



# SCINet: A Segmentation and Classification Interaction CNN Method for Arteriosclerotic Retinopathy Grading

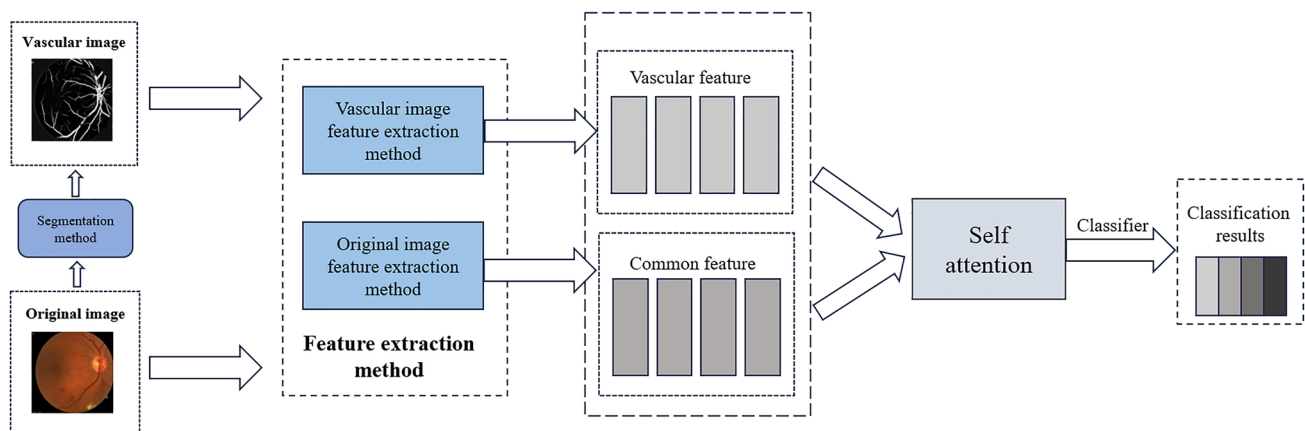
Xiongwen Quan<sup>1</sup> · Xingyuan Ou<sup>1</sup> · Li Gao<sup>2</sup> · Wenya Yin<sup>1</sup> · Guangyao Hou<sup>1</sup> · Han Zhang<sup>1</sup>

Received: 21 February 2024 / Revised: 9 August 2024 / Accepted: 12 August 2024  
© International Association of Scientists in the Interdisciplinary Areas 2024

## Abstract

As a common disease, cardiovascular and cerebrovascular diseases pose a great harm threat to human wellness. Even using advanced and comprehensive treatment methods, there is still a high mortality rate. Arteriosclerosis, as an important factor reflecting the severity of cardiovascular and cerebrovascular diseases, is imperative to detect the arteriosclerotic retinopathy. However, the detection of arteriosclerosis retinopathy requires expensive and time-consuming manual evaluation, while end-to-end deep learning detection methods also need interpretable design to high light task-related features. Considering the importance of automatic arteriosclerotic retinopathy grading, we propose a segmentation and classification interaction network (SCINet). We propose a segmentation and classification interaction architecture for grading arteriosclerotic retinopathy. After IterNet is used to segment retinal vessel from original fundus images, the backbone feature extractor roughly extracts features from the segmented and original fundus arteriosclerosis images and further enhances them through the vessel aware module. The last classifier module generates fundus arteriosclerosis grading results. Specifically, the vessel aware module is designed to highlight the important areal vessel features segmented from original images by attention mechanism, thereby achieving information interaction. The attention mechanism selectively learns the vessel features of segmentation region information under the proposed interactive architecture, which leads to reweighting the extracted features and enhances significant feature information. Extensive experiments have confirmed the effect of our model. SCINet has the best performance on the task of arteriosclerotic retinopathy grading. Additionally, the CNN method is scalable to similar tasks by incorporating segmented images as auxiliary information.

## Graphical Abstract



**Keywords** Arteriosclerotic retinopathy grading · Feature enhancement · Convolutional neural network · Interaction architecture

Extended author information available on the last page of the article

Published online: 02 September 2024

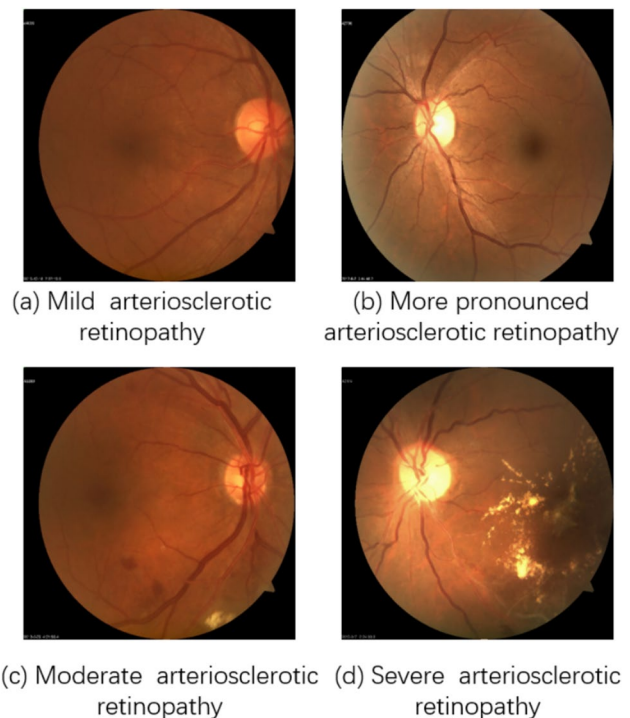
## 1 Introduction

Changes in the retinal vessels often reflect variation in some parts of human organs. Therefore, retinal arteriosclerosis has medical significance to show the changes of cerebral and renal vascular system. According to these studies [1–3], a large number of patients worldwide die from cardiovascular and cerebrovascular disease every year. Specifically, hypertension, diabetes, obesity and smoking, are the main causes of above diseases [4]. For experienced doctors, they can utilize the arteriosclerosis in blood vessels as an indicator of peripheral arteriosclerosis and cerebral artery characteristics to some extent.

The primary symptoms of arteriosclerosis include degenerative alterations in the middle arterial wall, thickening on the artery wall, constriction for blood vessels, etc. [5]. In the event that the blood supply to the central artery of the retina is impaired, patients' vision will be affected accordingly. Unless removing the vascular blockage immediately, patients may become blind.

In clinical practice, doctors usually identify changes in the fundus from small blood vessels [6]. There will be multiple bleeding, swelling of the optic disc and leakage of protein fluid in severe cases. Silver wire and copper wire, a bright and clear light band in the middle of eyes, are usually used to describe atherosclerosis of the retina. As for specific performance, it represents mild arteriosclerosis when blood vessels are pale and arterial wall is unobvious. The occurrence of silver wire represents a more severe form of arteriosclerosis, which is the result of arterial wall changes caused by chronic hypertension. Hypertension, hyperlipidemia and other systemic diseases are closely associated with the retinal vascular arteriosclerosis. The severity, onset types, and duration of hypertension can lead to the change of vessels, including vasoconstriction and arteriosclerosis. In addition, elevated blood pressure can lead to spiral-shaped small blood vessels in macular fibrosis and thinning of small arteries on the retina. In order to better quantify arteriosclerotic retinopathy, Scheie's classification identifies arteriosclerotic retinopathy as four stages [7], and examples of fundus arteriosclerosis images with different degrees are listed in Fig. 1.

As a nondestructive and low-cost method for ocular diseases screening, fundus retinal photography has become the primary way for ophthalmological experts to detect diseases and has been widely adopted in various research. However, grading these ophthalmic images for arteriosclerotic retinopathy can be extremely challenging for some doctors, requiring the expertise of experienced ophthalmologists to diagnose, which is inconvenient for underdeveloped and rural areas. Therefore, a fast, economical and



**Fig. 1** Four grades of retinal fundus images with arteriosclerotic retinopathy

convenient way for severity arteriosclerotic retinopathy assessment is in urgent need.

Over the past years, methods based on CNN have achieved gratifying performance in screening many common fundus diseases and play an irreplaceable role in the area of medical imaging [8–10]. Zhang et al. introduced a six-level cataract classification technique that focuses on multi-feature fusion [11]. Hong et al. proposed a CNN with 14 layers to recognize age-related macular degeneration disease and aid doctors in fundus screening [12]. Bai et al. first proposed a CNN model for the grading of arteriosclerotic retinopathy [13], and compared their model with several other recently network models [14–19]. The transfer learning with pretrained models trained on ImageNet datasets has been proven to be an outstanding method for ophthalmic disease classification tasks [20]. Bravo et al. introduced an effective transfer learning method using the VGG architecture for the detection of diabetic retinopathy [21]. However, there are almost no CNN methods have been applied to the grading task of fundus arteriosclerotic retinopathy.

Moreover, as an effective method in deep learning, attention mechanism has been combined with CNN frequently and achieved outstanding progress [22–24]. Wang et al. proposed a CNN model for diagnosing diabetic retinopathy with image-level labels, which can effectively capture salient region information from small high-resolution patches by attention mechanism [25]. He et al. proposed two attention

blocks for addressing unbalanced diabetic retinopathy grading, which can explore more discriminative region features and detailed global feature, respectively [26]. Xie et al. proposed a cross-attention module to classify various ocular diseases [27]. Their module incorporates spatial and channel attention maps from a two-branch ResNet under cross-fusion mode, effectively enhancing the representation capability of disease-specific features.

In addition, the combination of segmentation and classification networks has made outstanding progress in many medical image tasks [28, 29]. The classification network usually do not pay attention to the whole image information. At the same time, the segmented image can help classification network pay attention to the features of some key areas in the whole image during training process, thus improving the classification results. To this end, Xie et al. proposed a mutual bootstrapping CNN model for simultaneous skin damage classification and segmentation [30]. They used the coarse segmented images as masks to provide a prior bootstrapping for classification network, which can effectively enhance the ability of localization and classification for skin lesions. Wu et al. developed an innovative system that combines segmentation and classification for real-time and explainable diagnosis of COVID-19 using chest coherence tomography [31]. By training the classification and segmentation model together, this method also adopts data enhancement technology and image mixing technology to eliminate data bias, and ultimately improves the performance of classification and segmentation. Bai et al. proposed an ensemble learning model to combine segmented retinal vessel image and original image for fundus arteriosclerotic retinopathy grading [13]. They preprocessed the original image by segmented retinal vessel image at pixel level and then sent both the processed and the segmented retinal vessel image into the multi-branch model for training. However, both of them still remain in image-level preprocessing or concatenation, rather than the feature level interaction enhancement.

Therefore, in this paper, we propose a segmentation and classification interaction network (SCINet) based on attention mechanism as a solution to address the challenge of grading fundus arteriosclerotic retinopathy. The input to SCINet is fundus arteriosclerosis photographs and the output is the grading result for different fundus arteriosclerosis images. SCINet is composed of four modules as follows. Firstly, IterNet is used to segment retinal vessel from original fundus images. Then Backbone Feature Extractor is used for extraction features from the segmented and original fundus arteriosclerosis images. Subsequently, we design a vessel aware module to enhance the extracted features. Lastly, the grading results of fundus arteriosclerosis are generated from the classifier module. As sufficient experiments shown, SCINet achieves brilliant performance on the fundus arteriosclerosis grading task. We also explored different backbone

networks, including AlexNet, ResNet, VGG, DenseNet based on the CNN architecture and Vision Transformers based on the Transformer architecture to validate the feasibility of SCINet. We also investigated the impact of network depth on performance.

The major contributions of our research are highlighted as below: (1) we design a segmentation and classification interactive architecture for the fundus arteriosclerosis grading task. (2) We propose an innovative vessel aware module, which selectively learns the features of segmentation region information by the attention mechanism, fostering enhanced feature extraction through an interactive architecture that facilitates collaboration between segmentation and classification. (3) In comparison to the state-of-the-art methods, SCINet achieves the best overall performance.

The rest sections of this paper are structured as below: in Sect. 2, a review on relevant research is introduced. Next, we elaborate on our proposed segmentation and classification interaction CNN method. In Sect. 4, we elaborate various experimental designs and analyse the corresponding results. In the end, we conclude our proposed work.

## 2 Material and Methods

### 2.1 Network Architecture

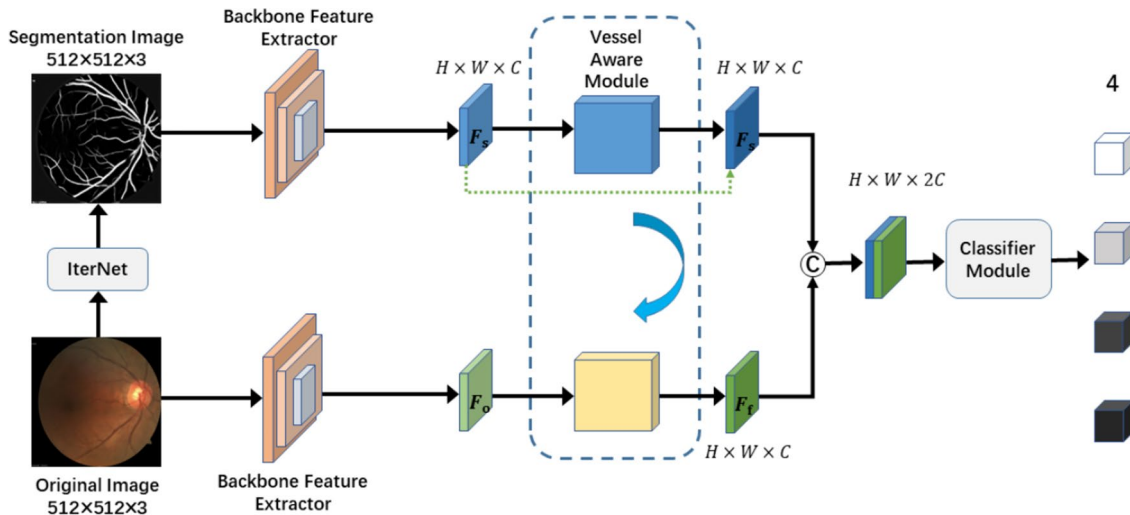
The overall framework of the SCINet is illustrated in Fig. 2. The network comprises four parts: InterNet, backbone feature extractor, vessel aware module, and classifier. In this section, we will give a detail introduction to each part of the SCINet.

#### 2.1.1 IterNet

The IterNet is used to segment the retinal vessel from the original fundus images. IterNet is a novel network model proposed by Li et al. [32] for the segmentation of retinal blood vessels, achieving the state-of-the-art result in dataset DRIVE [33] and CHASEDB1 [34]. This model utilizes a cascade structure consisting of multiple mini U-Nets and the original U-Net architecture. This design enables the detection of obscured details of blood vessels and further enhances the segmentation results. In order to achieve better segmentation images, we adopt the IterNet as our model for retinal vessel segmentation.

#### 2.1.2 Backbone Feature Extractor

The backbone feature extractor captures features from the input segmented retinal vessel images and original fundus photographs. For the segmented fundus images and the original fundus images  $I_s$  and  $I_o$ , the outputs of this module are



**Fig. 2** Overall architecture of SCINet. “C” is the concatenation operation

**Table 1** Fundus arteriosclerotic retinopathy grading performance of different backbone feature extractors

Model	Kappa coefficient	$F_1$	$A_{uc}$	$A_{cc}$
AlexNet	0.179	0.412	0.668	0.394
Vgg19	0.172	0.429	0.706	0.409
ResNet18	0.337	0.493	0.736	0.521
DenseNet121	0.309	0.501	0.729	0.507
VIT-L/16	0.382	0.585	0.792	0.591

$F_s$  and  $F_{s'}$ , respectively. Table 1 shows the classification performance of AlexNet, Vgg, ResNet, DenseNet, and Vision Transformer (VIT). Since ResNet and VIT get better result in overall, we adopt them as our backbone feature extractor. In addition, to investigate the impact of network depths, the comparative experiments were conducted on ResNet18, ResNet34, ResNet50, VIT-B/16, VIT-L/16.

### 2.1.3 Vessel Aware Module

The vessel aware module is designed to highlight the features of critical regions segmented from the original fundus images, thereby improving the model ability to discern subtle features in critical regions. Under the interactive architecture, we introduce segmented vessel images and utilize an attention mechanism to achieve the vessel information in classification tasks based on the segmentation images, which can model semantic interdependence in spatial, adaptively assign weights to features, and enhance vascular features information.

The vessel aware module takes two features from the backbone feature extractor as inputs and generates enhanced

features by improving the characteristic information of key areas within the original global information. The detail of this module is shown in Fig. 3. Inspired by non-local neural networks [35] and cross parallax attention network [36], the module uses two input feature maps,  $F_s$  and  $F_o$ , to enhance the important vessel features. The result of applying a  $1 \times 1$  convolution to  $F_s$  serves as  $Q$ , while the results of applying a  $1 \times 1$  convolution to  $F_o$  serves as  $K$  and  $V$ , respectively. The specific formula is as follows:

$$Q = \text{Conv}(F_s; \theta_s) \quad (1)$$

$$K = \text{Conv}(F_o; \theta_o) \quad (2)$$

$$V = \text{Conv}(F_o; \theta_o) \quad (3)$$

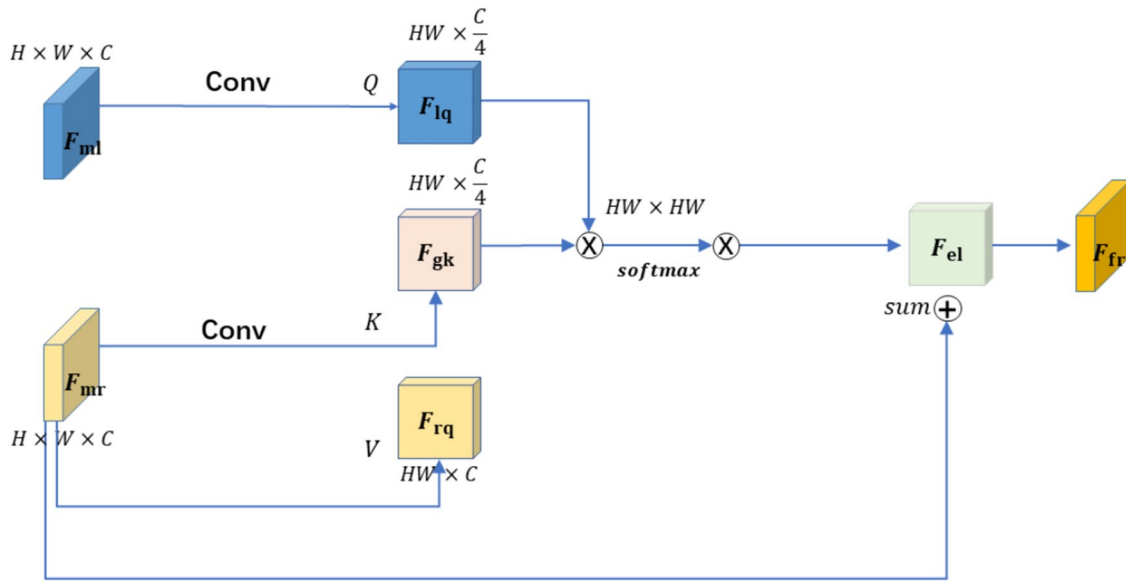
where  $Q$  is the segmented image feature,  $K$  and  $V$  are the original image feature, respectively.  $Q \in \mathbf{R}^{HW \times C'}$ ,  $K \in \mathbf{R}^{HW \times C'}$  and  $V \in \mathbf{R}^{HW \times C}$ ,  $C' = C/4$ . Conv represents a  $1 \times 1$  convolution, and  $\theta$  denotes the corresponding parameters.

To get the correlation weight between the feature maps of segmented and original images at the pixel level, as shown in Fig. 3, by calculating the inner product between query and key, we determine the weight assigned to the value and subsequently compute this matrix of  $F_e$  as

$$F_e = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (4)$$

where  $d_k$  is the dimension of key.

The final step of the vessel aware module is summing the enhanced feature  $F_o$  and the classification feature  $F_e$  to get final feature  $F_{t'}$ .



**Fig. 3** The detailed structure of proposed vessel aware module. “ $\times$ ” denotes matrix multiplication, and “ $+$ ” denotes element-wise sum

$$F_f = \text{Sum}(F_o, F_e) \quad (5)$$

where  $F_f \in \mathbf{R}^{H \times W \times C}$  is the output of this vessel aware module.

### 2.1.4 Classifier

Firstly, we concatenate the segmented images from the backbone feature extractor and the output features from the vessel aware module, which serves as the input of classifier. Compared with other frequently used splicing operation, the operations of summation and product, concatenation provide more information and retain the distinct features of each module. The classifier comprises two fully-connected layers. The initial layer employs an ReLU function, while the other does not. This initial fully-connected layer is designed for dimensionality reduction of the concatenated features, and following layer is to generate the final result. The size of exacted feature is determined by the backbone feature extractor used. Take the ResNet18 as example, the concatenated feature's dimension is 1024, which is then reduced to 512 dimensions by the initial fully-connected layer. Subsequently, the following fully-connected layer continuously reduces to four-dimensional features, which is equal to the four arteriosclerotic retinopathy grading. Finally, the features are compared with grading labels to calculate the network loss of the model.

## 2.2 Transfer Learning

In this task, the ResNet that we use has been trained on the normal ImageNet dataset. Moreover, we employ an ResNet with transfer learning to perform grading on retinal images and compare its performance with the model that does not utilize transfer learning. The network that adopts transfer learning achieves superior grading performance.

## 2.3 Loss Function

As the most popular loss function, cross-entropy can solve majority classification tasks, but this approach is not applicable when dealing with categories imbalanced datasets. Since our arteriosclerosis retinopathy grading data set is unbalanced, it is necessary to impose a higher penalty when misclassifying the smaller classes, which can incentivize the model to pay more attention to correctly classifying those minority classes and avoid the biases towards the majority classes. We finally use the weighted cross-entropy loss function, whose formula is as below:

$$\mathcal{L} = -w_{\text{label}} \log \frac{e^{x_{\text{label}}}}{\sum_{j=0}^{n-1} e^{x_j}} \quad (6)$$

where ‘label’ is the true category of the sample,  $w_{\text{label}}$  is the weight of the ‘label’ class.  $w_{\text{label}} = |N|/(|P| + |N|)$ ,  $|N|$  and  $|P|$  are the number of negative cases and the number of



positive cases with the category label. The  $j$  represents the category,  $x_j$  denotes the confidence probability score of class  $j$  that belongs to the output of the model.

## 2.4 Dataset

In all experiments in this paper, the dataset that we use for the grading of arteriosclerotic retinopathy was collected by Tianjin Huanhu Hospital from 2011 to 2019, and the grading labels were marked by ophthalmologist. In all cases, patients' information such as name, birth date, and therapy time is provided. Additionally, the high resolution of all fundus photographs in this dataset ensures that grading errors are not caused by the dataset itself. There are 706 fundus images in total, as shown in Table 2, the numbers of the four grades are 281, 186, 116, and 126, respectively. During the experiment, we randomly divided these data into training set, validation set, and testing set with a ratio of 8:1:1. We conducted five-fold cross-validation and calculated the average results.

## 2.5 Implementation Detail

During the experiment, we adjust the original images to a smaller resolution of  $512 \times 512$  to reduce the parameters of this model. Subsequently, we adopt some data augmentation techniques such as random cropping, random horizontal flipping, and vertical flipping to alleviate network overfitting. During test stage, we use center cropping instead of random cropping to ensure that the cropped photographs are centered, resulting in improved grading results.

The available framework Pytorch [37] was used to realize all our CNN model. Our experiments run on GeForce RTX 3080 Laptop GPU and adopt Stochastic gradient descent (SGD) optimizer for the networks' training. The initial learning rate  $\alpha_0$  is set at 0.002 and decays based on the poly decays policy  $\alpha = \alpha_0 \times (1 - \frac{t}{T})^m$  [38]. According to the experimental result  $m$  is set as 0.9. The total number of iterations ( $T$ ) in experiments is 30 epochs.

**Table 2** Summary of training, validation, and testing datasets

Class	Train	Valid	Test	Total
1	225	28	28	281
2	147	18	18	183
3	92	12	12	116
4	100	13	13	126
Total	571	64	71	706

## 3 Experiments

This section firstly introduces the evaluation metrics. Then two sets of experiments are conducted, one is ablation experiments, another one is to explore different depth CNNs' influence. Lastly, we compared and analyzed the experimental results.

### 3.1 Evaluation Metrics

To assess the grading performance of the SCINet, we adopt four evaluation metrics: the Kappa coefficient ( $K$  in Eq. (7)), the area under curve ( $A_{uc}$  in Eq. (8)), F1-score ( $F_1$  in Eq. (9)), and the accuracy of classification ( $A_{cc}$ ). The Kappa coefficient is universally employed to evaluate consistency, with a range from  $-1$  to  $1$ . The F1-score illustrates the harmonic score average about both of recall and precision. By these evaluation metrics, we can comprehensively assess the grading performance of our proposed network model. We use official sklearn package to calculate all of the above evaluation metrics.

The Kappa coefficient is defined as below:

$$K = \frac{p_o - p_e}{1 - p_e}$$

$$p_o = \frac{\sum_{i=1}^r X_{ii}}{N} \quad (7)$$

$$p_e = \frac{\sum_{i=1}^r X_{i+} \times X_{+i}}{N^2},$$

where  $p_o$  is the accuracy of prediction,  $p_e$  is the accidental consistency error,  $r = 4$  is the number of rows in the confusion matrix,  $N$  is the total number of fundus image samples,  $X_{ii}$  is the value located at the intersection of column  $i$  and row  $i$  in the confusion matrix,  $X_{i+}$  is the sum of all elements in the  $i$ -th row, while  $X_{+i}$  is the sum of all elements in the  $i$ -th column.

The formulas for AUC and F1-score are as follows:

$$A_{uc} = \int_{x=0}^1 r_{TP}(r_{FP}^{-1}(x))dx \quad (8)$$

$$r_{TP} = \frac{n_{TP}}{n_{TP} + n_{FP}}, r_{FP} = \frac{n_{FP}}{n_{FP} + n_{FN}}$$

$$F_1 = \frac{2n_{TP}}{2n_{TP} + n_{FN} + n_{FP}} \quad (9)$$

where  $n_{TP}$ ,  $n_{TN}$ ,  $n_{FP}$ ,  $n_{FN}$  are the number of true positive, true negative, false positive, false negative, respectively.  $r_{TP}$  is the true positive rate, and  $r_{FP}$  is the false positive rate.

## 3.2 Ablation Experiments

### 3.2.1 Grading Results Using ResNet18 after Transfer Learning

Firstly we compared the performance of original ResNet18 and ResNet18 with transfer learning for grading arteriosclerotic retinopathy. Then, we also compared the manner of inputting the concatenated segmented retinal vessel photographs and original images to a single CNN network. The result is listed in Table 5. We can see that ResNet18 achieved better results after ImageNet pre-training. The Kappa coefficient,  $F_1$ ,  $A_{uc}$  and  $A_{cc}$  scores improved by 0.110, 0.082, 0.091, and 0.100 respectively after using transfer learning. At the same time, inputting the concatenated images also realize better grading performance. The Kappa coefficient,  $F_1$ ,  $A_{uc}$  and  $A_{cc}$  scores improving by 0.015, 0.026, 0.036, and 0.015 respectively after using transfer learning.

### 3.2.2 The Effect of Vessel Aware Module on the Grading

After highlighting the features of segmented important areas from the original images, the vessel aware module highlights the features of the segmented area's relevant information. As described in Table 5, compared with the performance generated by complete SCINet, Kappa coefficient,  $F_1$ ,  $A_{uc}$  and  $A_{cc}$  declined 0.140, 0.099, 0.067 and 0.098 respectively, when we take out the vessel aware module.

## 3.3 Grading Performance and Computational Parameters

Table 3 lists the detail of the computational parameters for ResNet and VIT models with different depths. We can observe that the introduction of SCI (segmentation and classification interaction) will raise the network parameters and the calculation amount reaches the highest complexity when

**Table 3** Calculation parameters of network with/without SCI (segmentation and classification interaction)

Backbone	SCI	FLOPs ( $\times 10^9$ )	Params ( $\times 10^6$ )
ResNet18	Without	14.6	11.7
ResNet18	With	28.2	55.6
ResNet34	Without	29.4	21.8
ResNet34	With	57.3	70.2
ResNet50	Without	33.0	25.8
ResNet50	With	63.7	77.9
VIT-B/16	Without	350.9	85.8
VIT-B/16	With	653.7	164.7
VIT-L/16	Without	1271.8	303.4
VIT-L/16	With	2543.6	608.8

VIT-L/16 is used as backbone. Table 4 shows the detailed results of different depths' ResNet and VIT backbone. Through these results, we can observe that using deeper CNN feature extractor may not necessarily result in better performance. ResNet18 has achieved the best comprehensive result. Many researches have shown this similar situation that deep learning models cannot improve their performance through the linear increase of network depth [39]. The phenomenon may be caused by three aspects. The first is the issue of gradient disappearance. The difficulty of optimization will increase with the increase of network depth [39]. The second is related to the reduction in feature reuse, which results in incomplete utilization of many features obtained by CNNs [40]. Thirdly, due to the limitation of training samples, the network model may not be adequately trained.

On the other hand, for all VIT backbone networks, we found that the use of SCI has not significant impact on the results, and its overall performance was not as good as that of ResNet. Firstly, this is because SCI is mainly used to capture long-distance dependencies between segmented and original images, which can precisely compensate for the shortcomings of CNN in this regard. Therefore, the application of SCI in CNN leads to noticeable performance improvements. However, due to the strong ability of VIT itself to model long-distance dependencies, using SCI in VIT does not lead to significant improvements.

Secondly, VIT requires a large amount of data to train the model, while our dataset has a small amount of data with an imbalanced distribution. The use of VIT can cause overfitting problems, so the effectiveness of VIT is not outstanding.

## 3.4 Comparison with State-of-the-art Methods

In comparison to the state-of-the-art results of the grading on this fundus arteriosclerosis dataset, SCINet has obtained the best performance on the arteriosclerotic retinopathy grading task. As shown in Table 6 and Fig. 4,

**Table 4** Classification results of different backbone feature extractor with/without SCI (segmentation and classification interaction)

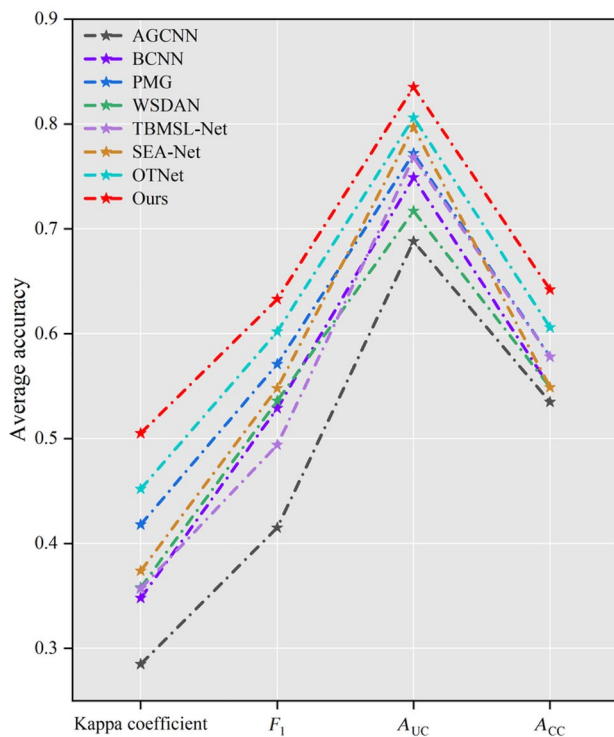
Backbone	SCI	Kappa coefficient	$F_1$	$A_{uc}$	$A_{cc}$
ResNet18	Without	0.337	0.493	0.736	0.521
ResNet18	With	0.505	0.633	0.835	0.642
ResNet34	Without	0.277	0.491	0.762	0.493
ResNet34	With	0.428	0.623	0.846	0.606
ResNet50	Without	0.268	0.482	0.750	0.485
ResNet50	With	0.420	0.618	0.839	0.592
VIT-B/16	Without	0.372	0.568	0.793	0.589
VIT-B/16	With	0.383	0.583	0.785	0.594
VIT-L/16	Without	0.382	0.585	0.792	0.591
VIT-L/16	With	0.396	0.602	0.811	0.598

**Table 5** Ablation experiments performed on fundus arteriosclerosis grading

Performance metric	ResNet18	ResNet18 + transfer learning	ResNet18 + transfer learning + concat segmented and original images	Our method without vessel aware module	Our method
Kappa coefficient	0.227	0.337	0.352	0.365	0.505
$F_1$	0.411	0.493	0.519	0.534	0.633
$A_{uc}$	0.645	0.736	0.762	0.768	0.835
$A_{cc}$	0.421	0.521	0.536	0.544	0.642

**Table 6** Comparisons with state-of-the-art methods

Model	Backbone	Kappa coefficient	$F_1$	$A_{uc}$	$A_{cc}$
AGCNN [15]	DenseNet121	0.285	0.415	0.688	0.535
BCNN [14]	ResNet18	0.348	0.529	0.749	0.549
PMG [17]	ResNet50	0.418	0.571	0.772	0.578
WSDAN [16]	ResNet18	0.358	0.536	0.717	0.549
TBMSL-Net [19]	ResNet50	0.357	0.494	0.768	0.578
SEA-Net [18]	ResNet50	0.374	0.548	0.797	0.549
OTNet [13]	ResNet18	0.452	0.602	0.806	0.606
Ours	ResNet18	0.505	0.633	0.835	0.642

**Fig. 4** The detailed structure of proposed vessel aware module

SCINet improves the Kappa coefficient,  $F_1$ ,  $A_{uc}$  and  $A_{cc}$  by 0.053, 0.031, 0.029 and 0.036, respectively. In comparison

to other methods, we use the segmented image feature to interactively enhance the original image features, which highlighting the features of segmented important areas from the original images, thereby achieving the information interaction, and improving the intelligence of model training.

## 4 Conclusion

In this paper, we develop a novel segmentation and classification interaction network called SCINet, which is a CNN architecture with feature level interaction enhancement between segmentation and classification for arteriosclerosis retinopathy grading. Specifically, the proposed vessel aware module can effectively highlight the features of segmented important areas from the original images, thereby achieving the information interaction, and improving the intelligence of model. To validate the effectiveness of the vessel aware module, we also conduct some ablation experiments. For backbone feature extractor, the grading performance is limited by increasing the network depth. On complex arteriosclerosis retinopathy grading task, SCINet achieves the best performance and can serve as a valuable tool for aiding clinical diagnosis.

However, there are several limitations in the network we proposed. Firstly, the overall structure of SCINet is not an end-to-end way, and we have to segment the original images in advance before subsequent processing. What's more, there exists a significant imbalance problem in the distribution of the four grading cases. Implementing over-sampling techniques for cases with small samples could be a viable approach to mitigate this issue.

The designed module exhibits excellent scalability and can be applied to more general situations. It can be used for many analogous assignments, such as thorax disease and skin diagnosis, which require the segmented images for auxiliary effect. Additionally, the interaction architecture can also be adopted to those multi-modal image crucial task, where different modal images' correlation is important for the research.



**Funding** Funding was supported by National Natural Science Foundation of China (Nos. 61973174, 62373200), Key project of the Natural Science Foundation of Tianjin City (No. 21JCZDJC00140).

**Data availability** The data that support the findings of this study are available on request from the corresponding author, Han Zhang, upon reasonable request.


## References

- Wong TY, Shankar A, Klein R et al (2005) Retinal arteriolar narrowing, hypertension, and subsequent risk of diabetes mellitus. *Arch Inter Med* 165(9):1060–1065. <https://doi.org/10.1001/archinte.165.9.1060>
- Stitt AW, Curtis TM, Chen M et al (2016) The progress in understanding and treatment of diabetic retinopathy. *Prog Retina Eye Res* 51:156–186. <https://doi.org/10.1016/j.preteyeres.2015.08.001>
- Liew G, Wang JJ, Cheung N et al (2008) The retinal vasculature as a fractal: methodology, reliability, and relationship to blood pressure. *Ophthalmology* 115(11):1951–1956. <https://doi.org/10.1016/j.ophtha.2008.05.029>
- Wong TY, Klein R, Sharrett AR et al (2002) Retinal arteriolar narrowing and risk of diabetes mellitus in middle-aged persons. *JAMA* 287(19):2528–2533. <https://doi.org/10.1001/jama.287.19.2528>
- Mendis S, Puska P, Norrving B et al (2011) Global atlas on cardiovascular disease prevention and control. World Health Organization, Geneva. <https://iris.who.int/handle/10665/44701>. Accessed 10 June 2022
- Lopez-Luppo M, Nacher V, Ramos D et al (2017) Blood vessel basement membrane alterations in human retinal microaneurysms during aging. *Investig Ophthalmol Vis Sci* 58(2):1116–1131. <https://doi.org/10.1167/iovs.16-19998>
- Nema HV, Nema N (2018) Gems of ophthalmology: cataract surgery. Jaypee Brothers Medical Publishers, New Delhi. <https://www.jaypeedigital.com/book/9789352704019>. Accessed 15 June 2022
- Litjens G, Kooi T, Bejnordi BE et al (2017) A survey on deep learning in medical image analysis. *Med Image Anal* 42:60–88. <https://doi.org/10.1016/j.media.2017.07.005>
- Yang J, Deng H, Huang X et al (2020) Relational learning between multiple pulmonary nodules via deep set attention transformers. In: 2020 IEEE 17th international symposium on biomedical imaging (ISBI), pp 1875–1878. <https://doi.org/10.1109/ISBI45749.2020.9098722>
- Li F, Chen H, Liu Z et al (2019) Deep learning-based automated detection of retinal diseases using optical coherence tomography images. *Biomed Opt Express* 10(12):6204–6226. <https://doi.org/10.1364/BOE.10.006204>
- Zhang H, Niu K, Xiong Y et al (2019) Automatic cataract grading methods based on deep learning. *Comput Methods Prog Biomed* 182:104978. <https://doi.org/10.1016/j.cmpb.2019.07.006>
- Tan JH, Bhandary SV, Sivaprasad S et al (2018) Age-related macular degeneration detection using deep convolutional neural network. *Future Gener Comput Syst* 87:127–135. <https://doi.org/10.1016/j.future.2018.05.001>
- Bai H, Gao L, Quan X et al (2021) OTNet: a CNN method based on hierarchical attention maps for grading arteriosclerosis of fundus images with small samples. *Interdiscip Sci* 2021:1–14. <https://doi.org/10.1007/s12539-021-00479-8>
- Lin TY, RoyChowdhury A, Maji S (2015) Bilinear CNN models for fine-grained visual recognition. In: Proceedings of the IEEE international conference on computer vision, pp 1449–1457. <https://doi.org/10.1109/ICCV.2015.170>
- Guan Q, Huang Y, Zhong Z et al (2020) Thorax disease classification with attention guided convolutional neural network. *Pattern Recogn Lett* 131:38–45. <https://doi.org/10.1016/j.patrec.2019.11.040>
- Hu T, Qi H, Huang Q et al (2019) See better before looking closer: weakly supervised data augmentation network for fine-grained visual classification. *arXiv*. <http://arxiv.org/abs/1901.09891>
- Du R, Chang D, Bhunia AK et al (2020) Fine-grained visual classification via progressive multi-granularity training of jigsaw patches. In: European conference on computer vision, pp 153–168. [https://doi.org/10.1007/978-3-030-58565-5\\_10](https://doi.org/10.1007/978-3-030-58565-5_10)
- Zhao Z, Chopra K, Zeng Z et al (2020) Sea-net: squeeze-and-excitation attention net for diabetic retinopathy grading. In: 2020 IEEE international conference on image processing (ICIP), pp 2496–2500. <https://doi.org/10.1109/ICIP40778.2020.9191345>
- Zhang F, Li M, Zhai G et al (2021) Multi-branch and multi-scale attention learning for fine-grained visual categorization. In: MultiMedia modeling: 27th international conference, pp 136–147. [https://doi.org/10.1007/978-3-030-67832-6\\_12](https://doi.org/10.1007/978-3-030-67832-6_12)
- Karri SPK, Chakraborty D, Chatterjee J (2017) Transfer learning based classification of optical coherence tomography images with diabetic macular edema and dry age-related macular degeneration. *Biomed Opt Express* 8(2):579–592. <https://doi.org/10.1364/BOE.8.000579>
- Bravo MA, Arbeláez PA (2017) Automatic diabetic retinopathy classification. In: 13th international conference on medical information processing and analysis, pp 446–455. <https://doi.org/10.1117/12.2285939>
- Zhang J, Xie Y, Xia Y et al (2019) Attention residual learning for skin lesion classification. *IEEE Trans Med Imaging* 38(9):2092–2103. <https://doi.org/10.1109/TMI.2019.2893944>
- Schlemper J, Oktay O, Schaap M et al (2019) Attention gated networks: learning to leverage salient regions in medical images. *Med Image Anal* 53:197–207. <https://doi.org/10.1016/j.media.2019.01.012>
- Al-Antary MT, Arafa Y (2021) Multi-scale attention network for diabetic retinopathy classification. *IEEE Access* 9:54190–54200. <https://doi.org/10.1109/ACCESS.2021.3070685>
- Wang Z, Yin Y, Shi J, et al (2017) Zoom-in-net: deep mining lesions for diabetic retinopathy detection. In: Medical image computing and computer assisted intervention-MICCAI 2017: 20th international conference, pp 267–275. [https://doi.org/10.1007/978-3-319-66179-7\\_31](https://doi.org/10.1007/978-3-319-66179-7_31)
- He A, Li T, Li N et al (2020) CABNet: category attention block for imbalanced diabetic retinopathy grading. *IEEE Trans Med Imaging* 40(1):143–153. <https://doi.org/10.1109/TMI.2020.3023463>
- Xie H, Zeng X, Lei H et al (2021) Cross-attention multi-branch network for fundus diseases classification using SLO images. *Med Image Anal* 71:102031. <https://doi.org/10.1016/j.media.2021.102031>
- Al-Masni MA, Kim DH, Kim TS (2020) Multiple skin lesions diagnostics via integrated deep convolutional networks for segmentation and classification. *Comput Methods Prog Biomed* 190:105351. <https://doi.org/10.1016/j.cmpb.2020.105351>
- Hasan MK, Elahi MTE, Alam MA et al (2022) DermoExpert: skin lesion classification using a hybrid convolutional neural network through segmentation, transfer learning, and augmentation. *Inform Med Unlocked* 28:100819. <https://doi.org/10.1016/j.imu.2021.100819>
- Xie Y, Zhang J, Xia Y et al (2020) A mutual bootstrapping model for automated skin lesion segmentation and classification. *IEEE Trans Med Imaging* 39(7):2482–2493. <https://doi.org/10.1109/TMI.2020.2972964>
- Wu YH, Gao SH, Mei J et al (2021) Jcs: an explainable Covid-19 diagnosis system by joint classification and segmentation. *IEEE*

- Trans Image Process 30:3113–3126. <https://doi.org/10.1109/TIP.2021.3058783>
32. Li L, Verma M, Nakashima Y, et al (2020) Iternet: retinal image segmentation utilizing structural redundancy in vessel networks. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision, pp 3656–3665. <https://doi.org/10.1109/wacv45572.2020.9093621>
  33. Staal J, Abramoff MD, Niemeijer M et al (2004) Ridge-based vessel segmentation in color images of the retina. IEEE Trans Med Imaging 23(4):501–509. <https://doi.org/10.1109/TMI.2004.825627>
  34. Owen CG, Rudnicka AR, Mullen R et al (2009) Measuring retinal vessel tortuosity in 10-year-old children: validation of the computer-assisted image analysis of the retina (CAIAR) program. Invest Ophthalmol Vis Sci 50(5):2004–2010. <https://doi.org/10.1167/iovs.08-3018>
  35. Wang X, Girshick R, Gupta A, et al (2018) Non-local neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 7794–7803. <https://doi.org/10.1109/CVPR.2018.00813>
  36. Ou X, Gao L, Quan X et al (2022) BFENet: a two-stream interaction CNN method for multi-label ophthalmic diseases classification with bilateral fundus images. Comput Methods Prog Biomed 219:106739. <https://doi.org/10.1016/j.cmpb.2022.106739>
  37. Ketkar N, Moolayil J, Ketkar N et al (2021) Introduction to pytorch. In: Deep learning with python: learn best practices of deep learning models with PyTorch. A Press, New York, pp 27–91. [https://doi.org/10.1007/978-1-4842-5364-9\\_2](https://doi.org/10.1007/978-1-4842-5364-9_2)
  38. Liu W, Rabinovich A, Berg AC (2015) Parsenet: looking wider to see better. arXiv. <http://arxiv.org/abs/1506.04579>
  39. Zagoruyko S, Komodakis N (2016) Wide residual networks. arXiv. <http://arxiv.org/abs/1605.07146>
  40. Srivastava RK, Greff K, Schmidhuber J (2015) Training very deep networks. arXiv. <http://arxiv.org/abs/1507.06228>

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

## Authors and Affiliations

Xiongwen Quan<sup>1</sup> · Xingyuan Ou<sup>1</sup> · Li Gao<sup>2</sup> · Wenya Yin<sup>1</sup> · Guangyao Hou<sup>1</sup> · Han Zhang<sup>1</sup> 

✉ Han Zhang  
zhanghan@nankai.edu.cn

Xiongwen Quan  
quanxw@nankai.edu.cn

Xingyuan Ou  
ouxingyuan@mail.nankai.edu.cn

Li Gao  
gaoli7506@163.com

Wenya Yin  
2120210426@mail.nankai.edu.cn

Guangyao Hou  
2120220551@mail.nankai.edu.cn

<sup>1</sup> National Key Laboratory of Intelligent Tracking and Forecasting for Infectious Diseases, Engineering Research Center of Trusted Behavior Intelligence, Ministry of Education, College of Artificial Intelligence, Nankai University, Tianjin 300000, China

<sup>2</sup> Ophthalmology, Tianjin Huanhu Hospital, Tianjin 300000, China