

URNet: System for recommending referrals for community screening of diabetic retinopathy based on deep learning

Kun Yang^{1,2}, Yufei Lu¹, Linyan Xue^{1,2}, Yueting Yang¹, Shilong Chang¹ and Chuanqing Zhou³ 

¹College of Quality and Technical Supervision, Hebei University, Baoding 071002, China; ²Hebei Technology Innovation Center for Lightweight of New Energy Vehicle Power System, Baoding 071002, China; ³College of Medical Instruments, Shanghai University of Medicine and Health Sciences, Shanghai 201318, China
Corresponding author: Chuanqing Zhou. Email: zhoucq@sumhs.edu.cn

Impact Statement

Diabetic retinopathy (DR) is the complication of diabetes and major cause of visual impairment in working adults, it is important to detect DR early and then to intervene with effective management strategies. We proposed URNet, a two-stage deep learning-based computer-aided classification algorithm for community screening of diabetes retinopathy. Different from the traditional direct image classification and prediction algorithm, URNet can effectively reduce the interference of nonpathological information in the images to the network by reconstructing low-quality images, thus improving the accuracy of the classification network. Abnormal exposure and blurring usually exist in the fundus images in community screening, particularly collected by personnel who is not well trained, and/or in the various illuminating conditions. Those low-quality images will interfere with the judgment of computer aided system. Thus, we hope that the URNet proposed in this article can be helpful for community screening of diabetes retinopathy and other fundus diseases as well.

Abstract

Diabetic retinopathy (DR) will cause blindness if the detection and treatment are not carried out in the early stages. To create an effective treatment strategy, the severity of the disease must first be divided into referral-warranted diabetic retinopathy (RWDR) and non-referral diabetic retinopathy (NRDR). However, there are usually no sufficient fundus examinations due to lack of professional service in the communities, particularly in the developing countries. In this study, we introduce UGAN_Resnet_CBAM (URNet; UGAN is a generative adversarial network that uses UNet for feature extraction), a two-stage end-to-end deep learning technique for the automatic detection of diabetic retinopathy. The characteristics of DDR fundus data set were used to design an adaptive image preprocessing module in the first stage. Gradient-weighted Class Activation Mapping (Grad-CAM) and t-distribution and stochastic neighbor embedding (t-SNE) were used as the evaluation indices to analyze the preprocessing results. In the second stage, we enhanced the performance of the Resnet50 network by integrating the convolutional block attention module (CBAM). The outcomes demonstrate that our proposed solution outperformed other current structures, achieving 94.5% and 94.4% precisions, and 96.2% and 91.9% recall for NRDR and RWDR, respectively.

Keywords: Referral-warranted diabetic retinopathy, deep learning, ResNet50, CBAM, retinal image, non-referral diabetic retinopathy

Experimental Biology and Medicine 2023; 248: 909–921. DOI: 10.1177/15353702231171898

Introduction

Diabetic retinopathy (DR) is a complication of diabetes and a major cause of visual impairment in working adults, with one in three people with diabetes worldwide suffering from DR.^{1,2} The blood viscosity increases with the amount of glucose in the blood, especially in DR patients. And the high viscosity of the blood causes fluid to leak into the tissues around the retina, thus threatening vision. As the disease progresses gradually, it leads to microaneurysms (MA), hemorrhage (HM), exudation, and neovascularization in the retina, which

are typical features of DR lesions.³ DR is difficult to cure in its severe stages. Therefore, it is important to detect DR early and then to intervene with effective management strategies.

In fact, there is no sufficient fundus examination due to lack of professional service in the communities, particularly in the developing countries.⁴ To overcome these limitations, computer-aided automated detection and diagnosis of DR has received increasing attentions. Automated detection can assist or replace ophthalmologists in screening the collected fundus images and grading or classifying DR.⁵ Yet, fundus photography is usually performed by a personnel who is

not well trained, and/or in various illuminating conditions. Therefore, abnormal exposure and blurring usually exist in the collected images, which will interfere with the judgment of the computer-aided system. In addition, it more often happens in practice that the collected images in the severe DR category of data set are more blurry and abnormally exposed than those in normal category. The abnormal exposure and blurring will be probably regarded as the characteristics of severe DR during the training process of the algorithm and result in misdiagnosis.

High-quality fundus images are important for the investigation and diagnosis of fundus diseases.^{6,7} To improve the image quality, Ashiba *et al.*⁸ presented a new algorithm, which decomposed the image into sub-bands by using the additive wavelet transform. The images with better visual details were obtained by enhancing the details of images in each sub-band. Xiong *et al.*⁹ proposed an image pixel extraction method combining Mahalanobis distance discrimination and global spatial entropy-based contrast enhancement. Experimental results demonstrated that the proposed method could perform well on illumination evenness, contrast enhancement, and color preservation. Although these methods belong to classic image processing algorithms, they will also probably facilitate the analysis and detection of computer-aided systems.

In the early stages of computer-aided detection (CAD) system, multilayer neural networks have been shown to be able to couple arbitrary nonlinear functions, and through some configurations, the networks can automatically extract features that previously require human participation. For example, classical deep learning algorithms such as multipath convolutional neural network (M-CNN) and Resnet50 are used to extract features of DR. Then, random forest classifier and support vector machine classifier are used to classify the level of fundus lesions according to the extracted color, texture, and other features.^{10,11} In this step-by-step learning, the feature extractor and classifier can obtain their own optimal solution through separate training but cannot obtain a globally optimal solution. With the development of deep learning neural networks, the end-to-end learning method makes the whole learning process unnecessary to divide sub-problems manually. Compared with step-by-step learning, the end-to-end deep learning method does not need to label data manually before executing each independent learning task, but directly learn the mapping from the original data to the required output. Luo *et al.*¹² proposed a convolutional network that fuses multi-view fundus images to make full use of the pathological features of the retina. The attention mechanism module was added to the network to increase the attention onto the important features in the fundus image. The important channels in the image were given larger weight values by the network, which allowed them to extract features more efficiently, thereby improving network performance. But the number of their network layers was increased. Although increasing the number of neural network layers is a common method to improve the accuracy of classification, it will lead to gradient disappearance and overfitting. To solve this problem, Al-Moosawi and Khudayer¹³ proposed a deep learning algorithm based on Resnet34 degree grading model for DR. They used residual network structure to balance the

relationship between network complexity and computing cost. However, ResNet-34 does not perform well for large-scale data sets. In addition, they used single neural network that can only solve a single task. The cascade of two neural networks can more effectively solve multiple tasks. Alyoubi *et al.*¹⁴ proposed a fully automatic diagnosis system consisting of two deep learning models. The first model (CNN512) used the entire image as the input of the CNN model to classify the five stages of DR, and the second model (YOLOv3) was used to detect and locate the DR lesions based on the results of the first model. However, through comparative experiments, Tseng *et al.*¹⁵ found that five classification methods would reduce the accuracy of detection, which was not conducive to rapid screening. Meanwhile, compared with the five classifications, the two classifications are more conducive to referral recommendations.

Although there are encouraging achievements in this field, no algorithm is particularly suitable for processing images obtained in community screening. Hence, a deep learning algorithm was proposed based on Resnet34 degree grading model for DR. In this case, when these images are used for training directly, the noise of the image can possibly be extracted and learned as the pathological feature of DR, which most likely leads to the misjudgment of DR. For example, the collected fundus images include defocused images, cataractous images, and abnormal exposure images, as shown in Figure 1. There are usually higher proportions of these low-quality images in referral-warranted diabetic retinopathy (RWDR) data set than that in non-referral diabetic retinopathy (NRDR) data set.¹⁶ Therefore, features such as low image brightness and blurring will be likely mistaken as the pathological features of DR in the process of network training.¹⁷ The existence of abnormal images will affect the accuracy of network training and judgment, which is not addressed in the single neural network classification algorithms.

In this work, we proposed a network model URNet, a new DR detection CAD system based on a generative adversarial network-convolutional neural network (GAN-CNN) two-stage network structure. In this URNet, the anomalous exposure of the image is repaired through the first stage's preprocessing operation, and the detailed information of the fundus image is better displayed, so that even with subpar data image quality, the second stage's classification network can still produce accurate classification results. We verified the feasibility of this scheme by testing it on public data sets. The main advantages over other cutting-edge algorithms are as follows: (1) the staged processing scheme of URNet can better cope with abnormal images in community screening, (2) we bypass lesion detection and directly train the classifier for DR referral to lighten the burden of data labeling, and (3) a convolutional block attention module (CBAM)¹⁸ is added between the first and last convolutional layers to improve the classification accuracy.

Materials and methods

The data sets and preparation techniques used in this work are presented in this section. The two network stages URNet is shown in Figure 2. To create a data set for the first-stage network and train UGAN, the fundus images with clear

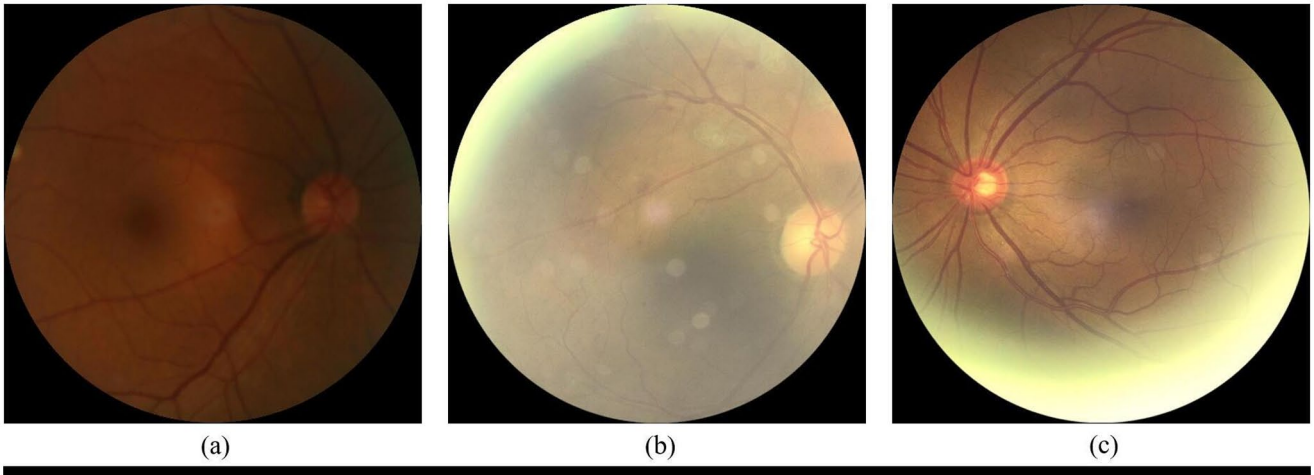


Figure 1. Low-quality fundus images in the data set, such as (a) defocused image, (b) cataractous image, and (c) abnormal exposure image.

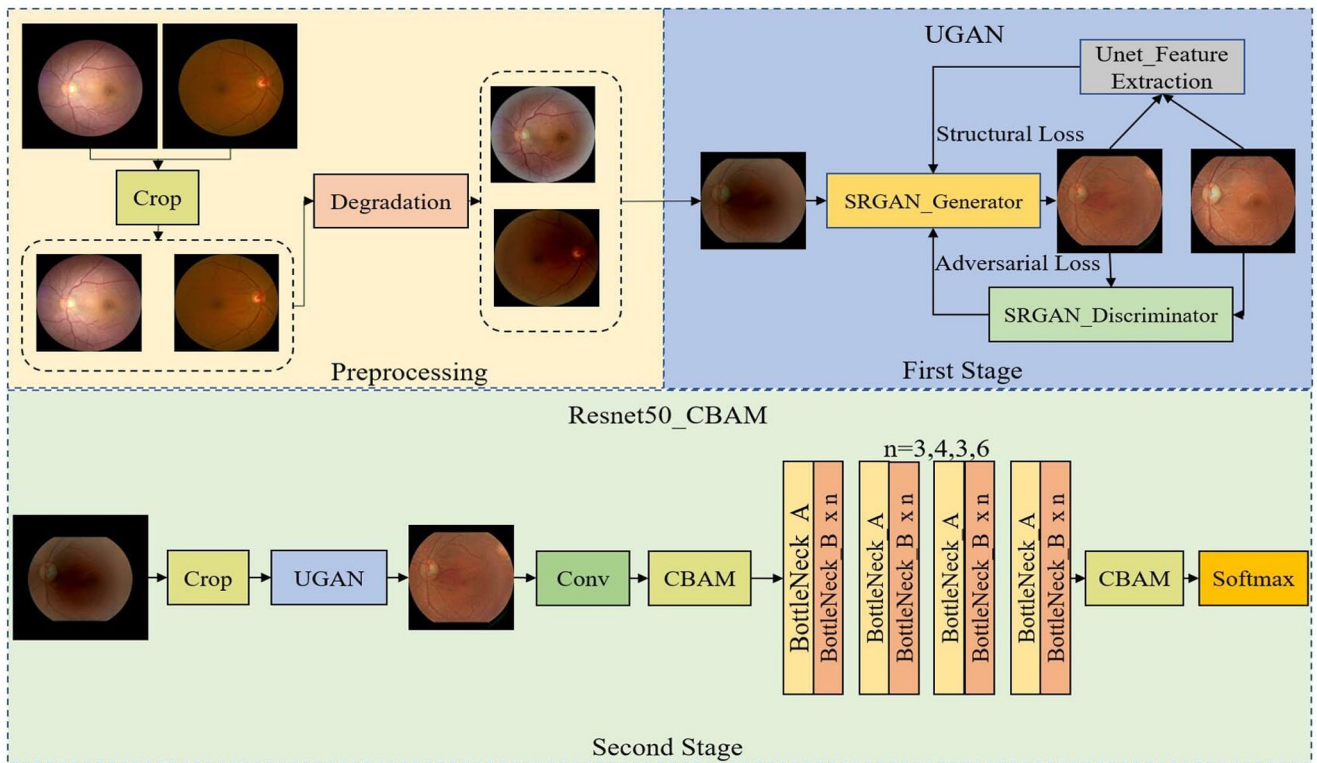


Figure 2. The URNet structure. The first stage is UGAN for image enhancement. The second stage is Resnet50_CBAM for image classification.

structure and good exposure quality were first screened out in the DDR data set. Then, the selected clear images were processed with the Crop module and the Degrad module sequentially to generate degraded images. The cropped clear images are used as labels and paired with degraded images to feed the network for training.

Data sets

This work utilized a publicly available DDR data set, as demonstrated in Table 1. The 13,673 fundus images in the DDR data set, which were taken at a 45-degree field of view (FOV), are classified into five stages according to the International

Clinical Diabetic Retinopathy Severity (ICDRS).¹⁹ The data set distribution is unbalanced since there are much more data in stages 0 and 2 than those in stages 1, 3, and 4. Stages 0 and 1 are categorized as NRDR; stages 2, 3, and 4 are categorized as RWDR; whereas stage 5 is categorized as other. As a result, the re-divided data set’s categorization distribution is fairly balanced.

Data preprocessing

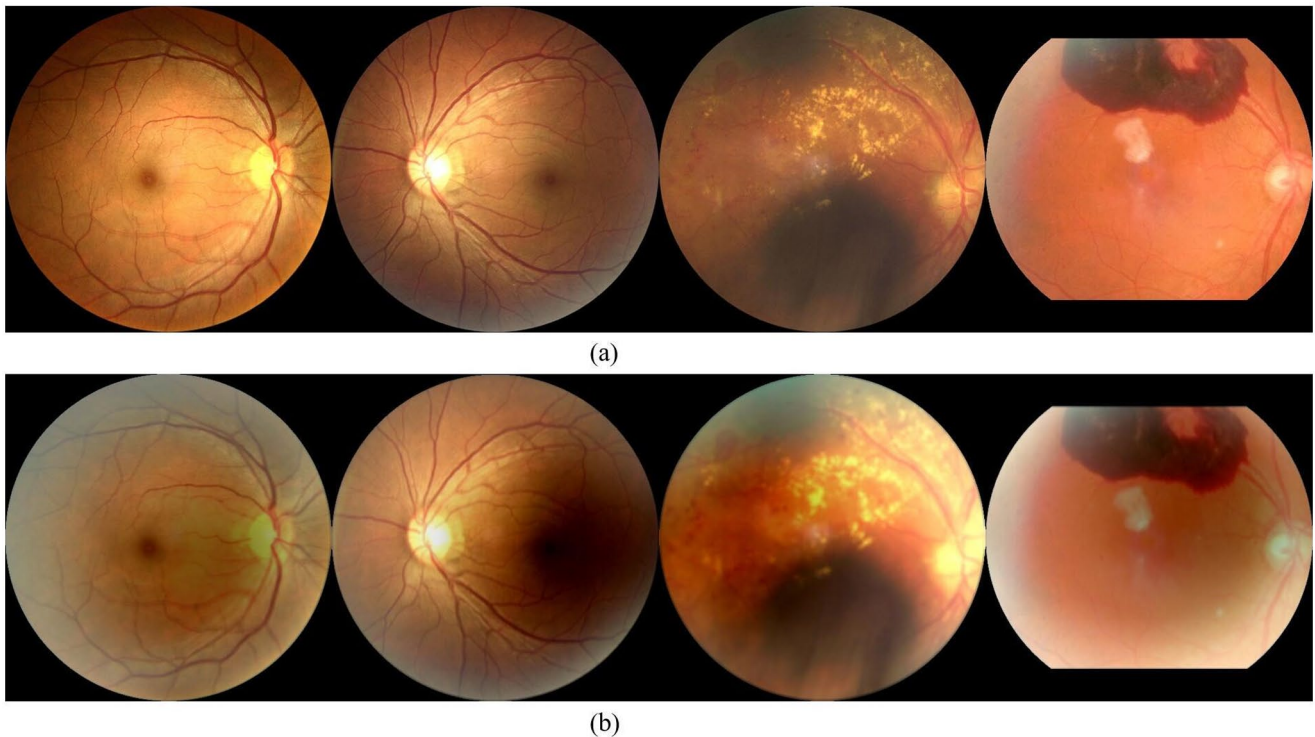
To extract more useful information from the input retinal image, data preprocessing is first conducted via cropping and degradation.

Table 1. The DDR data sets.

DDR	Stage	Train/ images	Valid/ images	Test/ images	Total
NRDR	0	3133	1253	1880	6266
	1	315	126	189	630
	Total	3448	1379	2069	6896
RWDR	2	2238	895	1344	4477
	3	118	47	71	236
	4	456	182	275	913
	Total	2812	1124	1690	5626
Other	5	575	230	346	1151
Total		6835	2733	4105	13,673

DDR is the name of a publicly available dataset (<https://github.com/nkicls/OIA>); NRDR: non-referral diabetic retinopathy; RWDR: referral-warranted diabetic retinopathy.

Stages 0 and 1 are categorized as NRDR; stages 2, 3, and 4 are categorized as RWDR, whereas stage 5 is categorized as other.

**Figure 3.** (a) Label images. (b) Degraded images.

Crop. To reduce the invalid information in the image, we cropped along the edge of the circular area of the fundus image and obtained a rectangular image with black edges removed. The Crop module cuts the input image based on pixel values and edge features, and outputs a fundus image without extra black filling, as shown in Figure 2(a).

Degradation. To acquire the UGAN data set, we first selected 1000 clear fundus images in the DDR public data set as labels, and then used the degradation algorithm²⁰ to add noise to the label images to obtain the corresponding noisy images, as shown in Figure 3. The generated noisy images and label images are made as data sets to train the UGAN network. We used contrast, brightness, and saturation interference to model the light degradation, and uses Gaussian blur to simulate image blur caused by defocus.

Given a label image x , its degraded paired image x' with light transmission disturbance is defined as:

$$x'_L = \text{clip}(\alpha(J_{(a,b)} \cdot G_L(r_L, \sigma_L) + x) + \beta; s) \quad (1)$$

$$x'_B = x \cdot G_B(r_B, \sigma_B) + n \quad (2)$$

$$x' = x'_L \cdot x'_B \quad (3)$$

where α , β , and $\text{clip}(;s)$ refer to the contrast factor, brightness, and saturation interference, respectively. J is defined as an illumination bias to be over-/under-illuminated at a panel centered at (a,b) with a radius of r . $G(r, \sigma)$ is a Gaussian filter, where r is radius, σ is spatial constant, and n is random Gaussian noise.

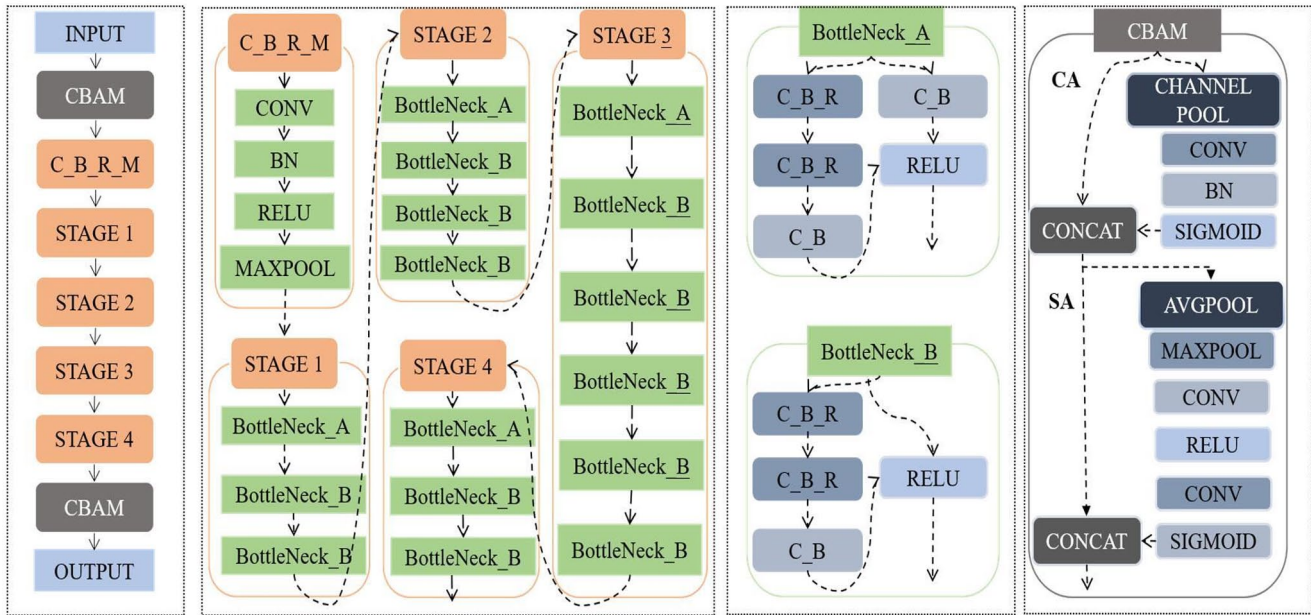


Figure 4. Main framework based on Resnet50 for RWDR and NRDR binary classification.

First stage – image enhancement based on GAN network. Based on the SRGAN image super-resolution algorithm, the Unet²¹ network is used to replace the original VGG (Visual Geometry Group)²² network for structural feature learning, as shown in Figure 2(b). The improved network UGAN conducts supervised learning with clear images as ground truth. The multi-scale fusion-based encoder–decoder structure²³ can learn more texture feature information and richer vascular structure information from noisy images.

In the encoder process, the multi-scale features of the image are obtained through continuous downsampling.²⁴ In the decoder process, the rich information in the encoder is connected to the layer corresponding to the decoder by skip connection, which efficiently avoids the image distortion caused by upsampling.

Second stage – classification network

The output of the image from the first stage is then input into the Resnet50-CBAM classification residual network,²⁵ as shown in Figure 2(c). The architecture of the network is shown in Figure 4, which is mainly composed of three parts: CBAM convolution attention mechanism, C_B_R_M (convolution layers [CONV], batch normalization layer, RELU layer, and max-pooling layer) convolution pooling module, and STAGE 1~4.

CBAM. CBAM is a lightweight attention mechanism that combines channel attention (CA) and spatial attention (SA), as shown in Figure 4. The CA module compresses the feature map on the spatial dimension, and generates two different spatial informations based on the width and height of the input feature map through max pooling and average pooling.²⁶ Then CA module input the spatial information of feature map into a shared network, which is composed of multilayer perceptron (MLP), and output the features of CA through element summation. Averaging pooling operates

on each pixel of the feature map, while max pooling focuses only on the local pixel maximum of the feature map in back-propagation. The mathematical model of the CA mechanism can be expressed as:

$$M_c(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) \quad (4)$$

where M_c is CA mechanism, F is feature map, and σ is sigmoid activation function.

Unlike CA, SA pays more attention to the location information of images. Through the CA, global max-pooling and global average-pooling operations are performed on the feature map based on its width and height, so as to obtain the attention feature in the spatial dimension. Then the spatial dimension of feature map has been changed from Height \times Width into 1×1 . After that, its dimension is reduced through CONV and RELU activation function. And then the feature map is upgraded to the original dimension after another convolution. Finally, the feature map is further processed with the sigmoid activation function and combined with the feature map of the CA output. At this stage, the recalibration of the feature map has been completed in the two dimensions of space and channel. The SA process can be expressed as following:

$$M_s(F) = \sigma(f^{7 \times 7} [AvgPool(F); MaxPool(F)]) \quad (5)$$

where M_s represents the SA mechanism, F represents the feature map, σ represents the sigmoid operation, 7×7 represents the convolutional kernel size, $AvgPool$ represents the global average pooling operations, and $MaxPool$ represents the global max pooling operations.

In fact, the pathological features only occupy small number of pixels in the DR image, and the backbone network has limited effective information on feature extraction. Therefore, after the CBAM convolutional attention module

is added to the backbone network, the channel information and spatial information of the small target are enhanced. So if we promote the network to learn more meaningful pathological information, the meaningless information is no longer transmitted downward, and detection accuracy can be improved.²⁷

C_B_R_M. The C_B_R_M architecture involves four main layers: CONV, batch normalization layer,²⁸ RELU layer,²⁹ and max-pooling layer. The CONV layer is to extract the features of the images with convolving filter, and the batch normalization layer is to normalize the inputs of a layer during training to increase the training speed and regularize the CNN. The function of the RELU layer is to add nonlinear factors, mapping features to high-dimensional nonlinear domain for interpretation, which solve problems that cannot be solved by linear models. At last, the max-pooling layer is used to reduce the dimension of the feature maps.

Stage 1~4. Stage 1~4 are composed of BottleNecks. The number of BottleNecks in every stage is 3, 4, 3, 6, respectively. The BottleNeck's role is to extract the features of the images by convolving different filters. Every BottleNeck consists of three convolutions: 1×1 , 3×3 , and 1×1 . The function of the 1×1 convolutional layer is to reduce and restore the dimension, and 3×3 layer is used to reduce input/output dimension. BottleNeck is used to reduce feature dimensionality, the number of layers of feature maps, and the number of parameters, and thus to reduce the amount of calculation.

Transfer learning

It is a common approach in deep learning to use pre-trained models as the starting point for new models in computer vision tasks and natural language processing tasks. Usually these pre-trained models consume huge time and computing resources when developing neural networks. Transfer learning can transfer the acquired powerful skills to related problems.³⁰ Using transfer learning can reduce training time and tuning effort for many hyperparameters. It transfers the knowledge from a pretrained network that was trained on large scale data set to a target network in which limited training data are available. In this article, the pretrained weights trained on the ImageNet data set³¹ by Resnet50 were used to initialize the parameters of the network layer other than the classification layer and the CBAM layer. we setted the random initialization parameters for the classification layer and the CBAM layer, and then retrained all the network layer parameters with the DR data set.

Loss function design

Since the proposed UGAN image enhancement algorithm is to make the denoised image contain more texture information and structural features of the original image, this article introduced a new structural loss function into the network model, which fused the structure, brightness and contrast features extracted by the convolutional network in the training process of the network. As part of the objective function, it constrains the training process and guides the

learning direction of the network. By reducing the difference between the generated image and the target image on the feature layer during the training process, the generated image can have higher consistency with the target image. Hence the finally reconstructed image has higher definition. We constrained the classification algorithm using the cross-entropy loss function.

Adversarial loss. Adversarial loss function is a popular loss function in GANs used to fool the discriminative network by making the generation network produce more realistic results. This loss function helps the network to converge and get a clearer fundus image, which can be expressed as follows:

$$I_{GAN}^u = \sum_{n=1}^N -\log D_{\theta D}(G_{\theta G}(P^{LR})) \quad (6)$$

where $D_{\theta D}$ represents the generator, $G_{\theta G}$ represents the discriminator, p^{LR} represents the low-quality images, and N represents the total number of samples.

Structural loss function. To improve the peak signal-to-noise ratio (PSNR) value, we used the L2 loss function. But this cannot effectively improve the perceptual resolution of the image. To make the color and brightness of the degraded image closer to the label image, we added structure similarity (SSIM) loss function³² as a constraint condition for network learning. The combination of L2 loss function and SSIM loss function can better preserve high-frequency information while maintaining the same color and brightness information, which is of great significance for fundus image reconstruction. These two loss functions are deduced as follows:

$$L_{Unet}^{l2} = \frac{1}{N} \sum_{x \in X} (x^{HR} - x^{LR})^2 \quad (7)$$

$$L_{Unet}^{SSIM} = 1 - SSIM(LR, HR) \quad (8)$$

$$SSIM(LR, HR) = \frac{(2\mu_{LR}\mu_{HR} + (K_1L)^2)(2\sigma_{LRHR} + (K_2L)^2)}{(\mu_{LR}^2 + \mu_{HR}^2 + (K_1L)^2)(\sigma_{LR}^2 + \sigma_{HR}^2 + (K_2L)^2)} \quad (9)$$

$$\mu_x = \frac{1}{N} \sum_{i=1}^N x_i \quad (10)$$

$$\sigma_x = \left(\frac{1}{N-1} \sum_{i=1}^N (x_i - \mu_x)^2 \right)^{1/2} \quad (11)$$

where x represents the pixel value, N represents the number of pixels, HR represents the label image, LR represents the corresponding degraded image, $K_1 = 0.01$, $K_2 = 0.03$, $L = 255$, x_i represents the value of each pixel, μ is the average brightness of the image, σ represents the contrast which is the standard deviation of the pixel values.

Cross-entropy loss function. In the case of two classifications, the results predicted by the model are only two cases. For each category, our predicted probabilities are p_i and $1-p_i$, and the expression is:

$$L = \frac{1}{N} \sum_i^{i=0,1} -[y_i \times \log_{10}(p_i) + (1 - y_i) \times \log_{10}(1 - p_i)] \quad (12)$$

where N is the total number of samples, y_i represents the label of the sample, p represents the probability of prediction, $i = 1$ represents positive sample, and $i = 0$ represents negative sample.

Experiments

Configuration

The proposed system was implemented using the Python language and torch framework built on top of Pytorch. All experiments were performed on GPU resources: NVIDIA RTX 2080TI GPU with 11 GB memory.

Resnet50, VGG16, Inception_resnet_v2,³³ Shufflenet_v2,³⁴ and Densenet121³⁵ were chosen for training. To speed up the training speed of the model and improve the stability and generalization ability of the model, each network model in this article used the weights learned on the ImageNet data set. The parameters of the attention module were randomly initialized. The network parameters were optimized by the Adam algorithm, the weight decay was $1e-5$, the learning momentum was 0.9, and the batch size was 128. Each model was trained for 100 epochs, where the initial learning rate was set to 0.01 and the learning rate was multiplied by 0.1 every 20 epochs.

Performance metrics

The metrics used to evaluate the performance of the noise reduction algorithm were SSIM and PSNR.

$$PSNR = 10 \times \log_{10} \left(\frac{(2^n - 1)^2}{MSE} \right) \quad (13)$$

where MSE is the mean square error between the original image and the processed image.

The metrics used to evaluate the performance of the classification algorithm are precision (pre), Recall, F1 score, accuracy (ACC), and receiver operating characteristic (ROC). Pre represents the proportion of the predicted positives that are true positives (equation (10)), Recall represents the proportion of the true positives that are correctly predicted (equation (11)), while F1 is an indicator to measure the accuracy of a binary classification model and defined as the harmonic average of precision and recall (equation (12)). ACC is the percentage of accurately classified images (equation (13)). The ordinate of ROC curve represents the true positive rate (TPR), while the abscissa represents the false positive rate (FPR). The magnitude of the area under the ROC curve is represented by the area under ROC curve (AUC) value. AUC is an indicator used to evaluate the performance of classification models, which typically ranges from 0.5 to 1.0. A larger AUC means better performance of the network model.

$$Pre = \frac{TP}{TP + FP} \quad (14)$$

$$Recall = \frac{TP}{TP + FN} \quad (15)$$

$$F1 = \frac{2 \times Recall \times Pre}{Recall + Pre} \quad (16)$$

$$ACC = \frac{TN + TP}{TN + TP + FN + FP} \quad (17)$$

where FP refers to the NRDR images that are classified as RWDR, FN means the RWDR images that are classified as NRDR, TP refers to the RWDR images that are classified as RWDR and true negative, and TN is the NRDR images that are classified as NRDR.

Results

Results of the first stage

For commonly used deep learning networks, it is generally considered to be a black box, and the interpretability is not strong. With Grad-CAM,³⁶ we can draw the heatmap shown in Figure 5, which visualizes the regions of interest to the network corresponding to a given category. Grad-CAM can help us analyze the network attention area of a certain class, and then we in turn analyze whether the network has learned the correct features or information through the network attention area.

Figure 5(a) demonstrates that when using raw data, the network pays close attention to the dark border, which does not include valuable fundus information. To minimize the interference of dark border, we proposed an image preprocessing scheme. The results are shown in Figure 5(b). The fundus image now occupies a larger portion of the image, which makes the network pay more attention to it. Therefore, the network will learn to use the fundus information to discriminate DR rather than other things during the network training process.

Using t-distribution and stochastic neighbor embedding (t-SNE) visualization, we conducted additional analysis on the synthetic images.^{37,38} High-dimensional data can be integrated into a two-dimensional space using the t-SNE dimensionality reduction algorithm. The distribution of the blue RWDR data and the red NRDR data are not clearly separated when we apply t-SNE to examine the original data as shown in Figure 6(a). This indicates that there is no clear differentiation between the image's features, which is not good for feature learning in a network. After processing by the UGAN algorithm, the feature distributions of RWDR and NRDR in the data set show a more separate trend, as shown in Figure 6(b). More distinct features will help the classification network better distinguish DR category.

To imitate the low-quality image in the data set, the clear image was degraded, and the low-quality image was then provided to the network for preprocessing. We added noise to the original color images to obtain the simulated low-quality images, and used the UGAN network to process it. To visually verify the effect from multiple perspectives, we further performed image grayscale processing and vessel segmentation on the original and simulated color image,

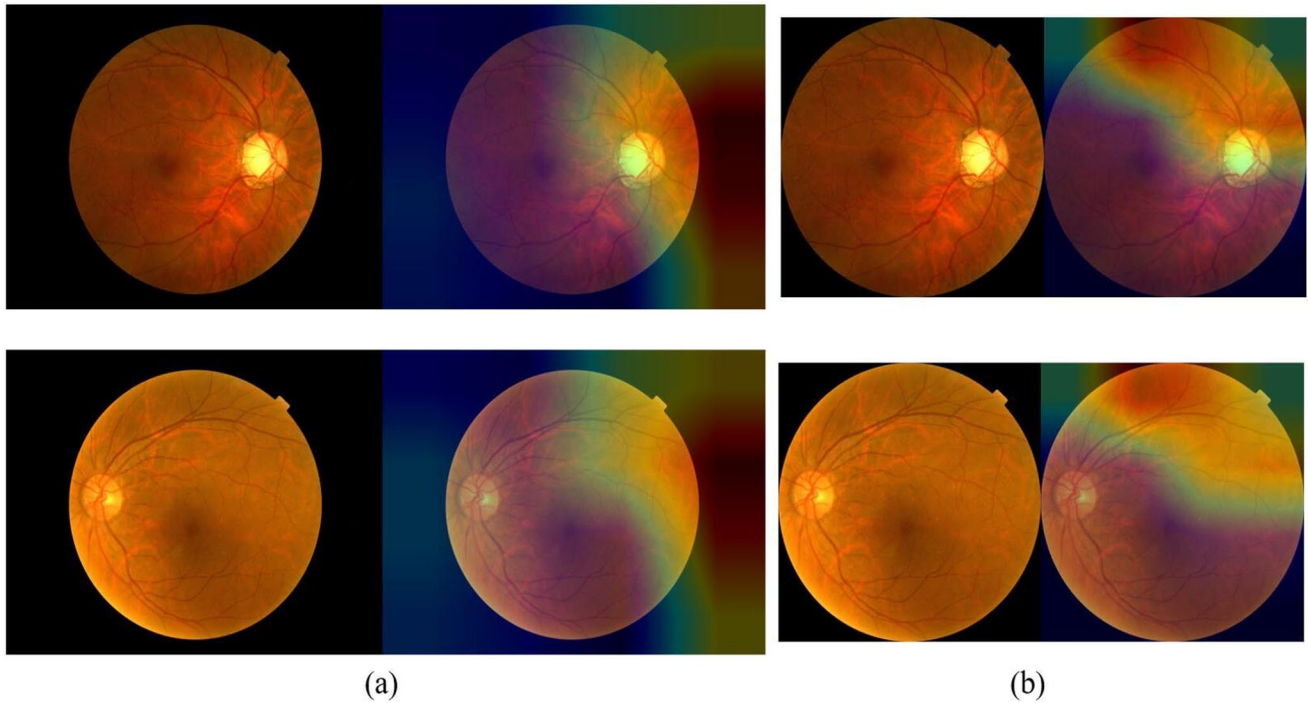


Figure 5. (a) The distribution of the network's attention in the original data. (b) The distribution of network attention in the processed data.

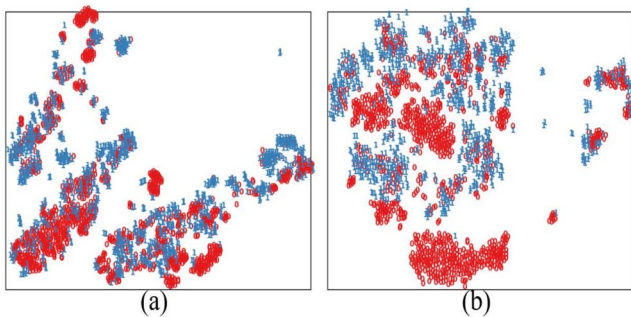


Figure 6. (a) t-SNE between RWDR (blue) and NRDR (red) in the original data set. (b) t-SNE between RWDR (blue) and NRDR (red) in the UGAN-enhanced data set.

respectively. The results are shown in Figure 7. The simulated low-quality images (Figure 7(b1) and (b2)) lose a lot of fine information when compared to the label images (Figure 7(a1) and (a2)), such as the absence of small blood vessels, local darkness, and so on. In contrast, after UGAN processing, the blood vessel outline has roughly been restored and the image clarity has also been improved (Figure 7(c1) and (c2)), even if the label image's clarity could not be completely recovered. At the same time, we introduced the PSNR and SSIM to evaluate the maximum signal and background noise of the image, and the similarity between the two images, respectively. The results are displayed in Table 2. The SSIM is enhanced by roughly 0.1, and the PSNR values of the first and second groups of images are both increased to more than 20. The PSNR improvement ratio is 37.2%, and the SSIM also shows improvement of nearly 0.1, for the whole data set.

To verify the reliability of the UGAN network, we combined it with other classification networks for validation, and trained and tested it on the same data set. The experimental

Table 2. Comparison of PSNR and SSIM Evaluation Scores Before and After Image Processing.^a

	a1		a2		All data set	
	b1	c1	b2	c2	$\sum_{i=1}^n b/n$	$\sum_{i=1}^n c/n$
PSNR	19.03	24.82	13.33	21.27	16.45	22.57
SSIM	0.80	0.89	0.69	0.77	0.74	0.83

PSNR: peak signal-to-noise ratio; SSIM: structure similarity.
^aa1/a2, c1/c2, and b1/b2 in the table represent the PSNR and SSIM evaluation scores of the corresponding RGB fundus images in Figure 7, respectively.

results are shown in Figure 8. The accuracy has been significantly improved, confirming the effectiveness of the UGAN network in improving the accuracy of the algorithm. Table 3 shows the results of the algorithm used in the experiment in the test data set, in which we highlight the value of best indicators of the network with and without UGAN with bold numbers. Comparisons were performed between the groups with and without UGAN, respectively, using student t-test, and $p < 0.05$ meant significant difference. Although Resnet50 and Densenet121 have their own advantages in all indicators in the classification network without adding UGAN, but UGAN_Resnet50 is better than UGAN_Densenet121 in all indicators.

Results of the second stage

With the cropped black-edge data set, we trained Densenet121, VGG16, Inception Resnet_v2, Shufflenet_v2, and Resnet50 to filter out the best classification network. All networks arrived at convergence after 50 epochs. Figure 8 displays the accuracy results for each of them, with

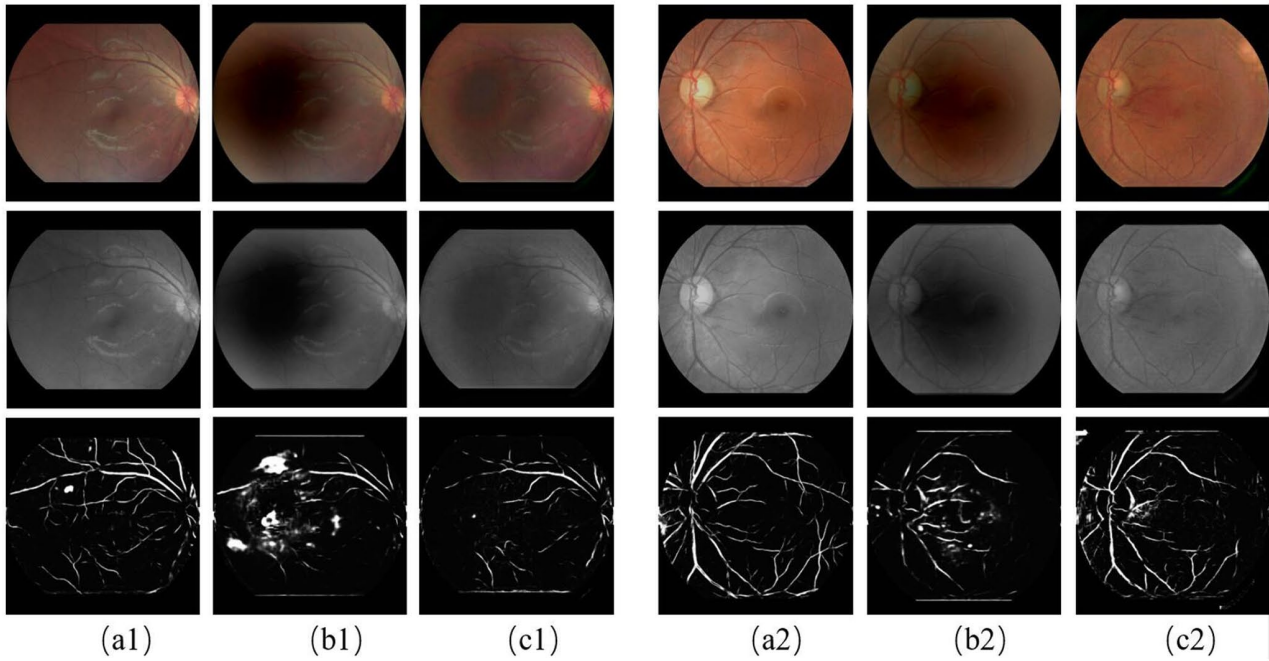


Figure 7. (a1, a2) Label images. (b1, b2) simulated low-quality images. (c1, c2) After UGAN, including RGB fundus images, grayscale fundus images, and vessel segmentation images.

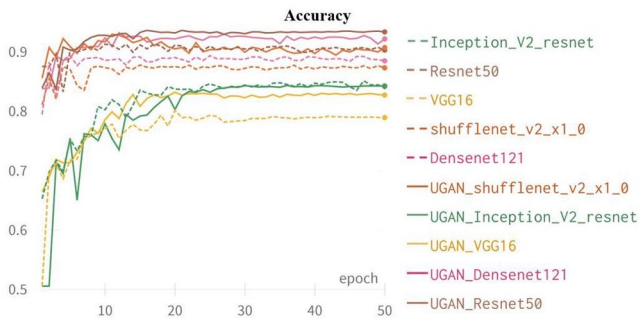


Figure 8. Curves of the accuracy for Shufflenet_v2, Densenet121, VGG, Inception_resnet_v2, and Resnet50. The dashed line is the classic classification network, and the solid line of the same color is the corresponding classification network with UGAN with preprocessing added.

Densenet121, Shufflenet_v2, and Resnet50 reaching the convergence state more quickly. The UGAN network makes better use of the image feature data owing to the image preprocessing operation. Resnet50 is noticeably superior to all other networks after the addition of the UGAN module, thus we decide to add the attention mechanism CBAM to Resnet50 to further raise the network’s performances. Figures 9 and 10 display the experimental results. Resnet50, UGAN_Resnet50, and URNet are shown by the green, blue, and red lines, respectively. The three networks all gradually converge after 20 epochs, with the red line stabilizing at the top value. The results also demonstrate that once the CBAM module was added, network prediction accuracy is significantly improved, showing that CBAM can effectively employ mining both the channel information and cross-space information to enhance the network’s capacity. When our URNet was compared with the original Resnet50, both AUC and accuracy obtained the best evaluation indices,

even though there was no linear link between them. As shown in Figure 11, it can be seen from the two sets of data (a) and (b) that even though the network classification was correctly predicted, the prediction results after adding UGAN were more confident than those of direct reasoning.

To verify the reliability of the algorithm, in addition to using the DDR data set, we also selected 5100 images from Kaggle data set and made a cross data set evaluation. It should be noted that for a fair comparison, all models were trained under the same training conditions and tested on the same data set. As can be seen from Table 4, when UGAN_Resnet50 performs image prediction after adding CBAM attention, all indicators are improved to a certain extent. Comparisons were performed between the groups with and without UGAN, as well as between those CBAM and without CBAM, respectively, using student t-test, and $p < 0.05$ meant significant difference. Our proposed model URNet achieved accuracy 94.4%, recall (96.2%, 94.4%), precision (94.5%, 91.9%), and F1 score (95%, 95%) in DDR data set, and achieved accuracy 93.1%, recall (93.4%, 93.6%), precision (95.4%, 90.7%), and F1 score (95%, 92%) in Kaggle data set.

Discussion

Due to the shortage of medical resources, it is impossible to avoid collecting poor-quality images in community screening, which will cause disturbances for computer-aided diagnosis. Therefore, we start from the algorithm itself to make it better adapted to the community screening environment. Before the classification network, we added an image preprocessing module, which was trained by comparing clear images with low-quality images. This module finds the conversion relationship between the low-quality image and the clear image, so that when the low-quality image is input, it can be reconstructed into a clear image.

Table 3. Comparison of Various Evaluation Indicators of Each Classification Network Before and after Adding UGAN.

Net	Accuracy			F1			Precision			Recall		
	ALL	NRDR	RWDR	NRDR	RWDR	NRDR	NRDR	RWDR	NRDR	RWDR		
VGG16	0.789 ± 0.004	0.79	0.79	0.793 ± 0.006	0.786 ± 0.003	0.784 ± 0.005	0.795 ± 0.006					
Inception_resnet_v2	0.842 ± 0.003	0.84	0.84	0.851 ± 0.003	0.833 ± 0.003	0.829 ± 0.004	0.855 ± 0.004					
Shufflenet_v2	0.873 ± 0.003	0.88	0.87	0.855 ± 0.003	0.893 ± 0.003	0.898 ± 0.003	0.848 ± 0.003					
Densenet121	0.885 ± 0.003	0.88	0.89	0.904 ± 0.003	0.875 ± 0.005	0.864 ± 0.003	0.906 ± 0.004					
Resnet50	0.887 ± 0.003(p<0.001)	0.90	0.87	0.872 ± 0.003	0.908 ± 0.004	0.931 ± 0.002	0.832 ± 0.006					
UGAN_VGG16	0.827 ± 0.003(p<0.001)	0.83	0.82	0.812 ± 0.002(p<0.001)	0.853 ± 0.003(p<0.001)	0.857 ± 0.003(p<0.001)	0.796 ± 0.003(p=0.581)					
UGAN_inception_resnet_v2	0.843 ± 0.004(p=0.347)	0.85	0.84	0.825 ± 0.002(p<0.001)	0.863 ± 0.004(p<0.001)	0.875 ± 0.003(p<0.001)	0.810 ± 0.004(p<0.001)					
UGAN_shufflenet_v2	0.903 ± 0.004(p<0.001)	0.90	0.90	0.907 ± 0.004(p<0.001)	0.899 ± 0.004(p=0.131)	0.898 ± 0.005(p=0.443)	0.908 ± 0.003(p<0.001)					
UGAN_densenet121	0.907 ± 0.003(p<0.001)	0.91	0.91	0.921 ± 0.003(p<0.001)	0.894 ± 0.003(p<0.001)	0.893 ± 0.003(p<0.001)	0.922 ± 0.004(p<0.001)					
UGAN_resnet50	0.932 ± 0.003(p<0.001)	0.94	0.92	0.941 ± 0.002(p<0.001)	0.915 ± 0.002(p=0.007)	0.939 ± 0.002(p=0.032)	0.930 ± 0.004(p<0.001)					

NRDR: non-referral diabetic retinopathy; RWDR: referral-warranted diabetic retinopathy.

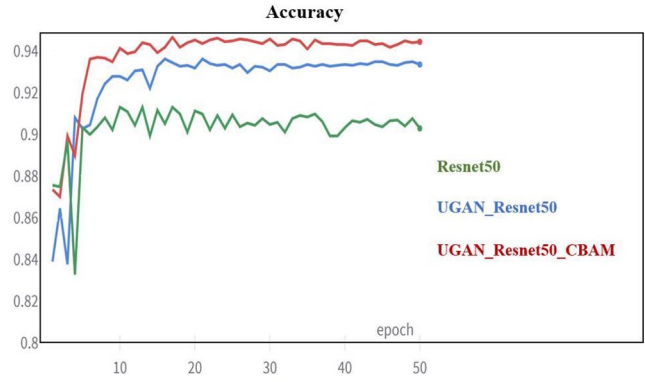


Figure 9. Curves of the accuracy for Resnet50 with CBAM and UGAN.

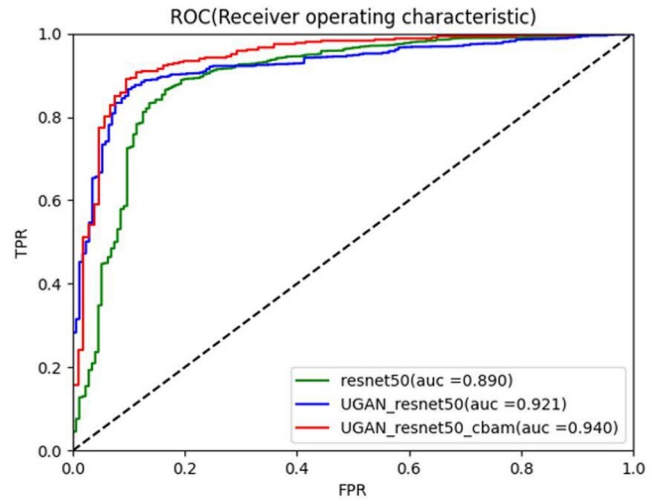


Figure 10. Curves of the ROC for Resnet50 with CBAM and UGAN.

This module can increase the PSNR of the image by six points and the SSIM value by 12%. At the same time, we applied this module to multiple classification networks, and the accuracy of the classification network has been improved by 5% on average, which proves that this module provides a favorable basis for subsequent classification network judgment.

In this work, we have established a fundus image enhancement data set, which contains 1000 sets of clear images and their corresponding low-quality images. The ratio of RWDR and NPDR images in the data set was 1:1, and the data set was divided into training set and test set with a ratio of 3:1. This data set can be used for related algorithm research. The work of image reconstruction is similar to that proposed by Shen *et al.*,²⁰ but we do not take into account the impact of dust on the lens of imaging system. There is a big difference between the interference caused by the dust and the anomalous exposure on the image. Removing these two kinds of interference at the same time will increase the huge data demand and computing cost. Considering that the collected images are seldom polluted with dust on lens, we choose to ignore the impact of this item.

However, the algorithm still has some limitations. In addition to the impact of dust on the lens, the types of image noise collected in community screening are more complex. The combination of impacts of different devices and environments gives rise to various types of noise, which cannot be

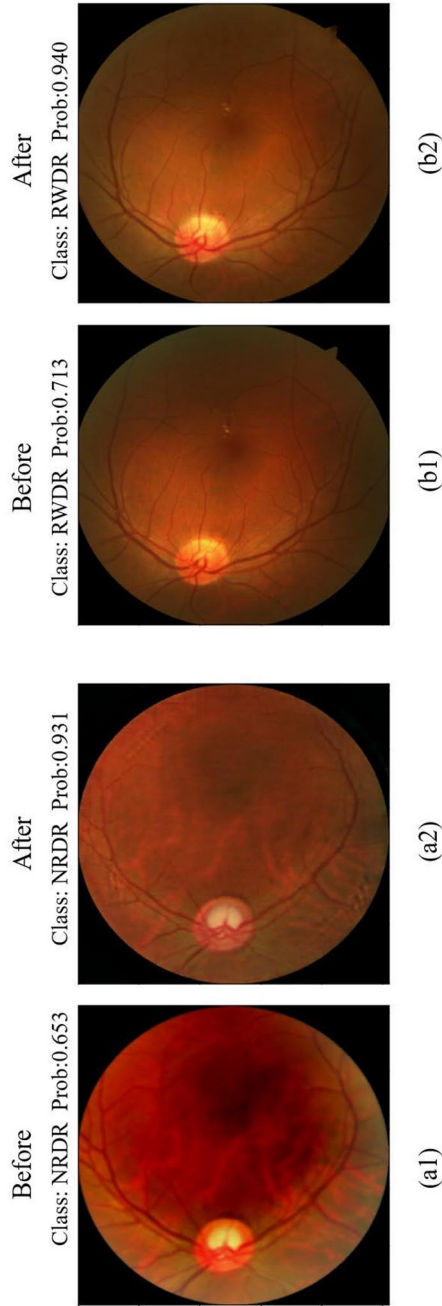


Figure 11. Confidence level of Resnet50 on DR prediction before and after UGAN processing. (a1, a2) Class: NRDR, confidence: 0.653–0.931; (b1, b2) Class: RWDR, confidence: 0.713–0.940.

Table 4. Ablation Experiment of Resnet50, UGAN, and CBAM.

Data sets	Net	Accuracy		Precision		Recall		F1	
		ALL	ALL	NRDR	RWDR	NRDR	RWDR	NRDR	RWDR
DDR	Resnet50	0.887 ± 0.004	0.872 ± 0.003	0.908 ± 0.002	0.908 ± 0.003	0.931 ± 0.004	0.832 ± 0.003	0.90	0.87
	UGAN_Resnet50	0.932 ± 0.003(p<0.001)	0.941 ± 0.003(p<0.001)	0.915 ± 0.003(p<0.001)	0.915 ± 0.003(p<0.001)	0.939 ± 0.003(p=0.003)	0.930 ± 0.003(p<0.001)	0.94	0.92
	URNet (Ours)	0.944 ± 0.003(p<0.001)	0.945 ± 0.002(p=0.142)	0.919 ± 0.003(p<0.001)	0.919 ± 0.003(p<0.001)	0.962 ± 0.003(p<0.001)	0.944 ± 0.003(p<0.001)	0.95	0.95
Kaggle	Resnet50	0.860 ± 0.004	0.881 ± 0.004	0.838 ± 0.004	0.838 ± 0.004	0.890 ± 0.003	0.826 ± 0.003	0.89	0.83
	UGAN_Resnet50	0.918 ± 0.004(p<0.001)	0.946 ± 0.004(p<0.001)	0.889 ± 0.003(p<0.001)	0.889 ± 0.003(p<0.001)	0.921(p<0.001)	0.923 ± 0.004(p<0.001)	0.93	0.91
	URNet (Ours)	0.931 ± 0.003(p<0.001)	0.954 ± 0.002(p<0.001)	0.907 ± 0.002(p<0.001)	0.907 ± 0.002(p<0.001)	0.934 ± 0.003(p<0.001)	0.936 ± 0.002(p<0.001)	0.95	0.92

sing t-test for significance analysis, p<0.05 means significant correlation; NRDR: non-referral diabetic retinopathy; RWDR: referral-warranted diabetic retinopathy.

completely simulated with our algorithm. For more complex noise components, it needs more investigations in the future.

Conclusions

To solve the problems of image blurring and abnormal exposure in DR community screening, we proposed an automatic DR classification method URNet based on deep learning. Using the DDR data set, we successively trained the two-stage URNet algorithm. First, we selected clear retinal images from the DDR data set for degradation, and used the clear images as labels of degraded images to create a data set for UGAN training. Then, we reclassified the DDR data set into RWDR and NRDR according to the severity of the DR lesions. The reclassified images were preprocessed by the trained UGAN network to filter out the features such as abnormal exposure and blurring in the images, and sent to the classification network as a new data set for training. The results showed that the proposed two-stage network URNet can solve the problem of low classification accuracy caused by low image quality. Therefore, the scheme is more suitable for applications in community screening of DR.

AUTHORS' CONTRIBUTIONS

All authors participated in the design, interpretation of the studies and analysis of the data, and review of the manuscript; YL and LX conducted the experiments; CZ, YL, and KY designed the study and wrote the manuscript. SC and YY analyzed the experimental data.


DECLARATION OF CONFLICTING INTERESTS

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

FUNDING

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was funded by the National Natural Science Foundation of China (grant no. 61875123), the Research Fund for Hebei University multidisciplinary (grant no. DXK201914), the President of Hebei University (grant no. XZJJ201914), the Natural Science Foundation of Hebei Province (grant no. H2019201378), the Special Project for Cultivating College Students' Scientific and Technological Innovation Ability in Hebei Province (grant no. 22E50041D), Guangdong Basic and Applied Basic Research Foundation (grant no. 2021A1515011654), and the Fundamental Research Funds for the Central Universities of China (grant no. 20720210117) and Shenzhen Science and Technology Program (grant no. 1210318663).

ORCID ID

Chuanqing Zhou  <https://orcid.org/0000-0002-9030-4993>

REFERENCES

- Kumar PNS, Deepak RU, Sathar A, Sahasranamam V, Kumar RR. Automated detection system for diabetic retinopathy using two field fundus photography. *Procedia Comput Sci* 2016;**93**:486–94
- Zhu CZ, Hu R, Zou BJ, Zhao RC, Chen CL, Xiao YL. Automatic diabetic retinopathy screening via cascaded framework based on image-and lesion-level features fusion. *J Comput Sci Technol* 2019;**34**:1307–18
- Li T, Gao Y, Wang K, Guo S, Liu H, Kang H. Diagnostic assessment of deep learning algorithms for diabetic retinopathy screening. *Inf Sci* 2019;**501**:511–22
- Hazin R, Colyer M, Lum F, Barazi MK. Revisiting diabetes 2000: challenges in establishing nationwide diabetic retinopathy prevention programs. *Am J Ophthalmol* 2011;**152**:723–9
- Tsiknakis N, Theodoropoulos D, Manikis G, Ktistakis E, Boutsora O, Berto A, Scarpa F, Scarpa A, Fotiadis DI, Marias K. Deep learning for diabetic retinopathy detection and classification based on fundus images: a review. *Comput Biol Med* 2021;**135**:104599–19
- Li F, Wang Y, Xu T, Dong L, Yang L, Jiang M, Zhang X, Jiang H, Wu Z, Zou H. Deep learning-based automated detection for diabetic retinopathy and diabetic macular oedema in retinal fundus photographs. *Eye* 2021;**7**:36–45
- Wang R, Chen B, Meng D, Wang L. Weakly-supervised lesion detection from fundus images. *IEEE Trans Med Imaging* 2019;**38**:1501–12
- Ashiba HI, Awadallah KH, El-Halfawy SM, Abd El-Samie FE. Homomorphic enhancement of infrared images using the additive wavelet transform. *Prog Electromagn Res C* 2008;**1**:123–30
- Xiong L, Li H, Xu L. An enhancement method for color retinal images based on image formation model. *Comput Methods Programs Biomed* 2017;**143**:137–50
- Veras R, Silva R, Araújo F, Medeiros F. SURF descriptor and pattern recognition techniques in automatic identification of pathological retinas. In: *Proceedings of the 2015 Brazilian conference on intelligent systems*, Natal, Brazil, 4–7 November 2015, pp.316–321. New York: IEEE
- Yaqoob MK, Ali SF, Bilal M, Hanif MS, Al-Saggaf UM. Resnet based deep features and random forest classifier for diabetic retinopathy detection. *Sensors* 2021;**21**:1–14
- Luo X, Pu Z, Xu Y, Wong WK, Su J, Dou X, Ye B, Hu J, Mou L. MVDRNet: multi-view diabetic retinopathy detection by combining DCNNs and attention mechanisms. *Pattern Recognit* 2021;**120**:108104
- Al-Moosawi NM, Khudayer RS. ResNet-34/DR: a residual convolutional neural network for the diagnosis of diabetic retinopathy. *Informatika* 2021;**45**:115–24
- Alyoubi WL, Abulkhair MF, Shalash WM. Diabetic retinopathy fundus image classification and lesions localization system using deep learning. *Sensors* 2021;**21**:1–22
- Tseng VS, Chen CL, Liang CM, Tai MC, Liu JT, Wu PY, Deng MS, Lee YW, Huang TY, Chen YH. Leveraging multimodal deep learning architecture with retina lesion information to detect diabetic retinopathy. *Transl Vis Sci Technol* 2020;**9**:41–12
- Tan CH, Kyaw BM, Smith H, Tan CS, Car LT. Use of smartphones to detect diabetic retinopathy: scoping review and meta-analysis of diagnostic test accuracy studies. *J Med Internet Res* 2020;**22**:1–12
- Chen Z, Bei Y, Rudin C. Concept whitening for interpretable image recognition. *Nat Mach Intell* 2020;**2**:772–82
- Woo S, Park J, Lee JY, Kweon IS. Cbam: convolutional block attention module Sanghyun. In: *European conference on computer vision*, Munich, 8–14 September 2018, pp.3–19. New York: IEEE
- Wilkinson CP, Ferris FL, III, Klein RE, Lee PP, Agardh CD, Davis M, Dills D, Kampik A, Pararajasegaram R, Verdaguer JT, Global Diabetic Retinopathy Project Group. Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales. *Ophthalmology* 2003;**110**:1677–82
- Shen Z, Fu H, Shen J, Shao L. Modeling and enhancing low-quality retinal fundus images. *IEEE Trans Med Imaging* 2021;**40**:996–1006
- Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: *Proceedings of the international conference on medical image computing and computer-assisted intervention*, Munich, 5–9 October 2015, pp.234–241. New York: IEEE
- Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition, <https://arxiv.org/abs/1409.1556>
- Ancuti CO, Ancuti C. Single image dehazing by multi-scale fusion. *IEEE Trans Image Process* 2013;**22**:3271–82
- Nebauer C. Evaluation of convolutional neural networks for visual recognition. *IEEE Trans Neural Netw* 1998;**9**:685–96
- He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and*

- pattern recognition*, Las Vegas, NV, 27–30 June 2016, pp.770–778. New York: IEEE
26. Litjens G, Kooi T, Bejnordi EB, Setio AAA, Ciompi F, Ghafoorian M, van der Laak JAWM van Ginneken B, Sánchez CI. A survey on deep learning in medical image analysis. *Med Image Anal* 2017;**42**:60–88
 27. Yan Y, Jin K, Gao Z, Huang X, Wang F, Wang Y, Ye J. Attention-based deep learning system for automated diagnoses of age-related macular degeneration in optical coherence tomography images. *Med Phys* 2021;**48**:4926–34
 28. Isola P, Zhu JY, Zhou T, Efros AA. Image-to-image translation with conditional adversarial networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, Honolulu, HI, 21–26 July 2017, pp.5967–5976. New York: IEEE
 29. Nair V, Hinton GE. Rectified linear units improve restricted Boltzmann machines. In: *Proceedings of the 27th international conference on international conference on machine learning*, Haifa, Israel, 21–24 June 2010, pp.807–814. New York: IEEE
 30. Silver DL, Bennett KP. Guest editor’s introduction: special issue on inductive transfer learning. *Mach Learn* 2008;**73**:215–20
 31. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Commun ACM* 2017;**60**:84–90
 32. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. *IEEE Trans Image Process* 2004;**13**:600–12
 33. Szegedy C, Ioffe S, Vanhoucke V, Alemi AA. Inception-v4, inception-ResNet and the impact of residual connections on learning. In: *Proceedings of the association for the advancement of artificial intelligence*, San Francisco, CA, 4–9 February 2017, pp.4278–4284. New York: IEEE
 34. Zhang X, Zhou X, Lin M, Sun J. ShuffleNet: an extremely efficient convolutional neural network for mobile devices. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, Salt Lake City, UT, 18–23 June 2018, pp.6848–6856. New York: IEEE
 35. Huang G, Liu Z, van der Maaten L, Weinberger KQ. Densely connected convolutional networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, Honolulu, HI, 21–26 July 2017, pp.2261–2269. New York: IEEE
 36. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: visual explanations from deep networks via gradient-based localization. *Int J Comput Vis* 2020;**128**:336–59
 37. van der Maaten L, Hinton G. Visualizing data using t-SNE. *J Mach Learn Res* 2008;**9**:2579–605
 38. Yoon D, Kong HJ, Mai N, Kim BS, Cho WS, Lee JC, Cho M, Lim MH, Yang SY, Lim SH, Lee J, Song JH, Chung GE, Choi JM, Kang HY, Bae JH, Kim S. Colonoscopic image synthesis with generative adversarial network for enhanced detection of sessile serrated lesions using convolutional neural network. *Sci Rep* 2022;**12**:261–73

(Received November 17, 2022, Accepted February 1, 2023)