difficult is the samples are to distinguish. Therefore, the focus loss is to increase the weight of hard samples in the loss function and reduce the weight of soft samples, so that the loss function pays more attention to hard samples, thus improving the overall classification accuracy. $\gamma(>0)$ is an adjustable factor. Its function is to adjust the steepness of the weight curve in an exponential manner [9].

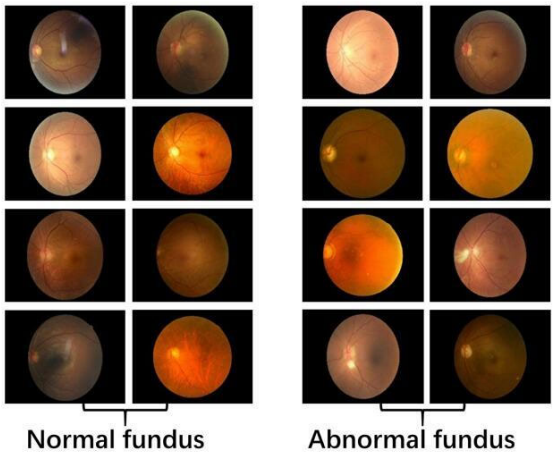

FIGURE 5. Sample images from two categories. The normal fundus on the left of the two types of images represents positive samples, and the other fundus on the right represents negative samples.

IV. EXPERIMENTS

The dataset OIA-ODIR used in this paper is a binocular fundus image dataset that can be used to detect various types of diseases. Its sample data is from a private clinical fundus database with more than 1.6 million images. These fundus images are collected from 487 clinical hospitals in 26 provinces of China. OIA-ODIR selected 10000 left and right fundus images of 5000 patients. In our experiments, we only used images of the left eye. All images are divided into normal fundus and abnormal fundus as shown in FIGURE 5. Three ophthalmologists with more than 2 years of clinical experience in ophthalmology and three doctors with more than 10 years of clinical experience in ophthalmology completed the annotation of the dataset within 10 months. In this series of experiments, we selected 3500 left-eye images in the train-site domain, 1000 left-eye images in the on-site domain, and 500 left-eye images in the off-site domain as training sets and test sets, respectively, to evaluate our method. During data preprocessing, we define the samples labeled normal fundus as positive samples and the others as negative samples. Their proportions are shown in TABLE 1. Before training, we adjust the non-RGB images in the data set to RGB images, uniformly adjust the pixels to 384 $\times$ 384 according to bilinear interpolation, and finally, normalize the pixels.

A. TRAINING DETAILS

Before training, we select resnext101, the pre-training model of the ImageNet dataset, as the backbone of the network

TABLE 1. The proportion of fundus images in training and testing datasets.

Labels	N(normal fundus)	A(abnormal fundus)
Training case	1578	1922
On-site testing cases	430	570
Off-site testing cases	227	273
All cases	**2235**	**2765**

model. the optimization function is Adaptive Motion Estimation (Adm). In the second stage of training, the on-site and off-site sets are used for fine-tuning, with a sample size of 50% of the total samples, which is 500 and 250 samples respectively. During training, we set epoch = 30, and the learning rate = 0.0001. When focal loss function is selected as the loss function, the hyper-parameter γ set to 0.25 to balance positive and negative samples. Label smoothing is used to reduce overfitting. The proposed method is implemented with PaddlePaddle of Baidu, network training using the v100 GPU with 32 GB memory of the Baidu AI Studio platform.

B. EXPERIMENTAL INDICATORS

In order to know how much of the evaluation result of a diagnostic test is due to opportunity factors, the kappa score is frequently used in clinical medicine to measure it. Its calculation method is as follows:

$$\text{KappaScore} = \frac{\text{Observationcr} - \text{Opportunitycr}}{100\% - \text{Opportunitycr}} \tag{5}$$

where cr represents compliance rate

F-1 score is an indicator used to measure the accuracy of binary classification model in statistics. It is defined as the harmonic mean of accuracy rate and recall rate, with a maximum of 1 and a minimum of 0. The expression is shown as follow:

$$\text{F1score} = 2 \cdot \frac{\text{accuracy} \cdot \text{recall}}{\text{accuracy} + \text{recall}} \tag{6}$$

AUC (Area Under Curve) is the area under the ROC curve, which is a performance indicator to measure the quality of classifier. The closer the AUC value is to 1.0, the better the reliability of the detection method is.

The final score is the average of kappa score, F-1 score and AUC value. We use it as a comprehensive evaluation indicator to declare the result of the method.

C. EXPERIMENTAL RESULTS

Through a series of experiments, we finally obtained the following results.

In TABLE 2, the first column is the distribution of training sets and test sets, where ''da'' represents data augmentation and ''fl'' represents focal loss. When only train-site is used as the training set, off-site is used as the test set, all indicators are the lowest except AUC Value. The effect can be improved to some extent by adding data augmentation or using the focal loss function during training. When data augmentation and focal loss are used simultaneously, the score is the highest, which just indicates that our method can effectively improve the recognition effect from train-site domain to off-site domain.

TABLE 2. The results obtained when the train-site is the training set and off-site is the test set.

	Kappa Score	F-1 Score	AUC Value	Final Score
Train-site(train)/off-site(test)	0.272	0.636	0.766	0.551
Train-site+da/off-site	0.392	0.696	0.764	0.617
Train-site+fl/off-site	0.376	0.688	0.765	0.609
Train-site+da+fl/off-site	**0.404**	**0.702**	**0.781**	**0.629**

TABLE 3. The results obtained when train-site is used as the training set and on-site is the test set.

	Kappa Score	F-1 Score	AUC Value	Final Score
Train-site(train)/on-site(test)	0.313	0.656	0.802	0.591
Train-site+da/on-site	**0.497**	**0.748**	0.821	**0.688**
Train-site+fl/on-site	0.430	0.715	0.797	0.648
Train-site+da+fl/on-site	0.483	0.742	**0.826**	0.684

TABLE 4. The results obtained with different training sets and test sets.

	Kappa Score	F-1 Score	AUC Value	Final Score
On-site(train)/train-site(test)	0.393	0.696	0.776	0.622
On-site+da+fl/train-site	**0.408**	**0.704**	**0.782**	**0.631**
On-site/off-site	0.364	0.682	0.738	0.594
On-site+da+fl/train-site	**0.388**	**0.694**	**0.769**	**0.617**
Off-site/train-site	0.334	0.667	0.733	0.578
Off-site+da+fl/train-site	**0.346**	**0.673**	**0.748**	**0.589**
Off-site/on-site	0.341	0.671	0.739	0.584
Off-site+da+fl/on-site	**0.368**	**0.684**	**0.764**	**0.605**

TABLE 3 shows the results when train-site is used as training set on-site is used as test set. Obviously, using da and fl can effectively improve the recognition results for cross-domain. Each corresponding score in TABLE 3 better than TABLE 2 due to on-site domain is closer to train-site domain than off_site domain. In TABLE 3, the use of focal loss leads to lower scores compared to not using it. This is due to the off-site domain having a larger disparity between positive and negative samples relative to the on-site domain.

TABLE 4 indicate that the model has been improved after using data augmentation and focal loss function. Compared with train-site domain to on-site domain or off-site domain, this improvement effect is not obvious, because the number of samples in on-site domain and off-site domain is much less than train-site domain. The insufficient number of samples and uneven distribution will inevitably lead to a poor recognition effect, which is also to be improved in the future.

TABLE 5 provides a comparison of the scores between the first-stage model and the second-stage model. In the second-stage model data, the terms ''on-site'' or ''off-site''

TABLE 5. Comparison of two-stage training data.

	Kappa Score	F-1 Score	AUC Value	Final Score
One-stage: train-site/on-site	0.388	0.694	0.789	0.624
One-stage:on-site/on-site	0.669	0.835	0.916	0.807
Two-stage(on-site): train-site/on-site	0.711	0.856	0.942	0.836
Two-stage(on-site): train-site/off-site	0.364	0.682	0.748	0.598
Two-stage(on-site+da): train-site/on-site(ours)	**0.845**	**0.923**	**0.974**	**0.914**
Two-stage(on-site+da): train-site/off-site	0.400	0.700	0.775	0.625
One-stage: train-site/off-site	0.382	0.691	0.781	0.618
One-stage:off-site/off-site	0.592	0.796	0.879	0.756
Two-stage(off-site): train-site/on-site	0.391	0.695	0.795	0.627
Two-stage(off-site): train-site/off-site	0.656	0.828	0.927	0.804
Two-stage(off-site+da): train-site/on-site	0.335	0.668	0.750	0.584
Two-stage(off-site+da): train-site/off-site(ours)	**0.792**	**0.896**	**0.961**	**0.883**

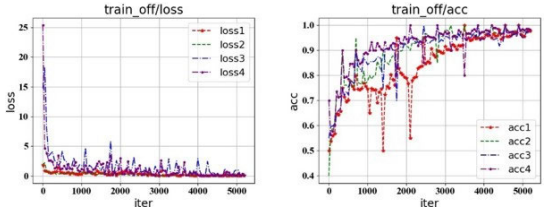

FIGURE 6. The loss and accuracy of training with train-site is the train set and off-site is the test set.

in parentheses indicate the use of transfer learning in a specific domain. It can be observed that the performance of the second-stage model exhibits significant advantages when transferred to the target domain. However, for non-transfer domains, the improvement is not significant and in some cases, even shows a decrease in performance. This is because the image knowledge learned in the first stage becomes more comprehensive after cross-domain adaptation, resulting in a more stable and reliable model.

The above-left figure shows the trend of loss change for the four training situations in TABLE 2. It can be seen from the observation that with the increase of the number of samples, the loss value shows an obvious downward trend. The curve loss4 representing our core method decays rapidly with the increase of sample numbe. The acc4 curve in the right figure shows that the final accuracy tends to the maximum value except for the large deviation of individual samples.

FIGURE 7. shows the trend of loss and acc when train-site is used as the training set on-site is used as the test set. They are very close to the trend of FIGURE 6. The difference is that the loss in FIGURE 7 decays faster and the acc reaches the peak using a shorter time, which is also because the gap between the train-site domain and the on-site domain is

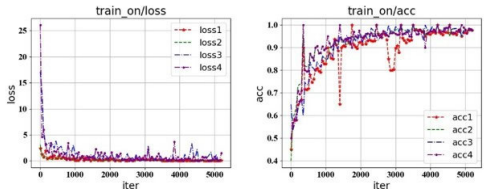

FIGURE 7. The loss and accuracy of training with train-site is the train set and on-site is the test set.

shorter than that between the train-site domain and the off-site domain.

V. CONCLUSION AND FUTURE WORK

To enhance the performance of cross-domain recognition in ophthalmic disease diagnosis, we initially opted for ViT as the backbone network, however, under the same conditions, the final score achieved was only about 0.6, significantly lower than the score of around 0.9 achieved by the proposed two-stage method in this paper. We believe this is due to the lack of prior knowledge in ViT compared to CNN.

The experimental results indicate that our method has better adaptability than the mainstream neural networks for cross-domain fundus screening. It can better help clinicians diagnose, and thus contribute to solving the pain of patients and their families.

However, the drawback of our two-stage method is the complexity it introduces to the training process, When the target domain changes, it requires the model to be readjusted once again to adapt to the new target domain, which poses a challenge for subsequent work. At the same time, GAN, as the mainstream framework of cross-domain methods at the present stage, has not been able to achieve cross-domain diagnosis in combination with medical images due to its unpredictability. In the future, we will focus on more datasets to carry out follow-up experiments and verify the effectiveness of our framework. In addition, we look forward to exploring more ideas and methods combined with medical image cross-domain recognition based on the advantages of the generation network

ACKNOWLEDGMENT

Author Contributions: Qiong Wang: conducting experiments and writing the article; Zhilin Guo: collecting data and writing the article; Jun Yao: data collection and result verification; and Nan Yan: analyzing the experimental results and improving the experiment.

REFERENCES

[1] W. Sun, B. Zheng, and W. Qian, "Computer aided lung cancer diagnosis with deep learning algorithms," *Proc. SPIE*, vol. 9785, pp. 241–248, Mar. 2016.

[2] R. Li, W. Zhang, H. I. Suk, L. Wang, J. Li, D. Shen, and S. Ji, "Deep learning based imaging data completion for improved brain disease diagnosis," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2014, pp. 305–312.

[3] S. W. Chung, S. S. Han, J. W. Lee, K.-S. Oh, N. R. Kim, J. P. Yoon, J. Y. Kim, S. H. Moon, J. Kwon, H.-J. Lee, Y.-M. Noh, and Y. Kim, "Automated detection and classification of the proximal humerus fracture by using deep learning algorithm," *Acta Orthopaedica*, vol. 89, no. 4, pp. 468–473, Jul. 2018.

[4] A. P. Adamis, L. P. Aiello, and R. A. D'Amato, "Angiogenesis and ophthalmic disease," *Angiogenesis*, vol. 3, pp. 3–14, Mar. 1999.
[5] D. Huang, E. A. Swanson, C. P. Lin, J. S. Schuman, W. G. Stinson, W. Chang, M. R. Hee, T. Flotte, K. Gregory, C. A. Puliafito, and J. G. Fujimoto, "Optical coherence tomography," *Science*, vol. 254, pp. 1178–1181, Nov. 1991.
[6] M. H. Freeman, "Ultrasonic pulse-echo techniques in ophthalmic examination and diagnosis," *Ultrasonics*, vol. 1, no. 3, pp. 152–160, Jul. 1963.
[7] I. Tsui, V. Franco-Cardenas, J.-P. Hubschman, and S. D. Schwartz, "Pediatric retinal conditions imaged by ultra wide field fluorescein angiography," *Ophthalmic Surg., Lasers Imag. Retina*, vol. 44, no. 1, pp. 59–67, Jan. 2013.
[8] B. Li, H. Chen, B. Zhang, M. Yuan, X. Jin, B. Lei, J. Xu, W. Gu, D. C. S. Wong, X. He, and H. Wang, "Development and evaluation of a deep learning model for the detection of multiple fundus diseases based on colour fundus photography," *Brit. J. Ophthalmol.*, vol. 106, pp. 1079–1086, Aug. 2022.
[9] T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, Jun. 2017, pp. 2980–2988.
[10] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 1–9.
[11] D. Shen, G. Wu, and H. I. Suk, "Deep learning in medical image analysis," *Annu. Rev. Biomed. Eng.*, vol. 19, pp. 221–248, Jun. 2017.
[12] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2017, pp. 1492–1500.
[13] H. Fu, Y. Xu, S. Lin, X. Zhang, D. W. K. Wong, J. Liu, A. F. Frangi, M. Baskaran, and T. Aung, "Segmentation and quantification for angle-closure glaucoma assessment in anterior segment OCT," *IEEE Trans. Med. Imag.*, vol. 36, no. 9, pp. 1930–1938, Sep. 2017.
[14] J. Son, J. Y. Shin, H. D. Kim, K.-H. Jung, K. H. Park, and S. J. Park, "Development and validation of deep learning models for screening multiple abnormal findings in retinal fundus images," *Ophthalmology*, vol. 127, no. 1, pp. 85–94, Jan. 2020.
[15] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, p. 25.
[16] R. Müller, S. Kornblith, and G. E. Hinton, "When does label smoothing help," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, p. 32.
[17] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.
[18] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.
[19] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-ResNet and the impact of residual connections on learning," in *Proc. AAAI Conf. Artif. Intell.*, 2017, vol. 31, no. 1, pp. 1–7.
[20] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[21] H. Xiao, K. Rasul, and R. Vollgraf, "Fashion-MNIST: A novel image dataset for benchmarking machine learning algorithms," 2017, *arXiv:1708.07747*.
[22] H. Caesar, J. Uijlings, and V. Ferrari, "COCO-stuff: Thing and stuff classes in context," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1209–1218.
[23] Y. He, A. Carass, Y. Liu, B. M. Jedynak, S. D. Solomon, S. Saidha, P. A. Calabresi, and J. L. Prince, "Fully convolutional boundary regression for retina OCT segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2019, pp. 120–128.
[24] R. Asgari, J. I. Orlando, S. Waldstein, F. Schlanitz, M. Baratsits, U. Schmidt-Erfurth, and H. Bogunovic, "Multiclass segmentation as multitask learning for drusen segmentation in retinal optical coherence tomography," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2019, pp. 192–200.
[25] M. Mokhtari, H. Rabbani, A. Mehri-Dehnavi, R. Kafieh, M.-R. Akhlaghi, M. Pourazizi, and L. Fang, "Local comparison of cup to disc ratio in right and left eyes based on fusion of color fundus images and OCT B-scans," *Inf. Fusion*, vol. 51, pp. 30–41, Nov. 2019.
[26] P. Mehta, A. Lee, C. Lee, M. Balazinska, and A. Rokem, "Multilabel multiclass classification of OCT images augmented with age, gender and visual acuity data," *bioRxiv*, May 2018, doi: 10.1101/316349.
[27] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, "Domain-adversarial training of neural networks," *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 2030–2096, 2016.
[28] K. Bousmalis, N. Silberman, D. Dohan, D. Erhan, and D. Krishnan, "Unsupervised pixel-level domain adaptation with generative adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2017, pp. 3722–3731.
[29] J. Hoffman, E. Tzeng, T. Park, J. Y. Zhu, P. Isola, K. Saenko, A. Efros, and T. Darrell, "CyCADA: Cycle-consistent adversarial domain adaptation," in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 1989–1998.
[30] Y. H. Tsai, W. C. Hung, S. Schulter, K. Sohn, M. H. Yang, and M. Chandraker, "Learning to adapt structured output space for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7472–7481.
[31] A. Ben-Cohen, E. Klang, S. P. Raskin, S. Soffer, S. Ben-Haim, E. Konen, M. M. Amitai, and H. Greenspan, "Cross-modality synthesis from CT to PET using FCN and GAN networks for improved automated lesion detection," *Eng. Appl. Artif. Intell.*, vol. 78, pp. 186–194, Feb. 2019.
[32] J. M. Wolterink, A. M. Dinkla, M. H. F. Savenije, P. R. Seevinck, C. A. van den Berg, and I. Isgum, "Deep MR to CT synthesis using unpaired data," in *Proc. Int. Workshop Simul. Synth. Med. Imag.* Cham, Switzerland: Springer, 2017, pp. 14–23.
[33] S. U. Dar, M. Yurt, L. Karacan, A. Erdem, E. Erdem, and T. Çukur, "Image synthesis in multi-contrast MRI with conditional generative adversarial networks," *IEEE Trans. Med. Imag.*, vol. 38, no. 10, pp. 2375–2388, Oct. 2019.
[34] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, "Adversarial discriminative domain adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2962–2971.
[35] H.-K. Hsu, C.-H. Yao, Y.-H. Tsai, W.-C. Hung, H.-Y. Tseng, M. Singh, and M.-H. Yang, "Progressive domain adaptation for object detection," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2020, pp. 738–746.
[36] T. Bek, "Regional morphology and pathophysiology of retinal vascular disease," *Prog. Retinal Eye Res.*, vol. 36, pp. 247–259, Sep. 2013.
[37] A. B. Jain, V. J. Prakash, and M. Bhende, "Techniques of fundus imaging," *Med. Vis. Res. Found*, vol. 33, p. 100, Jun. 2015.

QIONG WANG received the B.S. and M.S. degrees from Xidian University, Xi'an, China, in 2003 and 2006, respectively. He is currently a Teacher in computer science with the Engineering and Technical College, Chengdu University of Technology, China. His primary research interests include deep learning and microwave electronic technology. He has published a few journals in the related deep learning and microwave electronic technology fields.

ZHILIN GUO is currently pursuing the B.S. degree in data science and big data technology from the Engineering and Technical College, Chengdu University of Technology. His primary research interests include data analysis, machine learning, and computer vision. He has published two papers in related fields.

JUN YAO received the B.S. degree from the Southwest University of Science and Technology, Sichuan, China, in 2005, and the M.S. degree from the Chengdu University of Technology, Sichuan, in 2010. He is currently a Teacher in computer science with the Engineering and Technical College, Chengdu University of Technology. His primary research interests include machine learning and computer vision. He has published more than ten journals in the related machine learning and computer vision fields.

NAN YAN received the B.S. degree from Yunnan University, Yunnan, China, in 1999, and the M.S. degree from the Kunming University of Science and Technology, Kunming, China, in 2006. He is currently a Teacher in big data with the Engineering and Technical College, Chengdu University of Technology, China. His primary research interests include big data and machine learning. He has published more than five journals in the related machine learning and big data fields.