



CSANet: a lightweight channel and spatial attention neural network for grading diabetic retinopathy with optical coherence tomography angiography

Fei Ma¹, Xiao Liu¹, Shengbo Wang¹, Sien Li¹, Cuixia Dai², Jing Meng¹

¹School of Computer Science, Qufu Normal University, Rizhao, China; ²College Science, Shanghai Institute of Technology, Shanghai, China

Contributions: (I) Conception and design: F Ma, X Liu; (II) Administrative support: F Ma, J Meng; (III) Provision of study materials or patients: J Meng, C Dai; (IV) Collection and assembly of data: C Dai; (V) Data analysis and interpretation: X Liu, S Wang, S Li; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Fei Ma, PhD; Jing Meng, PhD. School of Computer Science, Qufu Normal University, No. 80 Yantai Road, Rizhao 276826, China. Email: mafei0603@163.com; jingmeng@qfnu.edu.cn.

Background: Diabetic retinopathy (DR) is one of the most common eye diseases. Convolutional neural networks (CNNs) have proven to be a powerful tool for learning DR features; however, accurate DR grading remains challenging due to the small lesions in optical coherence tomography angiography (OCTA) images and the small number of samples.

Methods: In this article, we developed a novel deep-learning framework to achieve the fine-grained classification of DR; that is, the lightweight channel and spatial attention network (CSANet). Our CSANet comprises two modules: the baseline model, and the hybrid attention module (HAM) based on spatial attention and channel attention. The spatial attention module is used to mine small lesions and obtain a set of spatial position weights to address the problem of small lesions being ignored during the convolution process. The channel attention module uses a set of channel weights to focus on useful features and suppress irrelevant features.

Results: The extensive experimental results for the OCTA-DR and diabetic retinopathy analysis challenge (DRAC) 2022 data sets showed that the CSANet achieved state-of-the-art DR grading results, showing the effectiveness of the proposed model. The CSANet had an accuracy rate of 97.41% for the OCTA-DR data set and 85.71% for the DRAC 2022 data set.

Conclusions: Extensive experiments using the OCTA-DR and DRAC 2022 data sets showed that the proposed model effectively mitigated the problems of mutual confusion between DRs of different severity and small lesions being neglected in the convolution process, and thus improved the accuracy of DR classification.

Keywords: Attention module; diabetes retinopathy; grading diabetic retinopathy (grading DR); optical coherence tomography angiography diabetic retinopathy data set (OCTA-DR data set)

Submitted Sep 05, 2023. Accepted for publication Dec 12, 2023. Published online Jan 23, 2024.

doi: 10.21037/qims-23-1270

View this article at: <https://dx.doi.org/10.21037/qims-23-1270>

Introduction

Diabetic retinopathy (DR) (1-5) is one of the most common eye diseases caused by diabetes. DR can cause vision loss and even blindness in the working age population worldwide (6). Fortunately, DR can be prevented and controlled. However, to minimize its damage, DR needs to be screened and detected early (7,8).

There are many ways to screen for DR, including fluorescein angiography (FA), fundus photography (FP), optical coherence tomography (OCT), and optical coherence tomography angiography (OCTA) (9). FA is an important medical imaging modality for the evaluation of DR; however, it is invasive, time consuming, and cumbersome (10). FP captures images of the inside of the eye through the pupil and can be used to examine the optic disc, retina, and lens. FP is a non-invasive technique that takes only one minute to administer. It enables doctors to observe subtle changes in the eye and recommend useful treatments for eye diseases (11). However, conventional FP cannot be reliably used to identify microvascular abnormalities that occur in the early stages of ocular diseases (12-14). OCT is a new non-invasive imaging technique that can be used to effectively observe subtle changes in the superficial and deep capillary plexus of the

human retinal microvasculature and has become popular in recent years (15). As an extension of OCT, OCTA is used to capture and analyze the movement of blood cells in the field of view by repeatedly acquiring images of the same retinal location to obtain an image of the capillary network (16). Numerous studies have shown that OCTA has many advantages over traditional imaging modalities, such as FP or FA, in the detection and diagnosis of various ocular diseases (17). *Figure 1* shows typical fundus images; the images in the top rows are representative of fundus images taken with conventional color fundus cameras; while the images in the bottom rows are representative of images taken with a swept-source OCTA camera.

Sandhu *et al.* (18) introduced a computer aided design system based on a random-forest classifier that was fed features extracted from OCT and OCTA images. Ramasamy *et al.* (19) extracted and fused retinal features from retinal images based on texture gray-level features and Ridgelet transform coefficients, and then used the sequential minimal optimization classification method to classify DR based on the retinal features obtained. The method achieved 97.05% accuracy on the DIARETDB1 data set and 91.0% accuracy on the KAGGLE data set. Abdelsalam *et al.* (20) developed a support vector machine-based model with multifractal geometry and lacunarity parameters to diagnose DR using

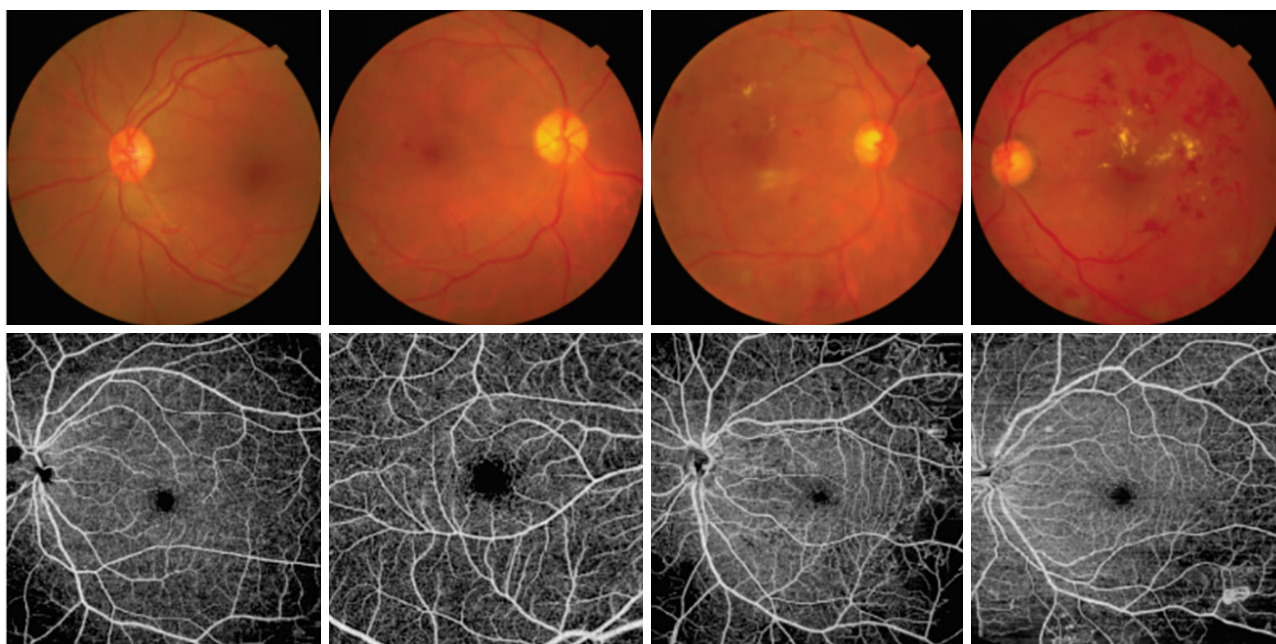


Figure 1 Typical retinal images. Top row: typical color fundus camera images. Bottom row: typical en-face optical coherence tomography angiography images.

OCTA images. Maqsood *et al.* (21) developed a new macular detection system based on contrast enhancement, top-hat transformation, and a modified Kirsch template method, which achieved state-of-the-art performance compared to other mainstream methods.

It takes a great deal of effort to manually extract features using machine-learning algorithms; however, deep-learning methods learn image features automatically during training. Recently, convolutional neural networks (CNNs) have been shown to be a powerful tool for learning features for DR (22–26). Chaurasia *et al.* (27) introduced an ensemble model for DR disease detection using transfer learning. Zang *et al.* (28) introduced a deep CNN called DcardNet with adaptive label smoothing to suppress overfitting using en-face OCT and OCTA images. Maqsood *et al.* (29) used a method that combined deep learning and machine learning to detect whether retinal fundus images were hemorrhagic and thus to determine whether patients suffer from DR. They conducted experiments on 1,509 images from five data sets and reported that their model had an average accuracy rate of 97.71%. Dong *et al.* (30) designed a fused network based on two networks [Inception-V3 and VGG16 (visual geometry group)] to improve the accuracy of the model. Ouyang *et al.* (31) introduced a contrastive self-learning algorithm that was first pre-trained with unlabeled retinal images using a convolutional network-based encoder, and then re-trained with small-scale annotated training data using a classifier to detect referable DR. Ryu *et al.* (32) developed a fully automated system based on the CNN model for early detection of DR using OCTA images. Durai *et al.* (33) developed a deformable ladder bi-attention U-shaped encoder-decoder network and deep adaptive CNN to classify DR. Tang *et al.* (34) designed an ordinal regularized module to represent the orderliness of disease severity that could be flexibly embedded into general classification networks. The above deep-learning methods have shown to be effective methods for grading DR and will be helpful for researchers and patients alike.

The CNN-based DR grading methods have achieved good performance; however, they still face challenges in grading DR tasks. In clinical practice, there are some differences between DR diseases with adjacent grades. Additionally, the lesions in DR images are relatively small. Thus, we sought to develop a fusion-attention module based on channel attention and spatial attention to obtain the discriminative features needed for fine-grained DR classification while suppressing the irrelevant features.

We used bottleneck blocks and skip connections in

the model to ensure the classification performance of the network while reducing the number of parameters and addressing the overfitting problem. Overfitting is a typical problem in computer vision applications (35–39). It is a problem that occurs in the training of CNNs and is caused by a lack of training data or the complexity of the network (40). Our main contributions can be summarized as follows:

- (I) We designed a new spatial attention block (SPAB) to obtain a set of spatial weights to alleviate the problem of small lesions being ignored during the convolution process.
- (II) We developed a novel channel attention module to explore the relationship in different channels of the proposed model that can learn a set of channel weights to focus on useful features and suppress irrelevant features.
- (III) We introduced a novel plug-and-play fused attention module to integrate the advantages of the spatial and channel attention modules.
- (IV) We performed extensive experiments on Dong's DR data set (the OCTA-DR data set). The extensive experimental results for the OCTA-DR data set showed that the channel and spatial attention network (CSANet) achieved state-of-the-art DR classification results.

Methods

Figure 2 shows the architecture of our proposed CSANet. The CSANet comprises two parts: (I) the hybrid attention module (HAM); and (II) the backbone network. The spatial attention module determines the spatial location weights, and the channel attention module obtains the feature channel weights. High weights can achieve good performance in grading DR tasks. As Figure 2 shows, our CSANet takes OCTA images as input, and outputs DR grades in an end-to-end manner.

In the following sections, we first introduce the attention modules and then describe our training and testing strategies for grading DR. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

Attention block

The attention block comprises the bottleneck layer, the attention-based module, the skip-connection layer, and the

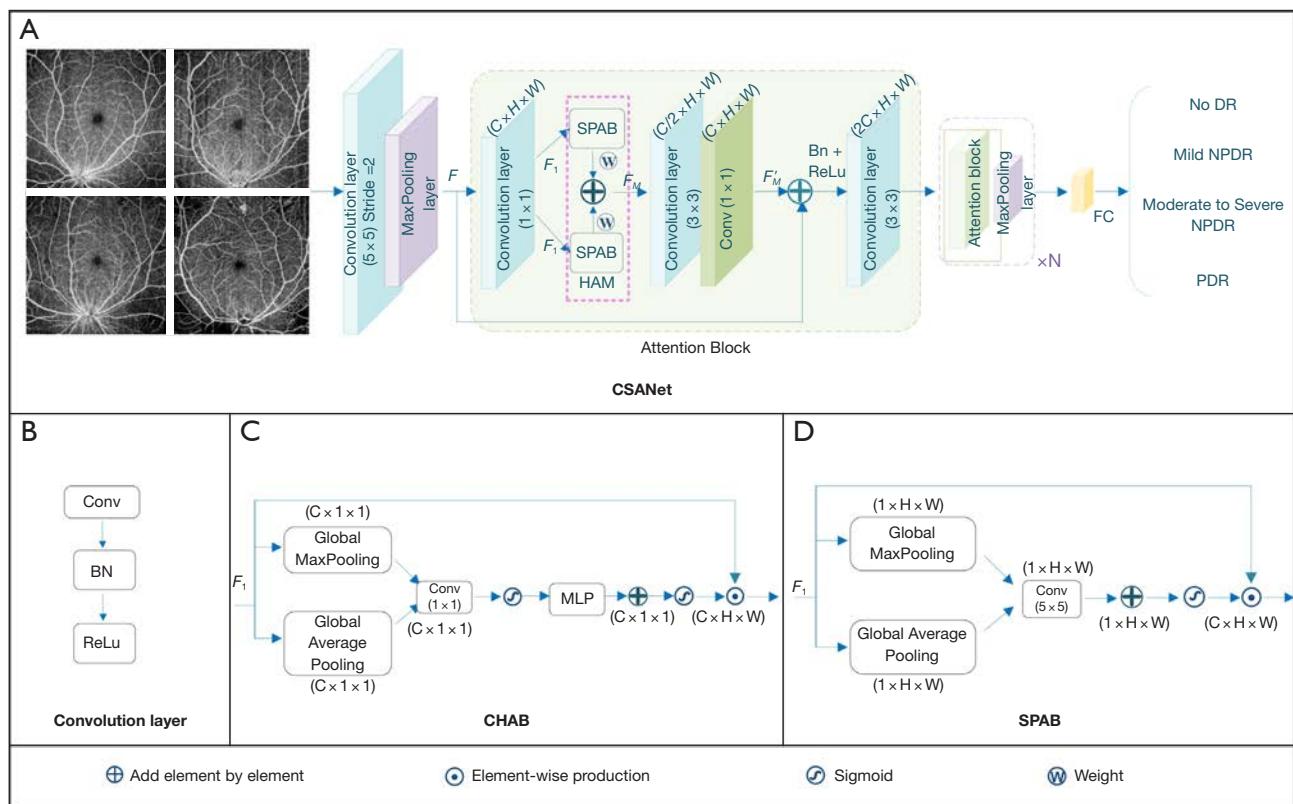


Figure 2 Illustration of our proposed CSANet. (A) The flowchart of the CSANet. (B) The convolutional layer. (C) The structure of the CHAB. (D) The structure of the SPAB. CSANet, channel and spatial attention network; CHAB, channel attention block; SPAB, spatial attention block; HAM, hybrid attention module; FC, fully connected layers; DR, diabetic retinopathy; NPDR, non-proliferative DR; PDR, proliferative DR; BN, batch normalization; MLP, multi-layer perception.

convolutional layer. First, the input feature map is denoted as $F \in \mathbb{R}^{C \times H \times W}$, where C is the number of channels, and H and W are the height and width of the feature map, respectively. To improve the computational performance of the model, the number of channels of F is halved by adding a convolutional layer. Here, the feature map is denoted as $F_1 \in \mathbb{R}^{C/2 \times H \times W}$. F_1 is then fed into the HAM module, which can obtain the spatial attention feature map and the channel attention feature map. The obtained spatial and channel attention feature maps are each weighted, respectively, and the feature map F_M is computed by element-wise addition to achieve a better classification effect. Next, the channel number of F_M is restored to C by a convolutional layer and a convolutional operation. We then obtain $F'_M \in \mathbb{R}^{C \times H \times W}$. The high and low semantic features are merged by the skip connections. Finally, in the last convolutional layer, the number of channels in the feature map becomes $2C$.

Channel attention block (CHAB)

Unlike the squeeze-and-excitation (41) attention mechanism, the efficient-channel-attention (ECA) (42) mechanism avoids dimensionality reduction and uses a one-dimensional convolution to efficiently implement local cross-channel interaction and extract inter-channel dependencies. The ECA mechanism first performs the global average pooling of the input feature map, then performs a one-dimensional convolution operation with a convolution kernel size of k , obtains the weights of each channel through the sigmoid activation function, and finally multiplies the weights by the corresponding elements of the original input feature map to obtain the final output feature map. The CHAB structure proposed in this article not only ensures the classification performance of the network while reducing the model parameter increase as much as possible, but also overcomes the shortcomings of the ECA structure

without considering the global channel correlation. The CHAB structure is shown in *Figure 2C*.

The feature map F_l is input to the CHAB module. Two feature vectors (i.e., p and q) can be obtained using the global average pooling and global maximum pooling operations for each channel as follows:

$$\begin{aligned} p &= \text{GAP}(F_l) \\ q &= \text{GMP}(F_l) \end{aligned} \quad [1]$$

where $\text{GAP}(\cdot)$ is the global average pooling operation, and $\text{GMP}(\cdot)$ is the global maximum pooling operation. The local channel relation features p' and q' can be obtained by the adaptive one-dim convolutional operation as follows:

$$\begin{aligned} p' &= \text{Sigm}(\text{Conv}(p)) \\ q' &= \text{Sigm}(\text{Conv}(q)) \end{aligned} \quad [2]$$

where $\text{Conv}(\cdot)$ is the convolution operation, and $\text{Sigm}(\cdot)$ is the sigmoid activation function.

Since the multi-layer perceptron (MLP) is a true global attention, the global channel-related features r and s are obtained from all local channel-related features p' and q' by the MLP. The global attention feature F_c is expressed as follows:

$$F_c = (\text{Sigm}(r \oplus s)) \odot F_l \quad [3]$$

where the \oplus operation denotes the element-wise addition, and \odot denotes the channel-wise weighting operation.

SPAB

Figure 2D shows the structure of the SPAB. The learned spatial position weights represent the importance of different spatial locations. Specifically, two weight vectors with $[1 \times H \times W]$ can be obtained from the global average pooling and global max pooling layers. The spatial attention feature A_c is obtained via a convolutional layer by merging two weight vectors with element-wise addition. Spatial attention W_s is expressed as follows:

$$W_s = \sigma(\text{Conv}(F_{\text{avg}}^S) \oplus \text{Conv}(F_{\text{max}}^S)) \quad [4]$$

where $\sigma(\cdot)$ is the activation function. A_c is multiplied by F_l on an element-by-element basis to generate a spatial attention feature map. The SPAB can capture the most important semantic information of the samples for grading DR tasks, mitigating the problem of small lesions being missed by convolution.

Results

In this section, we introduce two DR data sets [i.e., the OCTA-DR data set (30) and the diabetic retinopathy analysis challenge (DRAC) 2022 data set (43)], and the experimental settings and evaluation metrics, and then present the qualitative and quantitative results of the competing methods for the two DR data sets.

Data sets

Dong's OCTA-DR data set

The OCTA-DR data set comprised OCTA fundus images of 288 diabetic and 97 healthy individuals that were obtained using a swept-source OCT system with a 12 mm \times 12 mm single scan centered on the fovea (this data set is available at <https://kyanbis.github.io/OCTADR>). All the OCTA images were graded by two ophthalmologists (30). The size of each original image was 299 \times 299 pixels. Due to the similar clinical manifestations and consistent recommended treatment methods between moderate and severe non-proliferative DR (NPDR), two professional ophthalmologists graded these images into the following four categories based on the Early Treatment of Diabetic Retinopathy Study: (I) no DR; (II) mild NPDR; (III) moderate-to-severe NPDR; and (IV) proliferative DR (PDR).

As *Figure 3A-3F* shows, compared with normal eyes, mild NPDR eyes had a small amount of non-perfusion and microvascular tumors in the wide-field OCTA (WF-OCTA) images (44). As DR progresses from moderate to severe NPDR, the number of non-perfusion areas and microaneurysms increases, and blood vessels become distorted and dilated (45) (*Figure 3G-3I*). During PDR, ocular ischemia and hypoxia worsen, and new blood vessels are formed (46,47) (*Figure 3J-3L*).

To address the overfitting problem in the training process, we augmented samples using the same method as Dong *et al.* (30) and also normalized the data before augmentation. The number of augmented images was 2,693. *Table 1* shows the image distribution of the OCTA-DR data set.

DRAC 2022 data set

The DRAC was designed to provide a benchmark for evaluating the algorithms used to automatically analyze

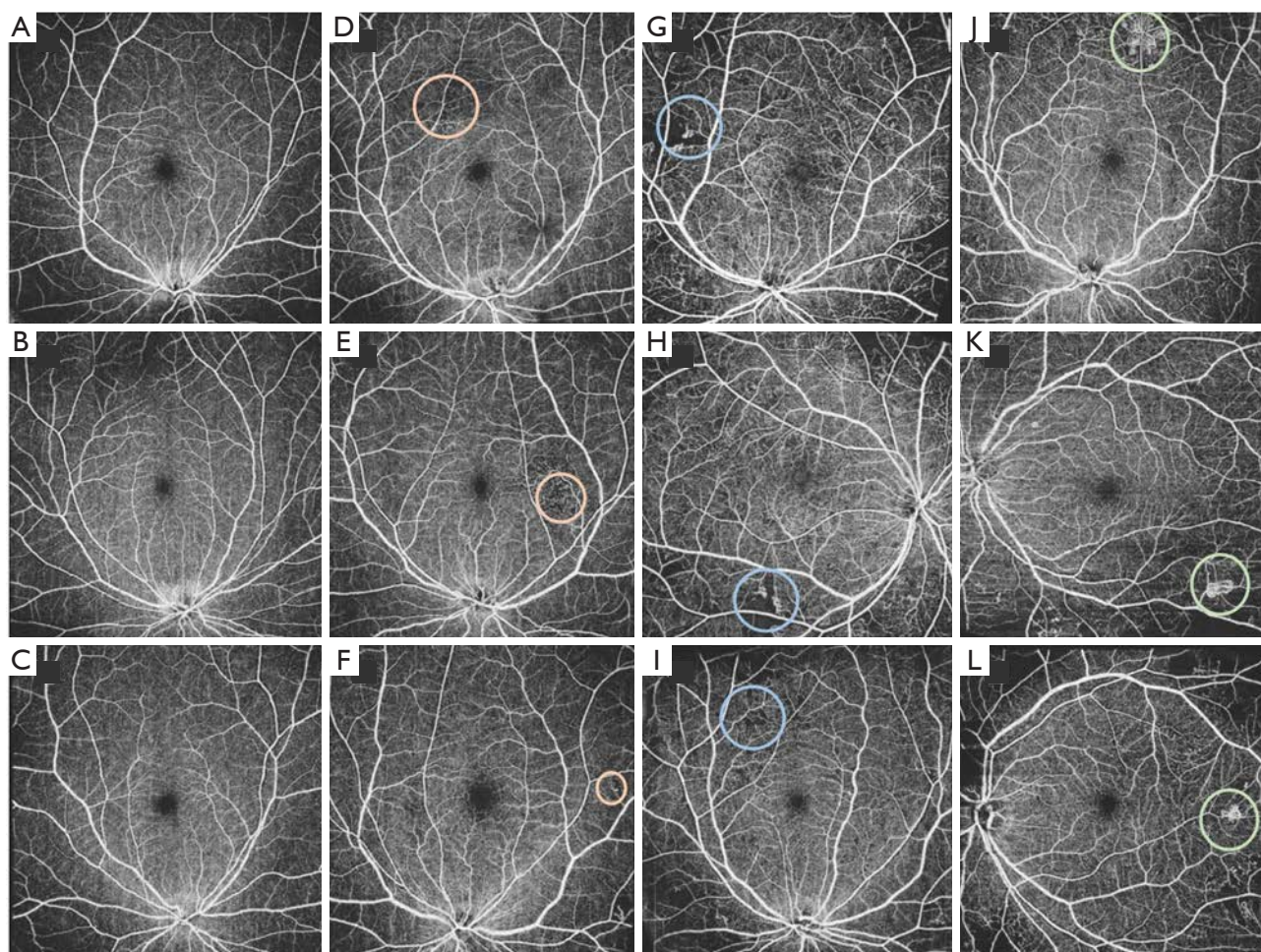


Figure 3 Representative optical coherence tomography angiography images of different severities of DR. (A) Normal sample 1. (B) Normal sample 2. (C) Normal sample 3. (D) Mild DR sample 1. (E) Mild DR sample 2. (F) Mild DR sample 3. (G) Moderate-to-severe DR sample 1. (H) Moderate-to-severe DR sample 2. (I) Moderate-to-severe DR sample 3. (J) Proliferative DR specimen 1. (K) Proliferative DR specimen 2. (L) Proliferative DR specimen 3. The orange circles indicate microaneurysms and areas of capillary non-perfusion. The blue circles indicate large areas of microaneurysms and intraretinal microvascular abnormalities. The green circle indicates neo angiogenesis. DR, diabetic retinopathy.

Table 1 Information for the OCTA-DR data set

Severities	Number
No DR	615
Mild NPDR	704
Moderate-to-severe NPDR	706
PDR	668

OCTA, optical coherence tomography angiography; DR, diabetic retinopathy; NPDR, non-proliferative DR; PDR, proliferative DR.

DR using ultra-wide OCTA images (43) (available at <https://drac22.grand-challenge.org>). The challenge was divided into three separate tasks as follows—task 1: the DR lesion segmentation task, which comprised 109 training images showing three types of lesions (i.e., an intraretinal microvascular abnormality, a non-perfusion area, and neovascularization), and 65 test images; task 2: the image quality assessment task, which comprised 665 training images of three different levels (i.e., poor, good,

Table 2 Information on the DRAC 2022 data set

Severities	Number
No DR	656
NPDR	639
PDR	420

DRAC, diabetic retinopathy analysis challenge; DR, diabetic retinopathy; NPDR, non-proliferative DR; PDR, proliferative DR.

and excellent) and 438 test images; task 3: the DR grading task, which comprised 611 training images, which were a subset of the task 2 training images divided into three classes (i.e., normal, NPDR and PDR), and 386 test images. No expert annotations were available for the participants. All the images had a resolution of 1,024×1,024 pixels; however, the images were re-sized to 512×512 pixels in this implementation. For the data in the DR grading task, we used horizontal flip for label 0, horizontal flip, and vertical flip for label 1, horizontal flip, vertical flip, rotation, and blur for label 2 to enhance the data. The number of enhanced images was 1,715. For this study, 80% of the images were used for training, and the remaining 20% were used for testing. The image distribution in the DRAC 2022 data set is shown in *Table 2*.

Implementation details

To examine the performance of our proposed model, we conducted experiments on two different DR data sets ($N=4$, where N is the number of attention block and maximum pooling layers). We used the Adaptive Moment Estimation (Adam) method to train the model for fast convergence. Of the images, 80% were used for training and the remaining 20% were used for testing. Cross-entropy loss was used as the loss function. The weights γ_1 and γ_2 of channel attention and spatial attention in the HAM module were set to 0.55 and 0.45, respectively. The model was trained for 80 epochs, and the learning rate, betas, and epsilon were set to 10^{-2} (0.9, 0.999), and $1e-8$, respectively. Of these, the betas were the momentum parameter in the Adam algorithm, and epsilon was used to maintain numerical stability. When performing the classification tasks on the OCTA-DR data set, the batch size was set to 8. When the model training process reached 3/4 on the OCTA-DR data set, the learning rate was reduced to 1/10 of the original learning rate. When performing classification tasks on the DRAC 2022 data set, the batch size was set to 4. When the model training

process reached 2/5 on the DRAC 2022 data set, the learning rate was reduced to 1/10 of the original learning rate. In addition, the benchmark model also underwent sufficient iterative training to achieve the best performance. We used Pytorch to implement the CSANet on a Window 10 workstation with an NVIDIA RTX 3090Ti (Santa Clara, CA, USA) with 24 GB of Graphics Processing Unit memory.

Evaluation metrics

To evaluate the performance of the proposed method, we used four evaluation metrics; that is, accuracy, precision, the F1-score, and the kappa coefficient. Accuracy, which represents the ratio of the number of correct predictions for classification to the total number of predictions is the most commonly used metric in classification tasks, and is expressed as:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad [5]$$

where TP (true positive) is the number of samples correctly predicted as positive examples; FN (false negative) is the number of samples correctly predicted as negative examples; FP (false positive) is the number of samples correctly predicted as positive examples; TN (true negative) is the number of samples correctly predicted as negative examples.

Precision represents the degree of prediction accuracy in the results of the correct sample, and is expressed as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad [6]$$

Recall is the ratio of being predicted as a positive sample to actual positive samples, and is expressed as:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad [7]$$

The F1-score is based on the harmonic mean of precision and recall, and is expressed as:

$$\text{F1-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad [8]$$

The kappa coefficient is used for consistency testing and can also be used to measure classification accuracy, and is expressed as:

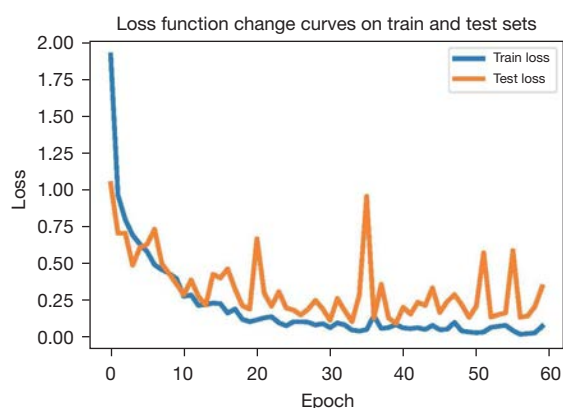
$$k = \frac{p_o - p_e}{1 - p_e} \quad [9]$$

Table 3 Comparison of accuracy, precision and F1-score of DR severity

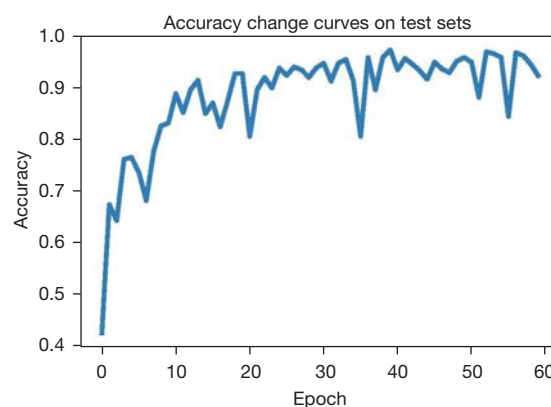
Metrics	No DR	Mild NPDR	Moderate-to-severe NPDR	PDR
Accuracy	0.9767	0.9444	1.0000	0.9774
Precision	0.9474	0.9855	0.9640	1.0000
F1-score	0.9618	0.9645	0.9817	0.9886

DR, diabetic retinopathy; NPDR, non-proliferative DR; PDR, proliferative DR.

	No DR	Mild NPDR	Moderate and severe NPDR	PDR
No DR	126	7	0	0
Mild NPDR	1	136	0	1
Moderate and severe NPDR	2	1	134	2
PDR	0	0	0	130

Figure 4 Confusion matrix for grading DR in the proposed model. DR, diabetic retinopathy; NPDR, non-proliferative DR; PDR, proliferative DR.**Figure 5** The loss curve of the proposed model.

where p_o is the sum of the number of samples correctly classified for each class divided by the total number of samples (i.e., the overall classification accuracy), and p_e is the sum of the “product of the actual and predicted

**Figure 6** The accuracy curve of the proposed model.

number” corresponding to all categories divided by the “square of the total number of samples”.

The results for the three metrics (i.e., accuracy, precision, and the F1-score) for the OCTA-DR data set are presented in *Table 3*. Due to the small non-perfusion areas caused by ischemia, there were small differences between the healthy OCTA images and the OCTA images with non-perfusion areas. Thus, the accuracy of mild NPDR was lower than other grades of DR.

The confusion matrix summarizes the performance of the DR grading algorithm. As *Figure 4* shows, the columns of the matrix represent the true classes, while the rows of the matrix represent the predicted classes. The confusion matrix of DR of different severity for the OCTA-DR data set is shown in *Figure 4*. Based on the confusion matrix and the classification accuracy of DR of different severity, the proposed model achieved good classification where the ground truth was moderate-to-severe NPDR and PDR, but requires improvement where the ground truth was no DR and mild NPDR.

In addition, we plotted the loss and accuracy curves of the proposed model in relation to the epochs for the OCTA-DR data set (*Figures 5,6*, respectively).

Table 4 Ablation study of the CSANet for the OCTA data set

Methods	Accuracy
Baseline	0.9537
Baseline + CHAB	0.9611
Baseline + SPAB	0.9667
Baseline + CBAM	0.963
Baseline + CHAB + SPAB	0.9703
Baseline + HAM	0.9741

CSANet, channel and spatial attention network; OCTA, optical coherence tomography angiography; CHAB, channel attention block; SPAB, spatial attention block; CBAM, convolutional block attention module; HAM, hybrid attention module.

Ablation study

We then performed ablation studies on DR grading to evaluate the effectiveness of each module in our proposed model. We analyzed the effect of the CHAB and SPAB on the OCTA-DR data set with the baseline as the backbone network.

Analysis of the spatial attention module

Table 4 shows the results of grading DR with the SPAB. Notably, the accuracy of SPAB achieved an improvement of 1.30% over the baseline. The advantage of the SPAB is that it captures the relationships in the spatial feature maps to alleviate the problem of small lesions being ignored during convolution.

CHAB

As Table 4 shows, our CHAB achieved an improvement of 0.74% over the baseline. To examine the relationship between the SPAB and CHAB, we simply connected these two attention modules in parallel. The baseline model with the SPAB and CHAB also achieved higher accuracy than the baseline model with the SPAB alone. We also compared our model with the most popular convolutional block attention module (48). The experimental results showed that our model outperformed all the baselines on the OCTA-DR data set.

Figure 7 shows the confusion matrices, which more intuitively represent the classification effect of each attention module. Notably, each attention module showed different degrees of improvement in the classification performance of no DR and mild NPDR, among which the baseline + CHAB + SPAB model showed the most

obvious improvement in the classification performance of mild NPDR. Compared with the SPAB and CHAB modules, simply connecting these two modules in parallel further improved the ability of the models to classify mild NPDR and PDR. However, for the classification of no DR, the performance of the baseline + CHAB + SPAB model was low. Among all the attention mechanisms, the SPAB achieved the best performance for moderate-to-severe NPDR, while the CHAB achieved the best performance for the no-DR class.

Based on the above analysis, we conducted further experiments on the DR data set by separately weighting the feature maps of the SPAB and CHAB. The feature-weighted attention mechanism was merged with the baseline data. The confusion matrix is shown in Figure 4. Compared to the unweighted attention module (baseline + CHAB + SPAB), the weighted spatial attention and the channel attention post-attention mechanism (HAM) led to improvement in the performance classification of no DR and moderate-to-severe NPDR. However, the effect on the performance of the model for the other categories was minimal.

In summary, our extensive experimental results showed that the proposed attention module could be used to perform fine-grained classification.

Comparisons with other state-of-the-art methods

To further evaluate the performance of our method for grading DR, we compared our proposed model with other representative neural networks. All the methods achieved their optimal performance with the corresponding epochs for the training models.

As Table 5 shows, the accuracy rate of our model for grading DR for the OCTA-DR data set was 97.41%, which was 2.78% higher than that of the ordinal regularization network (ORNet) (34) and 6.85% higher than that of the model proposed by Dong *et al.* (30). As Table 6 shows, the kappa value of our model for DR grading was 0.8813 for the DRAC 2022 data set, which was 0.0310 higher than that of the ORNet and 0.0515 higher than that of the model proposed by Dong *et al.* (30). We found that the accuracy of the model on the challenged data set was superior to other mainstream models but did not achieve very satisfactory results. Thus, we analyzed the data set and found that the OCTA images in the data set had different degrees of motion artifacts and mosaic-like patches. Among them, the number of severe image deletions and similar mosaics

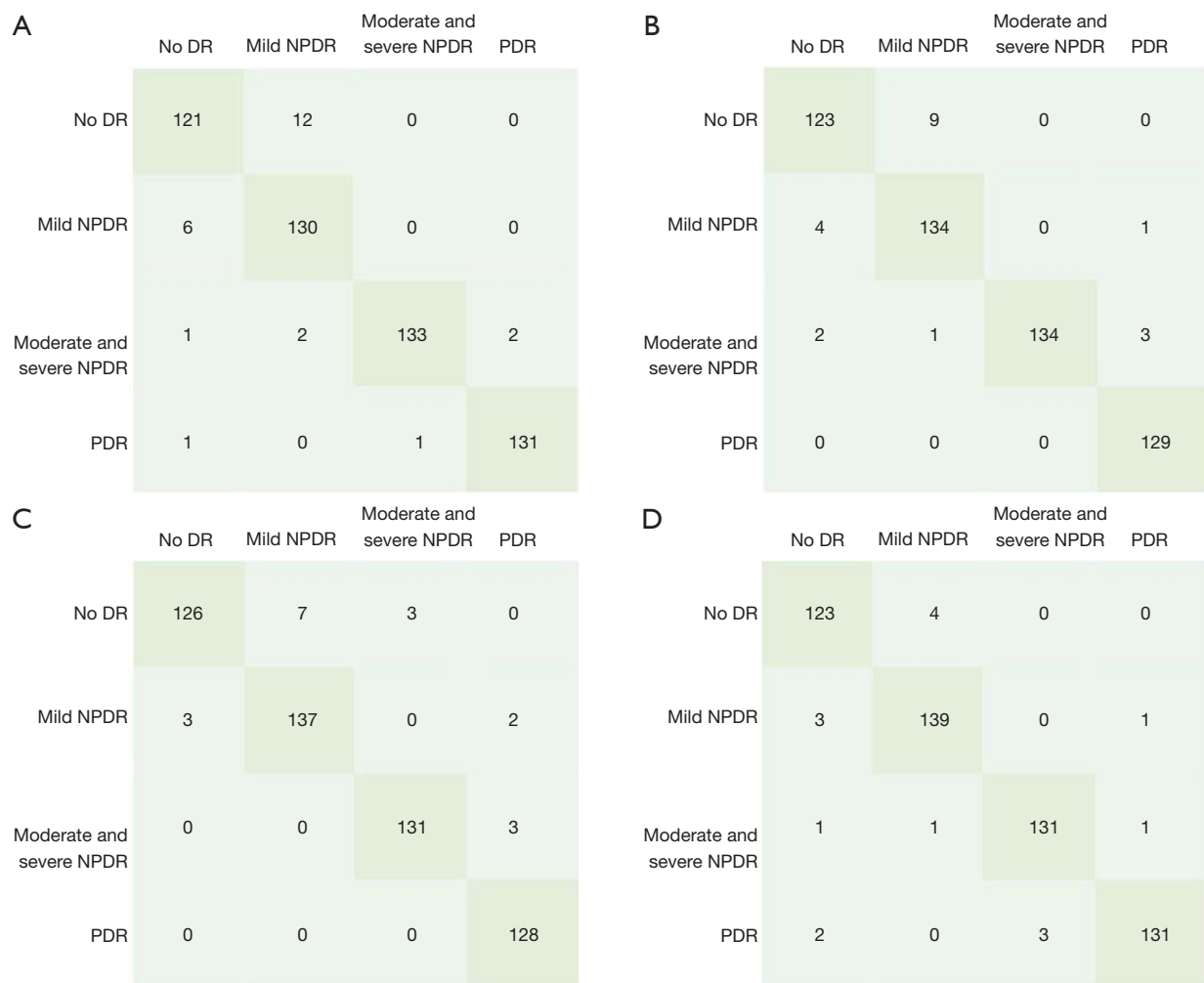


Figure 7 Confusion matrix comparison of DR classification of each attention module. (A) Baseline, (B) baseline + CHAB, (C) baseline + SPAB, (D) baseline + CHAB + SPAB. CHAB, channel attention block; SPAB, spatial attention block; DR, diabetic retinopathy; NPDR, non-proliferative DR; PDR, proliferative DR.

Table 5 The accuracy and loss comparison of the proposed model with other mainstream convolutional neural network models on the OCTA-DR data set

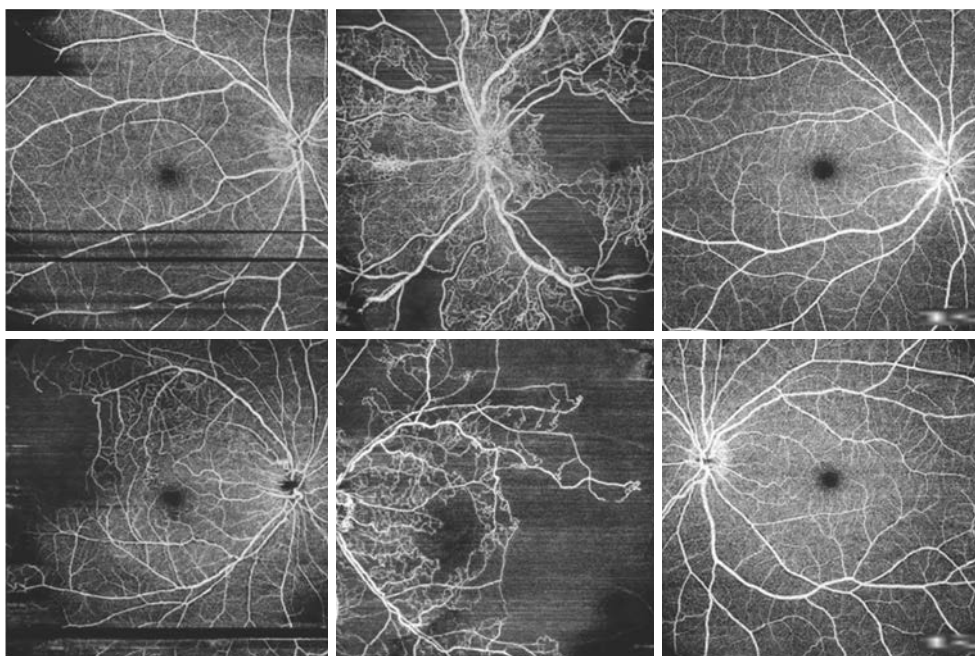
Metrics	Ours	ORNet	Dong (30)	Inception V3	VGG16	GoogLeNet	Resnet50
Accuracy	97.41%	94.63%	90.56%	81.25%	79.98%	79.78%	79.22%
Loss	0.0897	0.1552	0.2679	0.3882	0.4714	0.4901	0.4917
Epoch	60	60	70	60	60	50	55

OCTA-DR, optical coherence tomography angiography-diabetic retinopathy; ORNet, ordinal regularization network; VGG, visual geometry group.

Table 6 Kappa coefficient and accuracy comparison of the proposed model with other mainstream convolutional neural network models on the DRAC 2022 data set

Metrics	Ours	ORNet	Dong (30)	Inception V3	GoogLeNet	VGG16	Resnet50
Kappa	0.8813	0.8503	0.8298	0.8241	0.8082	0.7342	0.7676
AUC	0.9463	0.9394	0.9172	0.9130	0.9067	0.8882	0.8797
Accuracy	0.8571	0.8280	0.8017	0.7959	0.7638	0.7493	0.7289
Epoch	80	70	70	70	50	65	55

DRAC, diabetic retinopathy analysis challenge; ORNet, ordinal regularization network; VGG, visual geometry group; AUC, area under the receiver operating characteristic curve.

**Figure 8** Low-quality images in the DRAC 2022 data set. From left to right: the missing pixels, the motion artifacts, and the mosaic-like patches. DRAC, diabetic retinopathy analysis challenge.**Table 7** Number of parameters for all models

Model	Parameters (M)
VGG16	134.28
Dong (30)	97.76
AlexNet	57.02
Inception V3	27.46
ORNet	25.61
ResNet-50	23.52
Ours	30.13

M, million; VGG, visual geometry group; ORNet, ordinal regularization network.

in the no-DR category accounted for 11.3% of the total number of the original data sets and 21% of the no-DR category. *Figure 8* shows some images with artifacts and of poor quality. We believe that artifacts and image quality are the main factors affecting the performance of the proposed model.

The advantages of our approach are twofold: (I) our proposed model uses lightweight deep neural network parameters. As *Table 7* shows, our proposed model used only 30.13 million parameters, which is one-third of the amount of network parameters used by Dong *et al.* (30). Additionally, it achieves better performance with fewer parameters. Conversely, complex models will have

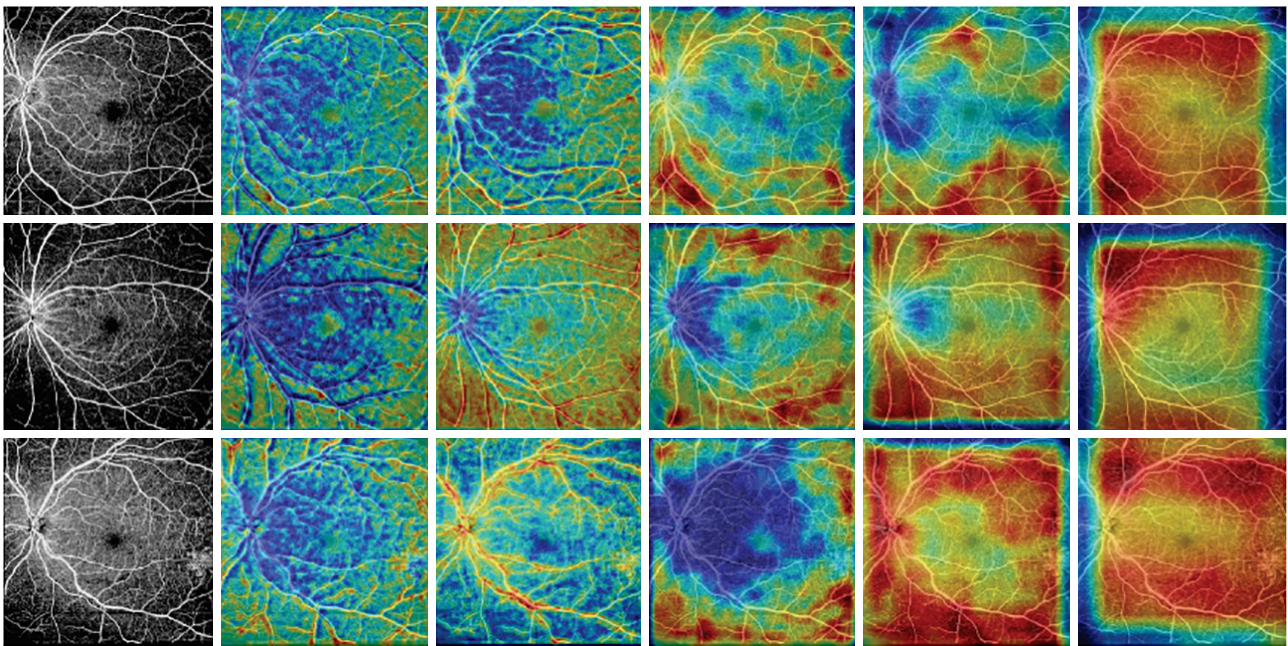


Figure 9 Visualization of different diabetic retinopathy severity levels for each hybrid attention module in the network using gradient-weighted class activation mapping. From left to right, the feature maps show the first to fifth layers. The first row is mild DR samples; the second row is moderate-to-severe DR samples, and the third row is proliferative DR samples. DR, diabetic retinopathy.

overfitting problems for a few-shot DR data set; and (II) we designed attention modules in the framework that can detect small lesions in DR images and improve the DR grading performance to some extent.

Visualizing the classification process of our model using gradient-weighted class activation mapping (Grad-CAM)

In this study, we use Grad-CAM (49), which can visualize the key regions of the feature maps of the models. The representative heat map generated by Grad-CAM is shown in *Figure 9*. The images in *Figure 9* show mild NPDR patterns, moderate-to-severe NPDR patterns, and PDR patterns in each HAM block module in the network. Grad-CAM can show the location of the discriminative features of the model during the training process. This will be useful for grading DR tasks. It should be noted that the proposed model is good at localizing the lesion location.

Discussion

Automated DR screening has become a research hotspot in medical imaging. Deep-learning methods have shown good performance in DR grading tasks; however, there is still

a certain gap in the clinical application of such methods. In this article, we developed the CSANet to solve the overfitting problem caused by small lesions and few-shot DR samples. To improve the interpretability of the model, we also obtained the location maps of suspicious lesions in OCTA images so that the results generated by the model could help ophthalmologists make the correct diagnosis.

The experimental results showed that our proposed model achieved state-of-the-art performance on the OCTA-DR and DRAC data sets. The advantages of our method are twofold: (I) the CHAB can capture a group of channel weights to focus on useful features and suppress irrelevant features; (II) the SPAB can capture a set of spatial weights, which addresses the issue of small lesions being ignored during the convolution process, and can learn richer features from the OCTA images.

Our method achieved state-of-the-art performance on the OCTA-DR data set and the DRAC 2022 data set; however, there is still room for improvement. First, the number of OCTA samples was relatively small, which led to the overfitting of the deeper neural network. Second, the entire network was only trained based on the image-level annotations, which made it very difficult to accurately locate small lesion areas.

Conclusions

In this article, we developed a hybrid attention network (CSANet) that incorporates channel attention and spatial attention. The experimental results for the OCTA-DR and DRAC data sets showed that our network outperformed other related methods in grading DR tasks. In the future, we intend to use adversarial networks to generate richer OCTA images, which will to some extent prevent overfitting when training the model. Additionally, we also intend to try to develop useful rules for deeper neural networks for DR grading tasks.

Acknowledgments

The authors would like to thank the editors and anonymous referees for their constructive criticism and valuable suggestions.

Funding: This work was supported by the Natural Science Foundation of Shandong Province (No. ZR2020MF105), Guangdong Provincial Key Laboratory of Biomedical Optical Imaging Technology (No. 2020B121201010), the Natural National Science Foundation of China (Nos. 62175156 and 61675134), Science and Technology Innovation Project of Shanghai Science and Technology Commission (Nos. 19441905800 and 22S31903000), and Qufu Normal University Foundation for High Level Research (No. 116-607001).

Footnote

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <https://qims.amegroups.com/article/view/10.21037/qims-23-1270/coif>). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with

the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. He A, Li T, Li N, Wang K, Fu H. CABNet: Category Attention Block for Imbalanced Diabetic Retinopathy Grading. *IEEE Trans Med Imaging* 2021;40:143-53.
2. Olvera-Barrios A, Heeren TF, Balaskas K, Chambers R, Bolter L, Egan C, Tufail A, Anderson J. Diagnostic accuracy of diabetic retinopathy grading by an artificial intelligence-enabled algorithm compared with a human standard for wide-field true-colour confocal scanning and standard digital retinal images. *Br J Ophthalmol* 2021;105:265-70.
3. Wang X, Xu M, Zhang J, Jiang L, Li L, He M, Wang N, Liu H, Wang Z. Joint Learning of Multi-Level Tasks for Diabetic Retinopathy Grading on Low-Resolution Fundus Images. *IEEE J Biomed Health Inform* 2022;26:2216-27.
4. Li X, Jiang Y, Zhang J, Li M, Luo H, Yin S. Lesion-attention pyramid network for diabetic retinopathy grading. *Artif Intell Med* 2022;126:102259.
5. Yang Y, Shang F, Wu B, Yang D, Wang L, Xu Y, Zhang W, Zhang T. Robust Collaborative Learning of Patch-Level and Image-Level Annotations for Diabetic Retinopathy Grading From Fundus Image. *IEEE Trans Cybern* 2022;52:11407-17.
6. Cho NH, Shaw JE, Karuranga S, Huang Y, da Rocha Fernandes JD, Ohlrogge AW, Malanda B. IDF Diabetes Atlas: Global estimates of diabetes prevalence for 2017 and projections for 2045. *Diabetes Res Clin Pract* 2018;138:271-81.
7. Ragkousis A, Kozobolis V, Kabanarou S, Bontzos G, Mangouritsas G, Heliopoulos I, Chatziralli I. Vessel Density around Foveal Avascular Zone as a Potential Imaging Biomarker for Detecting Preclinical Diabetic Retinopathy: An Optical Coherence Tomography Angiography Study. *Semin Ophthalmol* 2020;35:316-23.
8. Ghazal M, Ali SS, Mahmoud AH, Shalaby AM, El-Baz A. Accurate Detection of Non-Proliferative Diabetic Retinopathy in Optical Coherence Tomography Images Using Convolutional Neural Networks. *IEEE Access* 2020;8:34387-97.
9. Gerendas BS, Bogunovic H, Sadeghipour A, Schlegl T, Langs G, Waldstein SM, Schmidt-Erfurth U. Computational image analysis for prognosis determination

- in DME. *Vision Res* 2017;139:204-10.
10. Elsharkawy M, Elrazzaz M, Sharafeldeen A, Alhalabi M, Khalifa F, Soliman A, Elnakib A, Mahmoud A, Ghazal M, El-Daydamony E, Atwan A, Sandhu HS, El-Baz A. The Role of Different Retinal Imaging Modalities in Predicting Progression of Diabetic Retinopathy: A Survey. *Sensors (Basel)* 2022.
 11. Zahid S, Dolz-Marco R, Freund KB, Balaratnasingam C, Dansingani K, Gilani F, Mehta N, Young E, Klifto MR, Chae B, Yannuzzi LA, Young JA. Fractal Dimensional Analysis of Optical Coherence Tomography Angiography in Eyes With Diabetic Retinopathy. *Invest Ophthalmol Vis Sci* 2016;57:4940-7.
 12. Gramatikov BI. Modern technologies for retinal scanning and imaging: an introduction for the biomedical engineer. *Biomed Eng Online* 2014;13:52.
 13. Mendis KR, Balaratnasingam C, Yu P, Barry CJ, McAllister IL, Cringle SJ, Yu DY. Correlation of histologic and clinical images to determine the diagnostic value of fluorescein angiography for studying retinal capillary detail. *Invest Ophthalmol Vis Sci* 2010;51:5864-9.
 14. Cheng SC, Huang YM. A novel approach to diagnose diabetes based on the fractal characteristics of retinal images. *IEEE Trans Inf Technol Biomed* 2003;7:163-70.
 15. Akil H, Karst S, Heisler M, Etminan M, Navajas E, Maberley D. Application of optical coherence tomography angiography in diabetic retinopathy: a comprehensive review. *Can J Ophthalmol* 2019;54:519-28.
 16. Sambhav K, Grover S, Chalam KV. The application of optical coherence tomography angiography in retinal diseases. *Surv Ophthalmol* 2017;62:838-66.
 17. Jia Y, Bailey ST, Hwang TS, McClintic SM, Gao SS, Pennesi ME, Flaxel CJ, Lauer AK, Wilson DJ, Hornegger J, Fujimoto JG, Huang D. Quantitative optical coherence tomography angiography of vascular abnormalities in the living human eye. *Proc Natl Acad Sci U S A* 2015;112:E2395-402.
 18. Sandhu HS, Elmogy M, Taher Sharafeldeen A, Elsharkawy M, El-Adawy N, Eltanboly A, Shalaby A, Keynton R, El-Baz A. Automated Diagnosis of Diabetic Retinopathy Using Clinical Biomarkers, Optical Coherence Tomography, and Optical Coherence Tomography Angiography. *Am J Ophthalmol* 2020;216:201-6.
 19. Ramasamy LK, Padinjappurathu SG, Kadry S, Damaševičius R. Detection of diabetic retinopathy using a fusion of textural and ridgelet features of retinal images and sequential minimal optimization classifier. *PeerJ Comput Sci* 2021;7:e456.
 20. Abdelsalam MM, Zahran MA. A Novel Approach of Diabetic Retinopathy Early Detection Based on Multifractal Geometry Analysis for OCTA Macular Images Using Support Vector Machine. *IEEE Access* 2021;9:22844-58.
 21. Maqsood S, Damaševičius R, Shah FM, Maskeliūnas R. Detection of Macula and Recognition of Aged-Related Macular Degeneration in Retinal Fundus Images. *Computing and Informatics* 2021;40:957-87.
 22. Lim WX, Chen Z, Ahmed A. The adoption of deep learning interpretability techniques on diabetic retinopathy analysis: a review. *Med Biol Eng Comput* 2022;60:633-42.
 23. Ahsan MA, Qayyum A, Razi A, Qadir J. An active learning method for diabetic retinopathy classification with uncertainty quantification. *Med Biol Eng Comput* 2022;60:2797-811.
 24. Islam SMS, Hasan MM, Abdullah S. Deep learning based early detection and grading of diabetic retinopathy using retinal fundus images. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 2018. doi: 10.48550/arXiv.1812.10595
 25. 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. *Annu Int Conf IEEE Eng Med Biol Soc* 2018;2018:2724-7.
 26. Krause J, Gulshan V, Rahimy E, Karth P, Widner K, Corrado GS, Peng L, Webster DR. Grader Variability and the Importance of Reference Standards for Evaluating Machine Learning Models for Diabetic Retinopathy. *Ophthalmology* 2018;125:1264-72.
 27. Chaurasia BK, Raj H, Rathour SS, Singh PB. Transfer learning-driven ensemble model for detection of diabetic retinopathy disease. *Med Biol Eng Comput* 2023;61:2033-49.
 28. Zang P, Gao L, Hormel TT, Wang J, You Q, Hwang TS, Jia Y. DcardNet: Diabetic Retinopathy Classification at Multiple Levels Based on Structural and Angiographic Optical Coherence Tomography. *IEEE Trans Biomed Eng* 2021;68:1859-70.
 29. Maqsood S, Damaševičius R, Maskeliūnas R. Hemorrhage Detection Based on 3D CNN Deep Learning Framework and Feature Fusion for Evaluating Retinal Abnormality in Diabetic Patients. *Sensors (Basel)* 2021.
 30. Dong B, Wang X, Qiang X, Du F, Gao L, Wu Q, Cao G, Dai C. A multi-branch convolutional neural network for screening and staging of diabetic retinopathy based on wide-field optical coherence tomography angiograph. *IRBM* 2022;43:614-20.

31. Ouyang J, Mao D, Guo Z, Liu S, Xu D, Wang W. Contrastive self-supervised learning for diabetic retinopathy early detection. *Med Biol Eng Comput* 2023;61:2441-52.
32. Ryu G, Lee K, Park D, Park SH, Sagong M. A deep learning model for identifying diabetic retinopathy using optical coherence tomography angiography. *Sci Rep* 2021;11:23024.
33. Durai DBJ, Jaya T. Automatic severity grade classification of diabetic retinopathy using deformable ladder Bi attention U-net and deep adaptive CNN. *Med Biol Eng Comput* 2023;61:2091-113.
34. Tang W, Yang Z, Song Y. Disease-grading networks with ordinal regularization for medical imaging. *Neurocomputing* 2023;545:126245.
35. Zhang X, Wang D, Zhou Z, Ma Y. Robust Low-Rank Tensor Recovery with Rectification and Alignment. *IEEE Trans Pattern Anal Mach Intell* 2021;43:238-55.
36. Charilaou P, Battat R. Machine learning models and over-fitting considerations. *World J Gastroenterol* 2022;28:605-7.
37. Alexander Max B, Hostetler Z, Vavalle N, Armiger R, Coates R, Gayzik F. Hierarchical Validation Prevents Over-Fitting of the Neck Material Model for an Anthropomorphic Test Device Used in Underbody Blast Scenarios. *J Biomech Eng* 2021;143:014505.
38. Ding C, Li Y, Wen Y, Zheng M, Zhang L, Wei W, Zhang Y. Boosting Few-Shot Hyperspectral Image Classification Using Pseudo-Label Learning. *Remote Sens* 2021;13:3539.
39. Piao C, Lv M, Wang S, Zhou R, Wang Y, Wei J, Liu J. Multi-objective data enhancement for deep learning-based ultrasound analysis. *BMC Bioinformatics* 2022;23:438.
40. Qian L, Hu L, Zhao L, Wang T, Jiang R. Sequence-dropout block for reducing overfitting problem in image classification. *IEEE Access* 2020;8:62830-40.
41. Hu J, Shen L, Sun G. Squeeze-and-Excitation Networks. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA; 2018:7132-41.
42. Wang Q, Wu B, Zhu P, Li P, Zuo W, Hu Q. ECA-Net: efficient channel attention for deep convolutional neural networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 2020;11534-42.
43. Qian B, Chen H, Wang X, Che H, Kwon G, Kim J, et al. DRAC: Diabetic Retinopathy Analysis Challenge with Ultra-Wide Optical Coherence Tomography Angiography Images. 2023. doi: 10.48550/arXiv.2304.02389.
44. Zhang YS, Mucollari I, Kwan CC, Dingillo G, Amar J, Schwartz GW, Fawzi AA. Reversed Neurovascular Coupling on Optical Coherence Tomography Angiography Is the Earliest Detectable Abnormality before Clinical Diabetic Retinopathy. *J Clin Med* 2020.
45. Simonett JM, Scarinci F, Picconi F, Giorno P, De Geronimo D, Di Renzo A, Varano M, Frontoni S, Parravano M. Early microvascular retinal changes in optical coherence tomography angiography in patients with type 1 diabetes mellitus. *Acta Ophthalmol* 2017;95:e751-5.
46. Khalid H, Schwartz R, Nicholson L, Huemer J, El-Bradey MH, Sim DA, Patel PJ, Balaskas K, Hamilton RD, Keane PA, Rajendram R. Widefield optical coherence tomography angiography for early detection and objective evaluation of proliferative diabetic retinopathy. *Br J Ophthalmol* 2021;105:118-23.
47. Welikala RA, Fraz MM, Dehmeshki J, Hoppe A, Tah V, Mann S, Williamson TH, Barman SA. Genetic algorithm based feature selection combined with dual classification for the automated detection of proliferative diabetic retinopathy. *Comput Med Imaging Graph* 2015;43:64-77.
48. Woo S, Park J, Lee JY, Kweon IS. CBAM: Convolutional block attention module. *Proceedings of the European Conference on Computer Vision (ECCV)* 2018;7:3-19.
49. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: visual explanations from deep networks via gradient-based localization. 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy; 2017:618-26.

Cite this article as: Ma F, Liu X, Wang S, Li S, Dai C, Meng J. CSANet: a lightweight channel and spatial attention neural network for grading diabetic retinopathy with optical coherence tomography angiography. *Quant Imaging Med Surg* 2024;14(2):1820-1834. doi: 10.21037/qims-23-1270