Intelligent detection and applied research on diabetic retinopathy based on the residual attention network

Moye Yu1 | Yi Wang1,2,3

1Department of Information Center, Fudan University Shanghai Cancer Center, Shanghai, China
2Department of Oncology, Shanghai Medical College, Fudan University, Shanghai, China
3Shanghai Research Center of Artificial Intelligence Engineering Technology for Cancer, Shanghai, China

Correspondence
Yi Wang, Department of Information Center, Fudan University Shanghai Cancer Center, Shanghai, China.
Email: tonywangfz@163.com

Funding information
National key R&D project, Grant/Award Numbers: 2018YFC1314900, 2018YFC1314902; National Natural Science Foundation of China, Grant/Award Number: 81971708

Abstract
This study proposes a high-accuracy (ACC) algorithm to automatically detect diabetic retinopathy (DR) and diabetic macular edema (DME) in retinal fundus images. Three DR datasets were obtained for use in this study: EyePACS, Messidor, and IDRid. In the EyePACS dataset, two DR classifications and five classifications experiments were conducted. The Messidor and IDRid dataset were graded DR and DME. After preprocessing, enhancement, and normalizing, common convolutional neural networks (CNN) were used to obtain the classification results. Afterward, an optimization method residual attention network (RAN) was introduced that was based on the residual attention module, and incorporated dilated convolution, so as to optimize the experimental results. The focal loss was then added to solve the imbalance problem. Next, a five-fold cross-validation strategy was introduced so as to assess and optimize the proposed model, after which the prediction ACC, sensitivity, specificity, area under receiver operating curve, and Kappa score were assessed. The proposed method RAN was shown to achieve 89.2% ACC (95% confidence interval [CI], 0.8782–0.9123) for two DR classifications (normal and abnormal) on the EyePACS dataset, 89.8% ACC (95% CI, 0.8751–0.9275) for two DR classifications on the Messidor dataset. The IDRid dataset achieved an ACC of 71.5% (95% CI, 0.6941–0.7423) for the two DR classifications. RAN mainly improves the results of commonly used CNN methods on the same dataset. Therefore, the classification and diagnosis of DR may be improved by adopting the proposed method.

KEYWORDS
artificial intelligence, attention mechanism, CNN, diabetic retinopathy, dilated convolution, fundus image

1 INTRODUCTION

Diabetic retinopathy (DR) is a late manifestation of diabetes mellitus and is one of the most severe complications of diabetic microangiopathy. If it is not detected and treated early, DR can cause irreversible visual impairment or even blindness in severe cases.1 Fundus imaging is a vital method of inspection for the early detection of DR lesions. The corresponding fundus changes and grading standards used in DR are shown in Figures 1 and 2. As ophthalmologists in less developed areas are lacking, patients suffering from diabetes are unable to receive an early diagnosis and treatment for DR.2 Therefore, computerized screening technology...
based on fundus images is of great significance in delaying the progression of DR.

High-quality color retina images can assist doctors in investigating and diagnosing retinopathy. However, the diagnosis of DR requires a clinically experienced ophthalmologist, and DR screening is not performed in most grass-roots areas, which has significantly increased the risk of blindness due to diabetes. Therefore, adopting computer-assisted remote diagnostic technology in fundus imaging can effectively result in reductions in visual impairment among diabetic patients due to insufficient medical resources.

At present, most ophthalmology image analysis work focuses on DR classification, vessel segmentation, and detection of retina structures. Pratt et al. developed a network with convolutional neural networks (CNN) architecture and data augmentation that can identify intricate features involved in the classification of DR. They then trained the model on the EyePACS dataset, achieving a sensitivity (SE) of 95% and accuracy (ACC) of 75% for 5000 validation images. Rahim et al. presented an automatic detection method for DR and maculopathy in fundus images by employing fuzzy image processing techniques, circular hough transform, and several feature extraction methods were implemented. Eftekhar et al. adopted a two-step process with two online datasets to train CNN, which solved the problem of imbalance and reduced the training time while performing accurate detections. Seth et al. used CNN and linear support vector machines to train the network on the benchmark dataset EyePACS dataset, which demonstrated that the model had high SE and specificity (SP) in detecting DR. Dutta et al. proposed an automatic knowledge model to identify critical prerequisites for disaster recovery. After testing the model using a central processing unit CPU-trained neural network, three types of back-propagation neural networks were used. Accordingly, the model was able to quantify the characteristics of different types of blood vessels, exudates, bleeding, and microaneurysms. Adem et al. utilized a CNN model with a circular hough transform and applied it to retinal images. The results demonstrated that using a CNN model together with image processing can improve the ACC, with a success rate of exudate detection of 99.18%. Moreover, Li et al. developed a DL system for glaucomatous optic neuropathy (GON) classification to automatically classify GON on color fundus images. The
AUC of the system was 0.986, with a SE of 95.6% and SP of 92.0%. Wu et al. proposed a hierarchically Coarse-to-fine network (CF-DRNet) to classify the five stages of DR severity using CNN, which showed that CF-DRNet outperformed various state-of-the-art methods in the publicly available IDRiD and EyePACS datasets. Arenas-Cavalli et al. evaluated the automated DR screening tool DART, for which the receiver operating curve (ROC) analysis indicated a SE of 94.6%, SP of 74.3%, and AUC of 0.915. Furthermore, Dai et al. developed a system called DeepDR, which was able to detect early to late stages of DR. The grading of DR as mild, moderate, severe, and proliferative achieved an AUC of 0.943, 0.955, 0.960, and 0.972, respectively.

This article will factor in the needs of both ophthalmologists and diabetic patients by proposing a DL algorithm RAN to improve the performance of DR diagnosis in the model (Tables 1–3).

## MATERIAL AND METHODS

In this study, experiments were conducted on the EyePACS, Messidor, and IDRid datasets, as shown in Figure 3. Based on ResNet, the algorithm RAN proposed in this paper integrated the attention mechanism and added the attention guided module (AGM) and dilated convolution.

### 2.1 Database

This study utilized three public datasets. EyePACS's training set contained 35 125 fundus images released by the California Medical Foundation from eye-PACS users, including level 0 25 809 (74%), level 1 2443 (7%), level 2 5292 (15%), level 3 873 (2%), and level 4 708 (2%). Due to the excessive number of normal fundus images in this dataset, 40% of the normal images were selected for training and testing during the two classification experiments, while only 20% of the normal images were selected for training during the five classification experiments and tests. The Messidor dataset consisted of 1200 fundus images from three ophthalmology hospitals, 800 of which were images obtained following pupil dilation. Each picture was marked with a DR lesion grade of 0–3, and each picture had a DME lesion grade of 0–2. Table 4 lists the number distribution. The image sizes in the dataset were 1440 × 960, 2240 × 1488, and 2304 × 1536, respectively, in tif format. The IDRid dataset included lesion segmentation, disease classification, and optic disk and fovea detection. In this experiment, only disease classification data were used, including 413 pictures in the training set and 103 pictures in the test set. All pictures were 4288 × 2848 and in jpg format. Table 5 shows the number and proportion distribution. In these three datasets, 60%, 15%, and 25% of the images in each dataset were randomly selected as the training set, validation set, and test set, respectively.

As the number of abnormal pictures in the Messidor and IDRid dataset was too small, it was more meaningful to perform two classifications in terms of clinical application. Evidently, according to the above dataset, the most prominent feature of medical images is the imbalance in the distribution of data; the number of samples in normal images is much higher than that of abnormal images,
while the amount of grading data with the severity of the disease is becoming increasingly less. In order to solve this problem, data preprocessing and optimizing the loss function are most commonly utilized.

Commonly used data augmentation methods include translation, rotation, cropping, scaling, noise addition, affine transformation, and so forth. These methods usually do not change the type of object and are the earliest and most widely used types of image-enhancing methods. The color of the image can also be changed according to four areas: brightness, contrast, saturation, and tone (Figure 4).

In order to reduce the difference between the different images of the dataset, before sending the image to the...
network for training, normalization was implemented in each image:

\[ G_{i,j,k}^\prime = \frac{G_{i,j,k} - m_k}{\sigma_k}, \]  

(1)

where \( G_{i,j,k} \) is the input image, \( G_{i,j,k}^\prime \) is the normalized image, \( i \) and \( j \) is the coordinates of the pixel points, \( k \) represents the three channels of the image (blue, green, and red), \( m_k \) represents the average value of the \( k \)th channel pixel value, and \( \sigma_k \) represents the standard deviation of the \( k \)th channel pixel value.

The loss function in the neural network was used to measure the gap between the predicted value obtained by the model and the actual value of the data, which also served as the standard used to measure the generalization ability of the model. The smaller the loss function, the better the performance of the model, and the loss function used by different models are generally different. The most commonly used loss function is cross-entropy. Since the imbalance problem generally exists in DR datasets, focal loss was introduced in this experiment, which was modified on cross-entropy. It multiplies the original cross-entropy by an index that weakens the contribution of the easily detectable object to the model training.

### 2.2 Residual attention network

The core of RAN (Figure 5) served as the attention mechanism, which can increase the area information of the lesion and suppress other background information, thereby improving the ACC of the model in the DR classification. As module stacking becomes more in-depth,
different levels of attention information can be extracted from top to bottom, and the attention perception function from different modules will change adaptively. The added attention residual learning structure can train very deep residual attention networks, which may also be easily extended to hundreds of layers. By stacking this residual attention structure, the advantages of residual learning and attention mechanism can be thoroughly combined to achieve better results.

Each attention mechanism module is divided into two branches: the soft mask branch and trunk branch (Figure 6). The formula of the attention mechanism is

$$H_{ic}(x) = M_{ic}(x) \times T_{ic}(x),$$

(2)

where $T$ represents the main branch and $M$ represents the mask branch. The mask branch used several maximum pooling to increase the receptive field. After reaching the minimum resolution, a symmetric network structure was used to amplify the features back.

As shown in Figure 7, AGM was also added into RAN, which was composed of two 1x1 convolution layers with different activation functions in the adaptive average pooling layer. The specific operation was carried out according to the following. First, the input feature map passed through an adaptive average pooling layer, with an output feature map dimension of...

FIGURE 5 Residual attention network structure

FIGURE 6 Soft mask branch and trunk branch

FIGURE 7 AGM and RAN
Next, after a $1 \times 1$ convolution layer with the linear rectification function (ReLU) activation function, the output feature map dimension was $R^{1 \times 1 \times M/r}$, and the number of channels was reduced from $M$ to $M/r$. Then, after a $1 \times 1$ convolution layer with sigmoid activation function, the number of channels was expanded from $M/r$ to $M$, and a channel descriptor with dimension $R^{1 \times 1 \times M}$ was obtained so as to recalibrate the original feature map. Among them, the hyper-parameter $r$ can control the calculation amount of AGM, which was set to 16. Finally, by multiplying the obtained channel descriptor and input feature map, recalibration of the feature map was completed, and the importance of each channel was recalibrated by integrating the global information. The importance of different channels varied, highlighting the important information while suppressing background information.

2.3 | Dilated convolution module

In order to expand the receptive field and capture multi-scale contextual information, this article also adopted a dilated convolution module. As shown in Figure 8, the dilated convolution was equivalent to the filling $d − 1$ dilations between adjacent convolution kernel parameters. When the dilation rate $d = 1$, the dilated convolution degenerated into a standard convolution; the larger the $d$, the larger the receptive field of the convolution kernel. In this article, $1 \times 1$ standard convolution, $3 \times 3$ dilated convolution with dilation rate $d = 2$, $3 \times 3$ dilated convolution with dilation rate $d = 3$, $3 \times 3$ dilated convolution with dilation rate $d = 5$, and global average pooling was used to extract the features. Five levels of image information were subsequently extracted. The specific process in regard to using global average pooling to extract features was to initially utilize an adaptive average pooling layer in order to generate a $1 \times 1 \times 512$ dimension feature map. Then, a $1 \times 1$ convolution was used to change the number of channels to 256, after which the bilinear interpolation algorithm was adopted to expand its size to $14 \times 14$. The extracted feature maps of five levels were then spliced with the original feature maps to obtain a $14 \times 14 \times 1792$ dimension feature map. Finally, $1 \times 1$ convolution was done to change the number of channels to 512. After each convolution operation, a batch normalization layer and ReLU activation function were integrated. Before each dilated convolution extracted the features, the feature map carried out a padding operation in order to ensure that the resolution of the feature map before and after did not change.

2.4 | Transfer learning

Transfer learning is a method of machine learning that encompasses transplanting the model obtained from training one task to training other tasks. In this experiment, the pretrained EfficientNet weights were loaded on ImageNet to the three DR datasets so as to train the proposed model to obtain better results. In addition, in order to improve the performance and alleviate issues pertaining to the small amount of data in the Messidor and IDRid datasets, the weights RAN model trained on the EyePACS dataset was transferred to the Messidor and IDRid datasets.
### 2.5 Implementation details

The PyTorch framework and OpenCV image processing library were applied in this experiment, which was implemented on the Ubuntu16.04 operating system with a GeForce RTX 2080Ti graphics card. The Adam optimizer initial learning rate was 0.001, the batch size training phase was 16, the testing phase was 4, and a total of 60 epochs were trained. In addition, each image was initially scaled to \(512 \times 512\) pixels, which was then sent to the network for training and testing. The test set was tested after every epoch of training, only outputting models and results with the highest SE and ACC.

In this experiment, the relationship between the model prediction result and the true label of the data was evaluated according to the following criteria: True Positive, False Negative, False Positive, and True Negative. ACC, SE, SP, receiver operating curve, and AUC were also applied in order to evaluate the experimental results. In clinical settings, a missed diagnosis has a greater adverse effect on patients; hence, the SE in DR classification is more significant. In the DR five-category experiment, the Kappa coefficient was also added as an evaluation criterion (Table 6).

### 3 RESULTS

In this paper, according to the three DR datasets of EyePACS, Messidor, and IDRid, the commonly used DL methods, and the proposed RAN were experimented, respectively. Here, we utilized cross-entropy and focal loss to conduct classification and diagnosis experiments for DR and DME, which were then compared and analyzed, as shown in Tables 7–12.

In the EyePACS dataset, the SP, SE, and AUC of the RAN for DR two classifications reached 0.894 (95% CI, 0.8646–0.9108), 0.930 (95% CI, 0.9047–0.9486), and 0.917 (95% CI, 0.8976–0.9287), respectively, while the ACC reached 0.892 (95% CI, 0.8782–0.9123), which was 4.6%, 3.7%, 9.4%, and 5.6% higher than VGG-16, respectively. Moreover, RAN attained an excellent level of ACC in the DR classification. The ACC of RAN in the five DR classifications reached 0.815 (95% CI, 0.8024–0.8456), which was 5.3% higher than that of VGG-16. Meanwhile, the Kappa score reached 0.865, which was higher than the 0.829 obtained in the DR classification competition held...

---

#### Table 6

| CPU               | i7-9700k 8 cores3.6G |
|-------------------|-----------------------|
| RAM               | Kingston 16G DDR4 2666 |
| Main hard drive   | Samsung 256G M.2      |
| GPU               | RTX 2080ti 11GB       |
| Operating system  | Ubuntu 16.04          |
| Software environment | Python 3.6.2 PyTorch 1.0.0 Scikit-Learn 0.19.2 |

**Abbreviations:** GPU, graphics processing unit; RAM, random access memory.

#### Table 7

| Method         | ACC (95% CI) | SP (95% CI) | SE (95% CI) | AUC (95% CI) |
|----------------|--------------|-------------|-------------|--------------|
| VGG-16         | 0.836        | 0.848       | 0.893       | 0.823        |
| ResNet-50      | 0.844        | 0.862       | 0.907       | 0.876        |
| ResNeXt-50     | 0.866        | 0.867       | 0.919       | 0.892        |
| DenseNet-121   | 0.862        | 0.877       | 0.905       | 0.874        |
| SE-ResNet-50   | 0.889        | 0.889       | 0.924       | 0.896        |
| DPN-92         | 0.868        | 0.876       | 0.918       | 0.878        |
| EfficientNet-b0| 0.886        | 0.882       | 0.925       | 0.903        |
| RAN            | **0.892**    | **0.894**   | **0.930**   | **0.917**    |

**Note:** Bold values denote optimum performance for each project. **Abbreviations:** ACC, accuracy; AUC, area under receiver operating curve; RAN, residual attention network; SE, sensitivity; SP, specificity.

#### Table 8

| Method      | ACC (95% CI) | Kappa score |
|-------------|--------------|-------------|
| VGG-16      | 0.762        | 0.720       |
| ResNet-50   | 0.782        | 0.781       |
| ResNeXt-50  | 0.811        | 0.812       |
| DenseNet-121| 0.803        | 0.856       |
| SE-ResNet-50| 0.808        | 0.834       |
| DPN-92      | 0.802        | 0.827       |
| EfficientNet-b0 | 0.810    | 0.835       |
| RAN         | **0.815**    | **0.865**   |

**Note:** Bold values denote optimum performance for each project. **Abbreviations:** ACC, accuracy; RAN, residual attention network.
on the EyePACS website in 2015,\textsuperscript{28} and the performance was improved.

In the Messidor dataset, the SP and SE of the proposed RAN for the two DR classifications reached 0.887 (95% CI, 0.8695–0.9123) and 0.931 (95% CI, 0.9137–0.9498), while the ACC reached 0.898 (95% CI, 0.8751–0.9275), which was 3.7%, 4.9%, and 7.5% higher than VGG-16, respectively. In addition, the AUC result of EfficientNet was 0.912 (95% CI, 0.8965–0.9231), which was 0.5% higher than RAN. The ACC, SE, and AUC of RAN for the two DME classifications were 0.881 (95% CI, 0.8693–0.8964), 0.912 (95% CI, 0.8952–0.9413), and 0.882 (95% CI, 0.8697–0.9058), respectively, which were 4.2%, 4%, and 4.4% higher than VGG-16, respectively. The SP

| Method        | ACC (95% CI) | SP (95% CI) | SE (95% CI) | AUC (95% CI) |
|---------------|--------------|-------------|-------------|--------------|
| VGG-16        | 0.823        | 0.850       | 0.882       | 0.830        |
| ResNet-50     | 0.861        | 0.875       | 0.892       | 0.887        |
| ResNeXt-50    | 0.868        | 0.864       | 0.904       | 0.893        |
| DenseNet-121  | 0.854        | 0.878       | 0.895       | 0.864        |
| SE-ResNet-50  | 0.858        | 0.884       | 0.914       | 0.894        |
| DPN-92        | 0.864        | 0.876       | 0.908       | 0.869        |
| EfficientNet-b0 | 0.884    | 0.881       | 0.923       | 0.912        |
| RAN           | **0.898**    | **0.887**   | **0.931**   | **0.907**    |

Note: Bold values denote optimum performance for each project.

Abbreviations: ACC, accuracy; AUC, area under receiver operating curve; DR, diabetic retinopathy; RAN, residual attention network; SE, sensitivity; SP, specificity.

| Method        | ACC (95% CI) | SP (95% CI) | SE (95% CI) | AUC (95% CI) |
|---------------|--------------|-------------|-------------|--------------|
| VGG-16        | 0.839        | 0.824       | 0.872       | 0.838        |
| ResNet-50     | 0.852        | 0.833       | 0.876       | 0.851        |
| ResNeXt-50    | 0.861        | 0.833       | 0.883       | 0.858        |
| DenseNet-121  | 0.852        | 0.836       | 0.882       | 0.844        |
| SE-ResNet-50  | 0.866        | 0.843       | 0.897       | 0.863        |
| DPN-92        | 0.853        | 0.832       | 0.889       | 0.858        |
| EfficientNet-b0 | 0.871    | **0.862**   | 0.904       | 0.880        |
| RAN           | **0.881**    | 0.858       | **0.912**   | **0.882**    |

Note: Bold values denote optimum performance for each project.

Abbreviations: ACC, accuracy; AUC, area under receiver operating curve; DR, diabetic retinopathy; RAN, residual attention network; SE, sensitivity; SP, specificity.

| Method        | ACC (95% CI) | SP (95% CI) | SE (95% CI) | AUC (95% CI) |
|---------------|--------------|-------------|-------------|--------------|
| VGG-16        | 0.678        | 0.684       | 0.724       | 0.667        |
| ResNet-50     | 0.693        | 0.705       | 0.734       | 0.684        |
| ResNeXt-50    | 0.702        | 0.714       | 0.748       | 0.695        |
| DenseNet-121  | 0.694        | 0.706       | 0.736       | 0.685        |
| SE-ResNet-50  | 0.701        | 0.716       | 0.745       | 0.711        |
| DPN-92        | 0.692        | 0.708       | 0.734       | 0.702        |
| EfficientNet-b0 | 0.704    | 0.715       | 0.747       | 0.715        |
| RAN           | **0.715**    | **0.728**   | **0.752**   | **0.726**    |

Note: Bold values denote optimum performance for each project.

Abbreviations: ACC, accuracy; AUC, area under receiver operating curve; DR, diabetic retinopathy; RAN, residual attention network; SE, sensitivity; SP, specificity.
TABLE 12  DME two classifications results of different methods on the IDRid dataset

| Method      | ACC (95% CI) | SP (95% CI) | SE (95% CI) | AUC (95% CI) |
|-------------|--------------|-------------|-------------|--------------|
| VGG-16      | 0.702        | 0.714       | 0.748       | 0.693        |
| ResNet-50   | 0.725        | 0.736       | 0.768       | 0.714        |
| ResNeXt-50  | 0.746        | 0.749       | 0.785       | 0.738        |
| DenseNet-121| 0.735        | 0.740       | 0.779       | 0.737        |
| SE-ResNet-50| 0.742        | 0.751       | 0.792       | 0.736        |
| DPN-92      | 0.728        | 0.742       | 0.787       | 0.718        |
| EfficientNet-b0 | 0.756    | 0.763       | 0.807       | 0.739        |
| RAN         | **0.785**    | **0.796**   | **0.813**   | **0.772**    |

Note: Bold values denotes optimum performance for each project. Abbreviations: ACC, accuracy; AUC, area under receiver operating curve; DME, diabetic macular edema; RAN, residual attention network; SE, sensitivity; SP, specificity.

FIGURE 9  Comparison of the classification results of different methods with different loss functions on the EyePACS, Messidor, and IDRid dataset
of EfficientNet was 0.862 (95% CI, 0.8424–0.8762), which was 0.4% higher than RAN and demonstrated its good performance.

In the IDRid dataset, the ACC, SP, SE, and AUC of RAN for the two DR classifications reached 0.715 (95% CI, 0.6941–0.7423), 0.728 (95% CI, 0.6952–0.7461), 0.752 (95% CI, 0.7234–0.7883), and 0.726 (95% CI, 0.6882–0.7548). According to Table 12, the ACC, SP, SE, and AUC of the RAN for the IDRid dataset of the two DME classifications reached 0.785 (95% CI, 0.7658–0.8044), 0.796 (95% CI, 0.775–0.8227), 0.813 (95% CI, 0.6997–0.8341), and 0.772 (95% CI, 0.7529–0.8195). Compared to other methods, the performance of RAN in each index in the IDRid dataset improved by about 1–10 points.

As seen in Figure 9, due to the imbalance problem in the DR datasets, focal loss was more suitable than cross entropy as the loss function in each classification task. Accordingly, the ACC was noted to be greatly improved.

4 | DISCUSSION

This paper proposed a classification algorithm RAN for DR detection, in which the classification experiments were verified on the EyePACS, Messidor, and IDRid datasets. Since the imbalance between data categories will lead to overfitting during model training, data augmentation, and focal loss were introduced. The image augmentation method used in this experiment can make the amount of data in each classification of DR reach a relatively balanced state. Moreover, the focal loss also achieved satisfactory results in alleviating data imbalance issues. In order to address the minor differences between DR categories, normalization was also carried out on the original retinal image to highlight the bleeding and exudation in the fundus image. In addition, the attention mechanism, which focuses on features in the fine-grained image during classification, was added to the network in order to extract the features of the fine-grained images so that the network can better distinguish the differences between the types of lesions. Furthermore, dilated convolution was added to the network to increase the receptive field. The above results demonstrated that the intense competitiveness of CNN in clinical diagnostic applications and RAN achieved better performance in DR detection. In short, using the proposed RAN can enhance the ACC of DR classification and diagnosis of most fundus images. Through this combination of ResNet, attention mechanism, and dilated convolution, the ACC of classification of DR can be improved.

However, the rise in ACC of the proposed method is not significant enough. In our future studies, we will integrate additional information related to DR, such as age, blood glucose, blood pressure, intraocular pressure, and past history, into the DR classification model to effectively improve the diagnosis results. Moreover, multi-task experiments will be conducted to mutually promote the improvement of the experimental results. How to integrate the results of exudates, bleeding, microaneurysms detection, and blood vessel segmentation into the DR classification model will also be the focus of our subsequent works. Algorithm engineers and clinicians both aspire to build a robust and accurate DL model for DR detection, and this desire cannot be achieved without the joint efforts and cooperation of both parties.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (81971708) and the National Key Research and Development Program of China (2018YFC1314900, 2018YFC1314902).

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in References 28–30.

ORCID

Moye Yu https://orcid.org/0000-0002-3882-8810
Yi Wang https://orcid.org/0000-0002-6402-3637

REFERENCES

1. Matuszewski W, Bandurska-Stankiewicz E, Modzelewski R, Kamińska U, Stefanowicz-Rutkowska M. Diagnosis and treatment of diabetic retinopathy—historical overview. Clin Diabetol. 2017;6:182-188.
2. De Sonnaville JJJ, van der Sloot DF, Ernst L, Wijkel D, Heine RJ. Retinopathy screening in type 2 diabetes: reliability of wide angle fundus photography. Diabet Med. 1996;13:482-486.
3. Yau JKY, Rogers SL, Kawasaki R, et al. Global prevalence and major risk factors of diabetic retinopathy. Diabetes Care. 2012;35:556-564.
4. Chandrakumar T, Kathirvel R. Classifying diabetic retinopathy using deep learning architecture. Int J Eng Res Technol. 2016;5:19-24.
5. Wu Y, Xia Y, Song Y, Zhang D, Liu D, Zhang C, Cai W. Vesselnet: retinal vessel segmentation under multi-path supervision. Paper presented at: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI); 2019:11764.
6. Effekhari N, Pourreza HR, Masoudi M, Ghiasi-Shirazi K, Saeedi E. Microaneurysm detection in fundus images using a two-step convolutional neural network. Biomed Eng. 2019;18:1-16.
7. Pratt H, Coenen F, Broadbent DM, Harding SP, Zheng Y. Convolutional neural networks for diabetic retinopathy. Procedia Comput Sci. 2016;90:200-205.
8. Rahim SS, Palade V, Shuttleworth J, Jayne C. Automatic screening and classification of diabetic retinopathy and maculopathy using fuzzy image processing. *Brain Inf.* 2016;3:249-267.

9. Seth S, Agarwal B. A hybrid deep learning model for detecting diabetic retinopathy. *J Stat Manag Syst.* 2018;21:569-574.

10. Dutta S, Manideep BC, Bash SM, Caytles RD, Iyengar NCSN. Classification of diabetic retinopathy images by using deep learning models. *Int J Distrib Comput Sci Eng.* 2018;11:89-106.

11. Adem K. Exudate detection for diabetic retinopathy with circular Hough transformation and convolutional neural networks. *Expert Syst Appl.* 2018;114:289-295.

12. Li Z, He Y, Keel S, Meng W, Chang RT, He M. Efficacy of a deep learning system for detecting glaucomatous optic neuropathy based on color fundus photographs. *Ophthalmology.* 2018;125:1199-1206.

13. Wu Z, Shi G, Chen Y, et al. Coarse-to-fine classification for diabetic retinopathy grading using convolutional neural network. *Artif Intell Med.* 2020;108:101936.

14. Arenas-Cavalli JT, Abarca I, Rojas-Contreras M, Bernuy F, Donoso R. Clinical validation of an artificial intelligence-based diabetic retinopathy screening tool for a national health system. *Eye.* 2021;36:1-8.

15. Dai L, Wu L, Li H, et al. A deep learning system for detecting diabetic retinopathy across the disease spectrum. *Nat Commun.* 2021;12:1-11.

16. Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA.* 2016;316:2402-2410.

17. Li X, Pang T, Xiong B, Liu W, Liang P, Wang T. Convolutional neural networks based transfer learning for diabetic retinopathy fundus image classification. Paper presented at: International Congress on Image and Signal Processing; 2017:1-11.

18. Gargeya R, Leng T. Automated identification of diabetic retinopathy using deep learning. *Ophthalmology.* 2017;124:962-969.

19. Abrámoff MD, Lou Y, Erginay A, et al. Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. *Investig Ophthalmol Vis Sci.* 2016;57:5200-5206.

20. Ting DSW, Cheung CYL, Lim G, et al. 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.

21. Li Z, Keel S, Liu C, et al. An automated grading system for detection of vision-threatening referable diabetic retinopathy on the basis of color fundus photographs. *Diabetes Care.* 2018;41:2509-2516.

22. Abrámoff 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;1:1-8.

23. Wang Y, Wang G A, Fan W, Li J. A deep learning based pipeline for image grading of diabetic retinopathy. Paper presented at: International Conference on Smart Health; 2018: 240-248.

24. Zhou K, Gu Z, Liu W, Luo W, Cheng J, Gao S, Liu J. Multi-cell multi-task convolutional neural networks for diabetic retinopathy grading. Paper presented at: International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); 2018:2724-2727.

25. Doshi D, Shenoy A, Sidhpura D, Gharpure P. Diabetic retinopathy detection using deep convolutional neural networks. Paper presented at: International Conference on Computing Analytics and Security Trends; 2016: 261-266.

26. Alyoubi WL, Abulkhair MF, Shalash WM. Diabetic retinopathy fundus image classification and lesions localization system using deep learning. *Sensors.* 2021;21:3704.

27. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Paper presented at: IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016:770-778.

28. https://www.kaggle.com/c/diabetic-retinopathy-detection/data

29. https://www.acdis.net/en/third-party/messidor/

30. https://ieeexplore.ieee.org/portal/open-access/indian-diabetic-retinopathy-image-dataset-ridr

31. Schmidhuber J. Deep learning in neural networks: an overview. *Neural Netw.* 2015;61:85-117.

32. Lin T Y, Goyal P, Girshick R, He K, Dollár P. Focal loss for dense object detection. Paper presented at: IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2017:2980-2988.

33. Wang F, Jiang M, Qian C, Yang S, Li C, Zhang H, Wang X, Tang X. Residual attention network for image classification. Paper presented at: IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2017: 3156-3164.

34. Wang Z, Yin Y, Shi J, Fang W, Li H, Wang X. Zoom-in-net: deep mining lesions for diabetic retinopathy detection. Paper presented at: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI); 2017: 267-275.

35. Yu F, Kolton V, Funkhouser T. Dilated residual networks. Paper presented at: IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2017:472-480.

36. Pan SJ, Yang Q. A survey on transfer learning. *IEEE Trans Knowl Data Eng.* 2009;22:1345-1359.

37. Xie S, Girshick R, Dollar P, Tu Z, He K. Aggregated residual transformations for deep neural networks. Paper presented at: IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2017:1492-1500.

38. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. Paper presented at: IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2017:4700-4708.

39. Hu J, Shen L, Sun G. Squeeze-and-excitation networks. Paper presented at: IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2018:7132-7141.

40. Chen Y, Li J, Xiao H, Jin X, Yan S, Feng J. Dual path networks. *Adv. Neural Inf. Process. Syst.* 2017: 30. https://arxiv.org/abs/1707.01629

41. Tan M, Le Q. Efficientnet: rethinking model scaling for convolutional neural networks. Paper presented at: International Conference on Machine Learning (PMLR); 2019:6105-6114.

**How to cite this article:** Yu M, Wang Y. Intelligent detection and applied research on diabetic retinopathy based on the residual attention network. *Int J Imaging Syst Technol.* 2022;32(5):1789-1800. doi:10.1002/ima.22734