



# LKG-Net: lightweight keratoconus grading network based on corneal topography

SONG GAO,<sup>1,5</sup>  YINGJIE CHEN,<sup>2,5</sup> FEI SHI,<sup>1</sup>  YUANYUAN PENG,<sup>1</sup> CHENAN XU,<sup>3</sup>  ZHONGYUE CHEN,<sup>1</sup> WEIFANG ZHU,<sup>1</sup>  XIN XU,<sup>1</sup> WEI TANG,<sup>1</sup> ZHIWEI TAN,<sup>1</sup> YUE XU,<sup>2</sup> YARU REN,<sup>4</sup> XIAOFENG ZHANG,<sup>2,4,6</sup> AND XINJIAN CHEN<sup>1,3,\*</sup>

<sup>1</sup>MIPAV Lab, School of Electronics and Information Engineering, Soochow University, Suzhou, 215006, China

<sup>2</sup>Department of Ophthalmology, Dushu Lake Hospital Affiliated to Soochow University, Suzhou, 215006, China

<sup>3</sup>State Key Laboratory of Radiation Medicine and Protection, Collaborative Innovation Center of Radiological Medicine of Jiangsu Higher Education Institutions, and School for Radiological and Interdisciplinary Sciences (RAD-X), Soochow University, Suzhou 215006, China

<sup>4</sup>Department of Ophthalmology, Soochow University-Affiliated First Hospital, Suzhou, Suzhou, 215006, China

<sup>5</sup>These authors contributed equally to this paper

<sup>6</sup>zhangxiaofeng@suda.edu.cn

\*xjchen@suda.edu.cn

**Abstract:** Keratoconus (KC) is a noninflammatory ectatic disease characterized by progressive thinning and an apical cone-shaped protrusion of the cornea. In recent years, more and more researchers have been committed to automatic and semi-automatic KC detection based on corneal topography. However, there are few studies about the severity grading of KC, which is particularly important for the treatment of KC. In this work, we propose a lightweight KC grading network (LKG-Net) for 4-level KC grading (Normal, Mild, Moderate, and Severe). First of all, we use depth-wise separable convolution to design a novel feature extraction block based on the self-attention mechanism, which can not only extract rich features but also reduce feature redundancy and greatly reduce the number of parameters. Then, to improve the model performance, a multi-level feature fusion module is proposed to fuse features from the upper and lower levels to obtain more abundant and effective features. The proposed LKG-Net was evaluated on the corneal topography of 488 eyes from 281 people with 4-fold cross-validation. Compared with other state-of-the-art classification methods, the proposed method achieves 89.55% for weighted recall (W\_R), 89.98% for weighted precision (W\_P), 89.50% for weighted F1 score (W\_F1) and 94.38% for Kappa, respectively. In addition, the LKG-Net is also evaluated on KC screening, and the experimental results show the effectiveness.

© 2023 Optica Publishing Group under the terms of the [Optica Open Access Publishing Agreement](#)

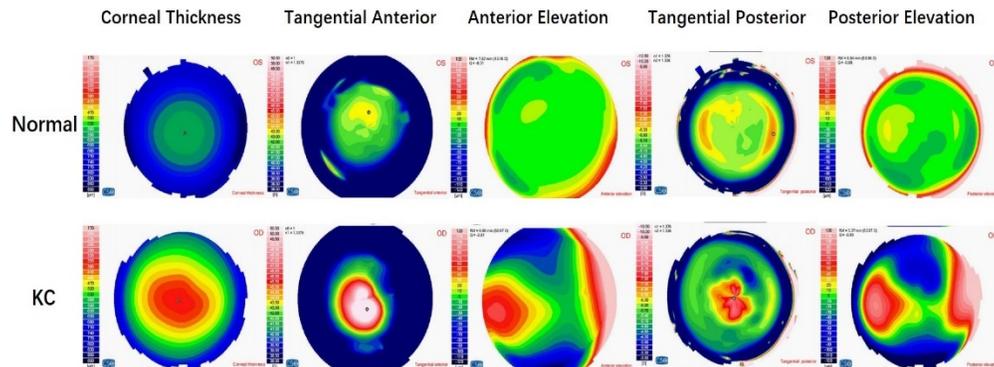
## 1. Introduction

Keratoconus (KC) is an ophthalmic disease characterized by a dilated, centrally thinned cornea protruding forward in a conical shape, which is usually asymmetric [1,2]. It tends to develop in adolescence and often results in highly irregular myopic astigmatism, with acute corneal edema, scarring, and significant loss of visual acuity in advanced stages [3–5]. Studies have shown that the prevalence in the general population is approximately between 5 and 23 per 10,000 [1,4,5]. Although the etiology of KC is still unknown, researchers have found that the incidence of this disease varies by ethnic, environmental and genetic factors. According to a study in the United Kingdom, a prevalence of 4:1, and an incidence of 4.4:1 was found in Asians compared to whites

[6]. Multiple genes are involved in the development of KC, and the prevalence of positive family history is relatively high [7–9].

The treatment of KC depends on the progression of the disease and its severity. Traditionally, spectacles have provided acceptable vision for mild to moderate patients. As the disease progresses and irregular astigmatism develops, contact lenses can correct irregular astigmatism to provide better vision for moderate patients [10,11], such as hybrid lenses, piggyback lenses, or scleral lenses [12]. Patients with severe KC can be treated with keratoplasty [10]. Other surgical treatment options include intra-corneal rings segments [13,14], corneal cross-linking [15,16], laser procedures [17,18], intra-ocular lens implants [19,20] or a combination of these [21]. It is evident that the development of an effective KC grading algorithm will provide effective guidance and assistance to ophthalmologists in the treatment plan for their patients.

Ophthalmologists often diagnose KC based on corneal topographies because they provide detailed data and morphological characteristics of the cornea. In this paper, we use the data obtained from a 3D corneal topography and anterior segment analysis system, which gives five color-coded maps, which are corneal thickness (170 to 890  $\mu\text{m}$ , 30  $\mu\text{m}$  step), tangential anterior (38.00 to 50.50 D, 0.50 D step), anterior elevation (–120 to 120  $\mu\text{m}$ , 10  $\mu\text{m}$  step), tangential posterior (–10.50 to –3.00 D, 0.50 D step), and posterior elevation (–120 to 120  $\mu\text{m}$ , 10  $\mu\text{m}$  step) maps, as shown in Fig. 1. Corneal curvature, elevation and thickness were measured by Scheimpflug camera and Placido corneal topography. During the scanning process, a series of 25 Scheimpflug images (meridian) and 1 Placido top view image were obtained. The annular edges were detected on the Placido image in order to calculate height, slope and curvature data using the arc-step method with conic curves, and the corresponding corneal topographic maps can be generated [22]. Based on the corneal data obtained from various devices, researchers have proposed various KC grading systems, such as the Amsler-Krumeich classification system [23], the Shabayek Ali ó grading system [24], the keratoconus severity score (KSS) system [25] and the ABCD grading system [26]. The Amsler-Krumeich classification system classifies cases based on the amount of myopia and astigmatism, corneal thickness or scarring, and central k-readings, and is a widely used classification standard by researchers.



**Fig. 1.** Examples of corneal topographies of normal and KC eyes. Five color-coded maps (corneal thickness, tangential anterior, anterior elevation, tangential posterior, and posterior elevation map) measured with an anterior segment analyzer.

In recent years, with the continuous development of artificial intelligence (AI) technologies, it has become more and more common to use AI techniques, especially machine learning (ML), to evaluate and detect various diseases. Especially in the field of deep learning (DL), AlexNet [27], VGGNet [28], GoogLeNet [29], ResNet [30], DenseNet [31] and other convolutional neural networks (CNNs) have been proposed for image recognition in many different scenes. In ophthalmology, DL methods have been widely used in the diagnosis and screening of diseases,

such as diabetes retinopathy (DR) [32], age-related macular degeneration (AMD) [33], retinopathy of prematurity (ROP) [34], etc., as described in [35]. In the past, many researches on KC mainly focused on the screening and detection of KC. Most of them used some machine learning algorithms to diagnose the disease by processing and analyzing the corneal indices obtained from the devices [22,36–39]. Accardo et al. used nine parameters obtained from corneal topography as the input of three-layer neural network to achieve clinical classification (normal, KC, other changes). The result demonstrated the quality and value of the neural network approach in identifying the topographic patterns of KC [37]. Arbelaez et al. used support vector machine to process the data from anterior and posterior corneal surfaces and pachymetry measured with a Scheimpflug camera combined with Placido corneal topography. The classification algorithm showed high accuracy, precision, sensitivity, and specificity in discriminating among abnormal eyes, eyes with KC or subclinical KC, and normal eyes. Including the posterior corneal surface and thickness parameters markedly improved the sensitivity in the diagnosis of subclinical KC [22]. With the emergence of advanced imaging devices, researchers began to use DL technology to research end-to-end automated KC detection algorithms based on corneal topography [40–43]. Lavric et al. integrated the SyntEyes KTC model to generate corneal topography to expand the experimental data and obtained excellent results on the dataset using CNN. This work is considered to be the first to introduce CNN into KC diagnosis [40]. Kuo et al. used three CNNs to detect KC based on corneal topography, and the sensitivity and specificity of all CNN models were over 0.90. The pixel-wise discriminative features and the heat maps of the prediction layer in the VGG16 model both revealed the focus on the largest gradient difference of topographic maps, which was corresponding to the diagnostic clues of ophthalmologists [42]. Kazutaka et al. separately trained six neural networks from each image of six color-coded maps, and integrated them by averaging the six outputs. The result showed an accuracy of 0.874 in classifying the stage of the disease [44].

In conclusion, DL is conducive to the severity grading of KC based on corneal topography. However, most of the current studies on KC use traditional CNNs to process each corneal map separately, without feature fusion to extract more effective features. Moreover, some models are complex, and some require to manually extract advanced features. Therefore, it is not suitable to be integrated into devices for automatic detection. Recently, various lightweight networks based on depth-wise separable convolution have been emerging, and have achieved good classification performance in different image classification tasks [45–48]. In addition, as seen in Fig. 1, we can notice these key features of the corneal topography with KC, and several studies have shown that the attention mechanism can help learn useful features to improve the performance of the network [49,50,52]. Based on CNNs, we propose an end-to-end lightweight KC grading network named LKG-Net. Using depth-wise separable convolution, we designed a self-attention based feature extraction block (SaFEB), which can significantly reduce the model complexity while obtaining more important features. Meanwhile, we propose a multi-level feature fusion module (MIFFM), which can adaptively calibrate features of upper and lower channel dimensions and fuse features. The whole diagnosis process is automated, unlike previous research, which requires manual data collection or complex manual feature extraction. The main contributions of this paper can be summarized as follows:

- (1) A lightweight classification network based on a novel feature extraction block is proposed for 4-level KC severity grading, which has excellent performance despite of a significant reduction in the model complexity. The block can use different dimensions of feature information to generate the attention weight matrix, so as to obtain more abundant and effective features.

- (2) A new multi-level feature fusion module is proposed, which can adaptively calibrate the features of the upper and lower channel dimensions and fuse them according to a certain weight, further improving the accuracy of KC severity grading.
- (3) Extensive experiments have been carried out to evaluate the effectiveness of the proposed method. Experimental results show that this method is superior to other most advanced classification methods in KC severity classification task. In addition, the method is also evaluated on KC screening, and the experimental results show the effectiveness.

The remainder of this paper is organized as follows: Section 2 introduces the proposed method of automatic KC severity grading. The experimental results are presented in detail in Section 3. In section 4, we conclude this paper and suggest future work.

## 2. Methodology

### 2.1. Overview

Figure 2 gives an overview of the novel end-to-end KC severity grading method based on LKG-Net. Using the corneal topography reports exported from the imaging device, we automatically extract the five color-coded maps from the report and further process them to generate the data format we need as input to the network, and LKG-Net will classify the samples into four categories: normal, mild, moderate and severe. All these steps can be integrated into a fully automated diagnostic approach from the corneal topography report.

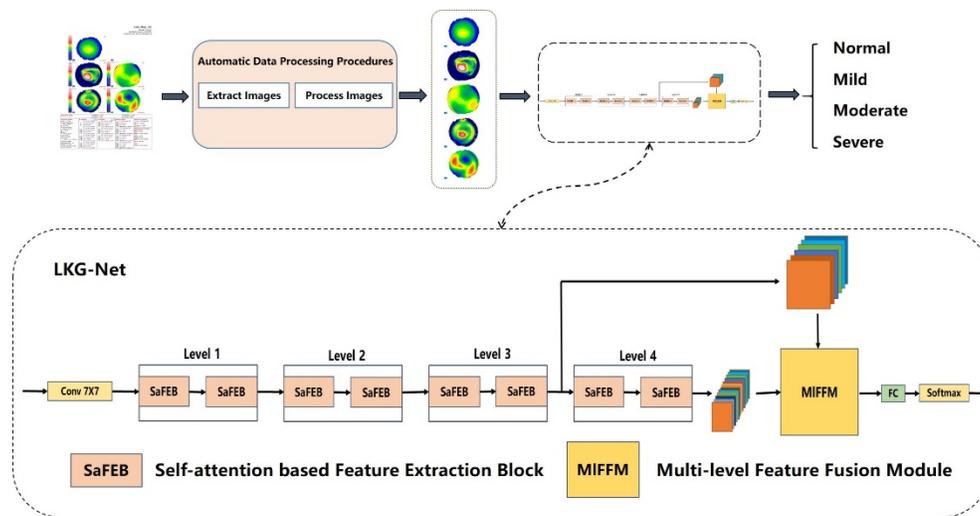


Fig. 2. Overview of the proposed method.

### 2.2. LKG-Net architecture

As shown in Fig. 2, using ResNet18 [30] as backbone, the proposed LKG-Net mainly includes self-attention based feature extraction blocks (SaFEB) and a multi-level feature fusion module (MIFFM). We use SaFEB to replace the basic block in ResNet18, and add MIFFM to fuse the feature maps output from level 3 and level 4 to obtain richer feature information.

#### 2.2.1. Self-attention based feature extraction block (SaFEB)

The most common practice to obtain a high-quality model is to increase the depth of the model, but the deeper the network, the more likely it is to suffer from gradient disappearance [29,51]. If more

effective feature extraction modules can be designed, a shallower network can be sufficient for the specific task. Recently, with the advent of Transformer [52], the modules based on self-attention have achieved performance comparable to or even better than the corresponding CNN modules on many vision tasks. The self-attention modules usually use a weighted averaging operation based on the input feature context to dynamically compute attention weights through a similarity function between relevant pixel pairs. This flexibility allows the attention module to adaptively focus on different regions and capture more features [53]. Based on the idea of self-attention, we design a novel feature extraction block, which uses different dimensions of feature information to generate its attention weight matrix, so as to obtain more abundant and effective features. In addition, to reduce the model complexity, we use depth-wise separable convolution (DSC) instead of traditional convolution. DSC consists of deep-wise (Dw) convolution and  $1 \times 1$  convolution. In Dw convolution, a convolution kernel has only one dimension, and a channel is convolved by only one convolution kernel. It can significantly reduce the number of parameters while maintaining the model performance, and thus is widely used in lightweight networks [45–47].

The feature extraction block based on the self-attention mechanism obtains the average and maximum information of the corresponding dimensions of each channel of the input features, and generates attention weights based on the rich information of these different dimensions, which can guide the network to extract the important feature information. As shown in Fig. 3, given the input feature maps  $F_i \in \mathbb{R}^{C_i, H_i, W_i}$ , it passes through batch normalization (BN) and ReLU6, which enhances the regularization of the model and makes it easier to optimize [54]. Then, the output feature maps are convolved by  $3 \times 3$  deep-wise separable convolutions to obtain the feature maps  $F_q \in \mathbb{R}^{C_o, H_o, W_o}$ ,  $F_k \in \mathbb{R}^{C_o, H_o, W_o}$ , and  $F_v \in \mathbb{R}^{C_o, H_o, W_o}$  as illustrated in Eq. (1). The average values of the specified dimension of the feature maps  $F_q$  and  $F_k$  are obtained as  $W_{qA} \in \mathbb{R}^{C_o, H_o, 1}$  and  $W_{kA} \in \mathbb{R}^{C_o, 1, W_o}$ , and the maximum values of the specified dimension of the feature maps  $F_q$  and  $F_k$  are obtained as  $W_{qM} \in \mathbb{R}^{C_o, 1, W_o}$  and  $W_{kM} \in \mathbb{R}^{C_o, H_o, 1}$ , as illustrated in Eq. (2). After multiplying and adding them, the attention weight matrix  $W_v \in \mathbb{R}^{C_o, H_o, W_o}$  is obtained by the Hard-sigmoid function, as illustrated in Eq. (3). Then the feature maps  $F_v$  are multiplied with  $W_v$  to apply the attention weights. Finally, the output feature maps  $F_o \in \mathbb{R}^{C_o, H_o, W_o}$  are obtained by adding the result with the feature maps  $F_i$  after  $1 \times 1$  convolution, as illustrated in Eq. (4). Because of the need to multiply feature dimensions in the first block of level 2, 3, and 4,  $1 \times 1$  convolution is used in the final summation to make the feature size consistent.

$$F_q = DSC_3(BR(F_i)) \quad F_k = DSC_3(BR(F_i)) \quad F_v = DSC_3(BR(F_i)) \quad (1)$$

$$W_{qA}, W_{kA} = Avg(F_q, F_k) \quad W_{kM}, W_{qM} = Max(F_q, F_k) \quad (2)$$

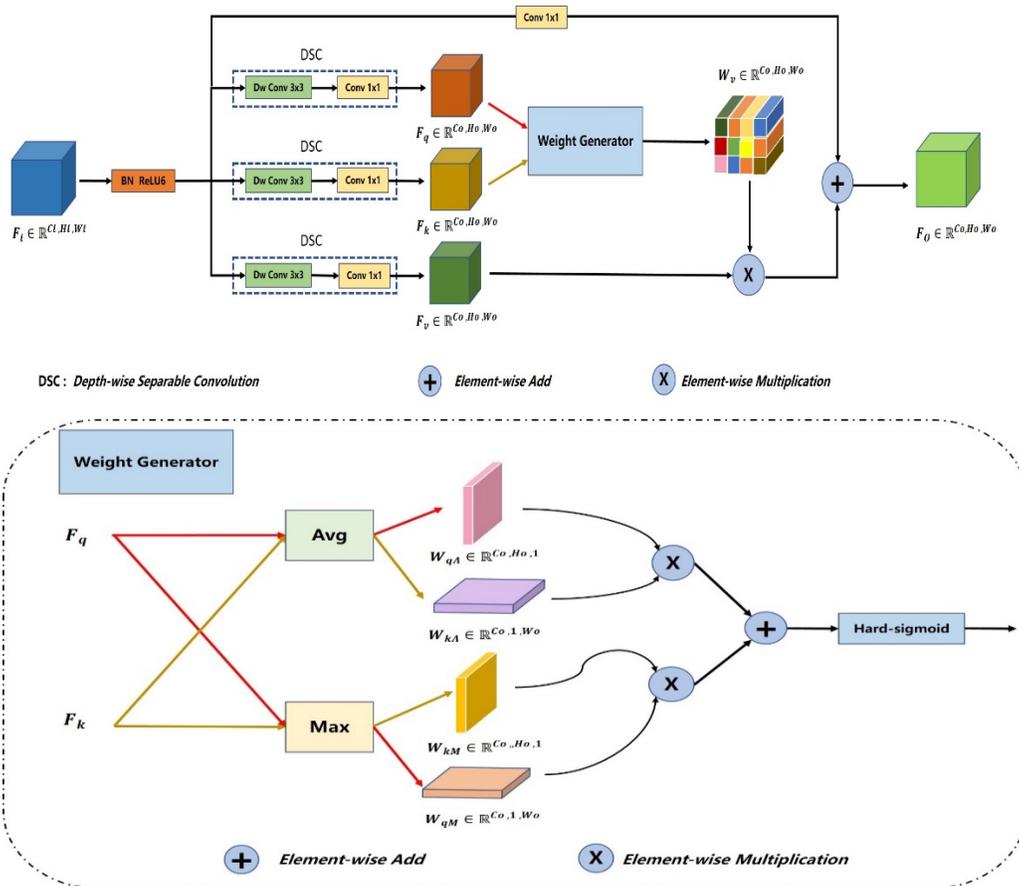
$$W_v = HS(EM(W_{qA}, W_{kA}) + EM(W_{kM}, W_{qM})) \quad (3)$$

$$F_o = EM(F_v, W_v) + C1(F_i) \quad (4)$$

where  $F_i \in \mathbb{R}^{C_i, H_i, W_i}$  denotes the input feature maps,  $C_i$  is the number of channels,  $H_i$  is the height and  $W_i$  is the width of feature maps,  $BR$  represents batch normalization and ReLU6,  $DSC_3$  represents the deep-wise separable convolution with kernel size of  $3 \times 3$ ,  $Max$  represents the maximum operation,  $Avg$  represents the average operation,  $EM$  represents the element-wise multiplication,  $HS$  represents the Hard-sigmoid function, and  $C1$  represents  $1 \times 1$  convolution.

### 2.2.2. Multi-level feature fusion module (MIFFM)

Similar to human attention, the essence of the attention mechanism in DL focuses on useful feature information. Many studies have shown that the performance of CNN models can be improved by focusing on important features and suppressing irrelevant features through the attention mechanism [55–57]. As shown in Fig. 1, the corneal topography maps of KC patients have distinct features compared with normal subjects. Through convolution operation, these



**Fig. 3.** The illustration of self-attention based feature extraction block (SaFEB).

features can be extracted to generate corresponding feature maps. However, some feature maps without key information may reduce the judgment ability of the model. The channel attention mechanism can obtain the importance of each feature map by automatic learning, and then assigns a weight to each feature, so that the network can focus on certain feature channels, improve the importance of the feature maps that are useful for the current task, and suppress the feature channels that are not useful for the current task [50]. In addition, sometimes some key information from the previous layer may be lost at higher level during the learning process. Therefore, we design a multi-level feature fusion module (MIFFM), which can adaptively calibrate the features of the upper and lower channel dimensions based on the channel attention mechanism, and fuse the calibrated features of the two levels according to a certain weight to reinforce the features.

As shown in Fig. 4, in the channel attention mechanism, the maximum pooled features and the average pooled features go through  $1 \times 1$  convolution and ReLU6, respectively, and the obtained results are added. Then channel attention weights are generated with a range of 0-1 through the hard-sigmoid activation function. Finally, the weights are multiplied with the feature map of corresponding channel to adjust their importance. The hard-sigmoid function is implemented based on ReLU6 and is very close to the sigmoid function, but it will become simpler in the calculation of derivatives [47]. Given the  $N-1$ <sup>th</sup> level output  $F_{N-1} \in \mathbb{R}^{C,H,W}$  and the  $N$ <sup>th</sup> level output  $F_N \in \mathbb{R}^{C \times 2, H/2, W/2}$ , the dimension and size of the  $N-1$ <sup>th</sup> level features are first adjusted to be consistent with those of the  $N$ <sup>th</sup> level using  $1 \times 1$  convolution, and then the feature maps

are rescaled separately using the channel attention mechanism. The calibrated results are fused according to a fusion ratio  $w$  to obtain richer feature information. Finally,  $1 \times 1$  convolution is used to eliminate information redundancy. The output  $F_{N+1} \in \mathbb{R}^{C \times 2, H/2, W/2}$  of MIFFM can be computed as:

$$F_{N+1} = C1((CA(C_2(F_{N-1}))) \times w + (CA(F_N)) \times (1 - w)) \quad (5)$$

where  $C$  is the number of channels,  $H$  is the height and  $W$  is the width of feature map.  $CA$  represents the channel attention mechanism,  $C1$  represents the  $1 \times 1$  convolution, and  $w \in [0, 1]$  represents the fusion ratio of the  $N-1^{\text{th}}$  level.

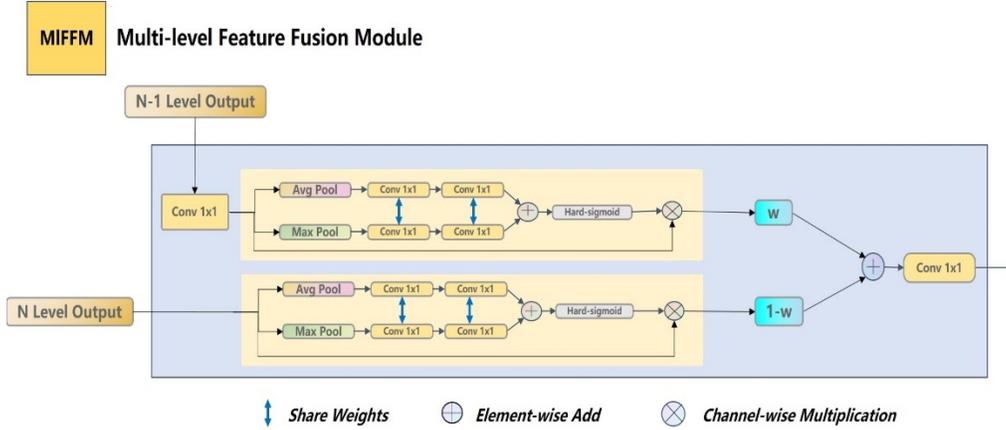


Fig. 4. The illustration of multi-level feature fusion module (MIFFM).

### 2.3. Loss functions

Our proposed network is aimed at grading the severity of KC, which is a multi-classification task. In the supervised classification tasks of deep learning, people often use cross entropy (CE) loss function. It can be used not only for binary classification tasks but also for multi-category tasks, and can effectively measure the difference between the distribution learned by the model and the true distribution of labels. CE loss is generally defined as follows:

$$L_m = -\frac{1}{N} \sum_{i=0}^{N-1} \sum_{k=0}^{K-1} y_{i,k} \ln p_{i,k} \quad (6)$$

where  $y_{i,k}$  denotes the ground truth label indicating the  $i$ -th sample belongs to category  $k$ ,  $K$  is the number of categories, and  $N$  is the number of samples per mini-batch.  $p_{i,k}$  denotes the predicted probability that the  $i$ -th sample belongs to the category  $k$ .

## 3. Experiments and results

### 3.1. Dataset

The dataset of this study was collected in the Ophthalmology Center of the First Affiliated Hospital of Soochow University from 2017 to 2021. Corneal curvature, elevation, and thickness measurements were obtained by means of a Scheimpflug camera combined with Placido corneal topography (Sirius, software version 1.2, CSO, Firenze, Italy). The measurements were carried out by an experienced examiner according to the guidelines. An informed consent was obtained from the guardians of each subject to perform all the imaging procedures. The collection and analysis of the examination reports were approved by the review committee of the First Affiliated Hospital of Soochow University and complied with the principles of the Declaration of Helsinki.

On the premise of ensuring the quality of the corneal topographic maps, examination reports were collected from 488 eyes of 281 people, including 236 normal and 252 KC eyes. The processed data of each eye contained five corneal topographic maps (shown in Fig. 1), which were corneal thickness (TK), tangential anterior (TA), anterior elevation (AE), tangential posterior (TP), and posterior elevation (PE) maps, and the resolution of the images was  $640 \times 480$ . Each type of topographic map used a fixed color scale where one color represented a given range of intervals. The ground truth of the data was given by a professional ophthalmologist, who identified KC eyes from normal ones, and classified KC eyes into three categories: mild (138 eyes), moderate (63 eyes), and severe (51 eyes), based on both clinical experience and analysis of indicators from the examination reports, with reference to the Amsler-Krumeich classification system [23]. Considering the amount of data, a 4-fold cross-validation strategy was used in order to assess the validity of the proposed method. To reduce the computational cost, all corneal topographic maps were down sampled to  $256 \times 256$  using bilinear interpolation and the intensities of each color channel were normalized to  $[0,1]$ .

### 3.2. Experimental setup

#### 3.2.1. Parameter setting

The implementation of the proposed LKG-Net is based on the PyTorch platform. A NVIDIA GeForce RTX 3090 GPU with 24GB memory was used to train the model with back-propagation algorithm by minimizing the loss function as shown in Eq. (6). The adaptive momentum (Adam) algorithm with an initial learning rate of 0.0001, momentum of 0.9 and weight decay of 0.0001 was used to optimize the network. Besides, the batch size was set to 8 and the number of epochs was 200.

#### 3.2.2. Evaluation metrics

As can be seen from the dataset, the number of samples in each category is unbalanced. In order to comprehensively and fairly evaluate the classification performance of different methods, we use common metrics including weighted recall (W\_R), weighted precision (W\_P), weighted F1 score (W\_F1) and Kappa [58] index to evaluate the KC severity grading performance. In addition, the indicators of accuracy, precision, recall, F1 score and area under the ROC curve (AUC) are used to evaluate the performance of the proposed method on KC screening.

### 3.3. Comparison experiments

with a 4-fold cross validation strategy, we compare the proposed method with other excellent CNN based classification networks, including VGG16 [28], InceptionV2, ResNet34 [36], ResNet101 [36], ResNext50 [59], SE\_ResNet50 [50], SE\_ResNext50 [50], DenseNet121 [31], EfficientNet\_b0 [33], EfficientNet\_b2 [33], MobileNet\_v3\_small [47], MobileNet\_v3\_large [47], Mutil-ResNet18\_FF and Mutil-ResNet18\_DF [44]. The Mutil-ResNet18\_FF means we use five weight-sharing networks to process images separately, and then perform feature fusion by concatenating features of the last layers. The Mutil-ResNet18\_DF means we use five weight-sharing networks to process images separately, and then perform decision-layer fusion by averaging the predicted probability at decision layers. For convenience, we call the basic ResNet18 [36] as the Baseline method. Except for Mutil-ResNet18\_FF and Mutil-ResNet18\_DF, we feed five corneal topographic maps as a 15-channel input into the network for training and testing, and the experimental results are shown in Table 1.

As we can see from Table 1, the proposed LKG-Net outperforms the mentioned state-of-the-art CNN based methods. Firstly, compared to the Baseline, the performance of our method has been improved, which improves the W\_R, W\_P, W\_F1 and Kappa by 2.46%, 2.15%, 3.45% and 1.67%, respectively. The complexity of our model is greatly reduced and the number of parameters is only one fifth of that of Baseline, making it a lightweight model. Secondly, our method has significant

Table 1. Comparison of KC Grading Results of Different Methods

Methods	W_R	W_P	W_F1	Kappa	Parameter (M)
VGG16 [28]	0.8668 ± 0.0169	0.8658 ± 0.0210	0.8637 ± 0.0180	0.9265 ± 0.0106	165.75
InceptionV2 [51]	0.8811 ± 0.0331	0.8816 ± 0.0321	0.8790 ± 0.0335	0.9241 ± 0.0302	7.37
ResNet34 [36]	0.8689 ± 0.0115	0.8678 ± 0.0249	0.8590 ± 0.0214	0.9225 ± 0.0129	21.32
ResNet101 [36]	0.8648 ± 0.0156	0.8647 ± 0.0218	0.8577 ± 0.0189	0.9278 ± 0.0114	42.55
ResNext50 [59]	0.8689 ± 0.0177	0.8697 ± 0.0156	0.8673 ± 0.0177	0.9224 ± 0.0195	23.03
SE_ResNet50 [50]	0.8729 ± 0.0271	0.8739 ± 0.0279	0.8715 ± 0.0295	0.9289 ± 0.0261	26.08
SE_ResNext50 [50]	0.8832 ± 0.0103	0.8835 ± 0.0115	0.8798 ± 0.0098	0.9283 ± 0.0321	26.08
DenseNet121 [31]	0.8771 ± 0.0241	0.8905 ± 0.0285	0.8753 ± 0.0250	0.9221 ± 0.0118	6.99
EfficientNet_b0 [33]	0.8627 ± 0.0270	0.8605 ± 0.0288	0.8582 ± 0.0292	0.9122 ± 0.0199	4.02
EfficientNet_b2 [33]	0.8361 ± 0.0444	0.8297 ± 0.0741	0.8172 ± 0.0557	0.8869 ± 0.0470	9.11
MobileNetV3_small [47]	0.8156 ± 0.0171	0.8094 ± 0.0163	0.8090 ± 0.0174	0.8928 ± 0.0171	1.83
MobileNetV3_large [47]	0.8423 ± 0.0226	0.8590 ± 0.0233	0.8412 ± 0.0244	0.9075 ± 0.0257	4.21
Mutil-ResNet18_FF	0.8136 ± 0.0310	0.8083 ± 0.0414	0.7867 ± 0.0398	0.8739 ± 0.0305	12.49
Mutil-ResNet18_DF [44]	0.8156 ± 0.0303	0.8163 ± 0.0398	0.8123 ± 0.0366	0.8712 ± 0.0533	11.18
Baseline [36]	0.8709 ± 0.0155	0.8783 ± 0.0206	0.8605 ± 0.0223	0.9271 ± 0.0193	11.22
LKG-Net (Ours)	<b>0.8955 ± 0.0263</b>	<b>0.8998 ± 0.0274</b>	<b>0.8950 ± 0.0280</b>	<b>0.9438 ± 0.0180</b>	<b>2.32</b>

advantages in terms of model performance and complexity compared to other ResNet-based networks with attention mechanism including SE\_Resnet50 and SE\_ResNext50. Compared with Mutil-ResNet18\_FF and Mutil-ResNet18\_DF, the result of Baseline is significantly improved with 15-channel input. These corneal maps are generated based on the same eye surface and reflect the morphological features of the cornea from different aspects, so the 15-channel input is more convenient for the network to extract the correlation features among the corneal maps to achieve better classification result. Compared to MobileNetV3\_small, which is a typical representative of lightweight networks with comparable model complexity, our method has also made great improvement in performance. In addition, comparing the experimental results of different versions of the same network in Table 1, it can be found that the performance of the model will gradually decrease with the increase of network depth, especially for the small and large versions of MobileNetV3. Although VGG16 has a simple structure with fewer layers, its performance is relatively high. As shown in Fig. 1, corneal topographic maps are relatively simple color-coded images, which are not full of many complex targets, so complex networks are not necessary. Based on these experimental results and data analysis, we infer that the dataset is more suitable to be processed by the network with fewer layers, which is why we choose simple ResNet18 as the backbone, and the experimental results also verify our inference.

### 3.4. Experiments for different input combinations of corneal topography

To compare the effects of different combinations of the topographic maps in KC grading, we conduct the experiments by feeding the LKG-Net with different inputs, and the results are shown in Table 2. The first five rows give the results when we input one image, and the sixth and seventh rows show the results when the network receives the posterior and anterior maps as a whole input respectively. The last row gives the results when the network uses five maps as a whole input for classification.

Firstly, compared to other schemes of input data, the network obtains the best performance with five-map input. Secondly, for single image input, the network has excellent performance

Table 2. KC Grading Results of Different Input Combinations of Corneal Topography

TK	LKG-Net Input				W_R	W_P	W_F1	Kappa
	TA	AE	TP	PE				
√					0.7500 ± 0.0280	0.7356 ± 0.0393	0.7355 ± 0.0365	0.8190 ± 0.0550
	√				0.7480 ± 0.0350	0.7433 ± 0.0434	0.7337 ± 0.0412	0.8227 ± 0.0577
		√			0.7992 ± 0.0345	0.7975 ± 0.0400	0.7917 ± 0.0303	0.8790 ± 0.0276
			√		0.7608 ± 0.0375	0.7641 ± 0.0425	0.7674 ± 0.0365	0.8540 ± 0.0355
				√	0.7805 ± 0.0232	0.7811 ± 0.0337	0.7724 ± 0.0207	0.8788 ± 0.0395
	√	√			0.8382 ± 0.0123	0.8365 ± 0.0164	0.8333 ± 0.0163	0.9093 ± 0.0063
			√	√	0.8256 ± 0.0082	0.8271 ± 0.0091	0.8254 ± 0.0088	0.8930 ± 0.0240
√	√	√	√	√	<b>0.8955 ± 0.0263</b>	<b>0.8998 ± 0.0274</b>	<b>0.8950 ± 0.0280</b>	<b>0.9438 ± 0.0180</b>

using the elevation maps of anterior or posterior surfaces. Compared to other corneal maps in Fig. 1, where the features are mainly concentrated in the central region, AE and PE maps have more rich and diverse features, so the model has better classification results. And several studies based on corneal data have also shown that the elevation indices have a significant role in the diagnosis of KC [60,61]. Comparing the experimental results of the last three rows, we find that the corneal maps on the posterior surface play an important role for the grading, which also shows the importance of the data analysis of the corneal posterior surface in clinical research of KC [62,63]. In addition, to explore the areas of attention of the network on different corneal maps of KC eyes, we obtained the heat maps with Score-CAM separately for single corneal topographic map input, as shown in Fig. 5. The comparison results show that the attention of the network is focused on the regions with significant corneal changes, which is consistent with the physician's attention in the diagnosis, indicating the effectiveness of the deep learning method.

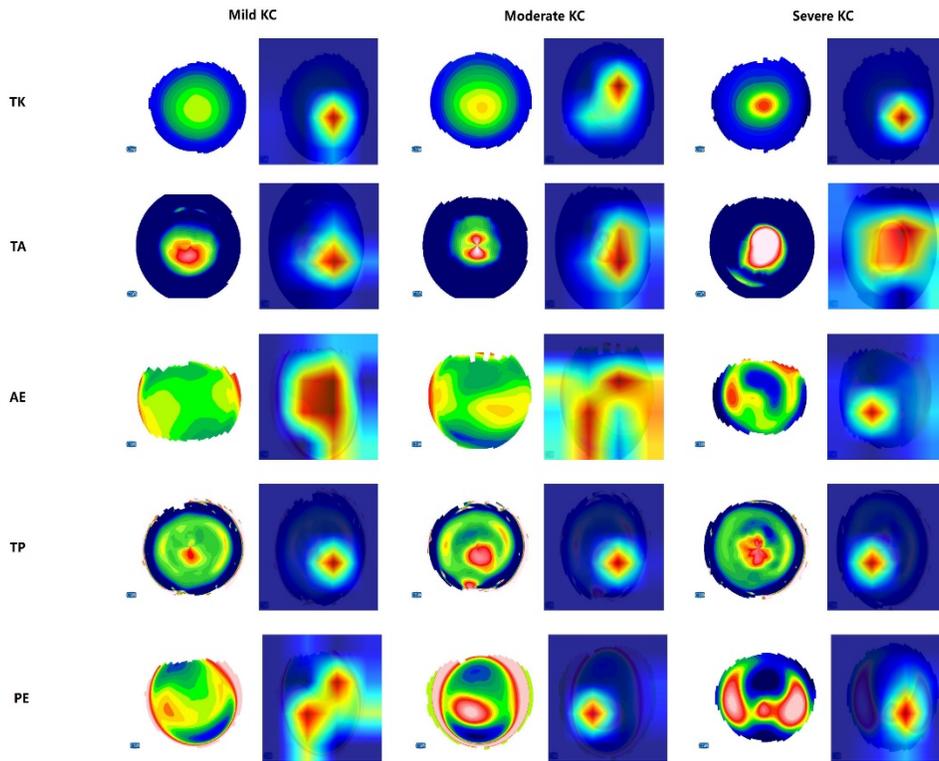
### 3.5. Ablation experiments

In order to verify the validity of the proposed self-attention based feature extraction block (SaFEB) and multi-level feature fusion module (MIFFM), we have conducted the ablation experiments as shown in Table 3, where Baseline is the ResNet18. As can be seen from Table 3, using the proposed MIFFM achieves substantial improvement over the Baseline in W\_R, W\_P and W\_F1 metrics. Meanwhile, using SaFEB to replace the basic blocks in ResNet18 also helps to improve the performance with the significant decrease in the number of model parameters. We also use the Score-CAM [64] to further demonstrate the effectiveness of the proposed method. As can be observed from Fig. 6, benefiting from the modules of SaFEB and MIFFM, our method can explicitly exploit information from the discriminative area for KC grading.

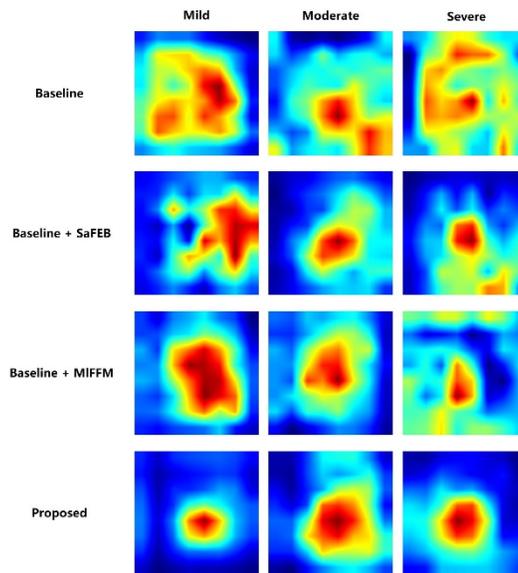
Table 3. The Results of Ablation Experiments based on LKG-Net

Methods	W_R	W_P	W_F1	Kappa	Parameter (M)
Baseline	0.8709 ± 0.0155	0.8783 ± 0.0206	0.8605 ± 0.0223	0.9271 ± 0.0193	11.22
Baseline + MIFFM	0.8873 ± 0.0316	0.8949 ± 0.0271	0.8835 ± 0.0344	0.9291 ± 0.0204	11.68
Baseline + SaFEB	0.8893 ± 0.0106	0.8894 ± 0.0081	0.8865 ± 0.0103	0.9370 ± 0.0115	<b>1.86</b>
Proposed	<b>0.8955 ± 0.0263</b>	<b>0.8998 ± 0.0274</b>	<b>0.8950 ± 0.0280</b>	<b>0.9438 ± 0.0180</b>	2.32

In the multi-level feature fusion module (MIFFM), the value of  $w$  controls the fusion ratio of the upper layer features. Results with different values for  $w$  are shown in Table 4. From the experimental results, we can see that the network has the best performance when  $w$  is 0.3. In convolutional neural networks, the features extracted from the deeper layers of the network have



**Fig. 5.** Score-CAM visualization results of different corneal map inputs. For each input, we selected three maps with different severity from the corneal maps of KC eyes.



**Fig. 6.** Score-CAM visualization results of different methods.

stronger semantic information that facilitates classification, and the features obtained from the shallower layers of the network contain more detailed information [65]. By fusing a certain proportion of low level features, we can obtain more abundant feature information, which is conducive to improving the model performance. However, an excess of the low level features will cause certain interference to the network, reducing the model performance, as shown in our experimental results.

**Table 4. Ablation Study for the Influence of  $W$  in MIFFM**

LKG-Net	W_R	W_P	W_F1	Kappa
$w = 0.1$	$0.8750 \pm 0.0309$	$0.8718 \pm 0.0336$	$0.8710 \pm 0.0298$	$0.9324 \pm 0.0236$
$w = 0.3$	<b><math>0.8955 \pm 0.0263</math></b>	<b><math>0.8998 \pm 0.0274</math></b>	<b><math>0.8950 \pm 0.0280</math></b>	<b><math>0.9438 \pm 0.0180</math></b>
$w = 0.5$	$0.8811 \pm 0.0303$	$0.8829 \pm 0.0261$	$0.8789 \pm 0.0298$	$0.9364 \pm 0.0199$
$w = 0.7$	$0.8853 \pm 0.0334$	$0.8868 \pm 0.0333$	$0.8845 \pm 0.0346$	$0.9293 \pm 0.0404$
$w = 0.9$	$0.8791 \pm 0.0375$	$0.8901 \pm 0.0362$	$0.8761 \pm 0.0403$	$0.9335 \pm 0.0164$

### 3.6. Experiments for KC screening

To verify the effectiveness of this method in KC screening, we divided the data into a new dataset containing 252 KC and 236 normal eyes. We also fed different combinations of inputs into the network for the experiments, and the results are shown in Table 5. As can be seen from Table 5, our method still achieves good results. Similar to the KC classification results in Table 2, when inputting a single map, the corneal maps of anterior and posterior elevation all have excellent performance in KC screening, and the corneal posterior data is also particularly important for KC detection, which is also consistent with some medical research results on KC [61,62]. Moreover, comparing the experimental results of the last three rows, it can be found that our method achieves high accuracy in detection of KC. Both the anterior and the posterior maps can distinguish KC well, and all five maps only slightly improved the model performance.

**Table 5. KC Screening Results of Different Input Combinations of Corneal Topography**

TK	LKG-Net Input				Accuracy	Precision	Recall	F1_score	AUC
	TA	AE	TP	PE					
✓					$0.9037 \pm 0.0194$	$0.8858 \pm 0.0111$	$0.8889 \pm 0.0130$	$0.9053 \pm 0.0171$	$0.9042 \pm 0.0201$
	✓				$0.8955 \pm 0.0155$	$0.8551 \pm 0.0227$	$0.8492 \pm 0.0304$	$0.8935 \pm 0.0162$	$0.8971 \pm 0.0156$
		✓			$0.9529 \pm 0.0123$	$0.9210 \pm 0.0126$	$0.9206 \pm 0.0129$	$0.9528 \pm 0.0123$	$0.9540 \pm 0.0124$
			✓		$0.9119 \pm 0.0205$	$0.8649 \pm 0.0284$	$0.8572 \pm 0.0343$	$0.9093 \pm 0.0213$	$0.9137 \pm 0.0203$
				✓	$0.9467 \pm 0.0195$	$0.9241 \pm 0.0310$	$0.9246 \pm 0.0327$	$0.9469 \pm 0.0201$	$0.9475 \pm 0.0191$
	✓	✓			$0.9631 \pm 0.0082$	$0.9434 \pm 0.0152$	$0.9445 \pm 0.0159$	$0.9636 \pm 0.0080$	$0.9638 \pm 0.0083$
			✓	✓	$0.9693 \pm 0.0140$	$0.9512 \pm 0.0182$	$0.9524 \pm 0.0184$	$0.9697 \pm 0.0138$	$0.9699 \pm 0.0139$
✓	✓	✓	✓	✓	$0.9734 \pm 0.0079$	$0.9669 \pm 0.0126$	$0.9683 \pm 0.0129$	$0.9742 \pm 0.0073$	$0.9735 \pm 0.0083$

## 4. Conclusion and discussions

In this paper, we propose an end-to-end deep learning network named LKG-Net for KC severity grading. The proposed self-attention based feature extraction block (SaFEB) can not only extract more abundant and important information, but also significantly reduce the complexity of the model. The multi-level feature fusion module (MIFFM) is designed to adaptively calibrate the features of the last two levels channel dimensions based on the channel attention mechanism, and the calibrated two-level features are fused to enhance the features. The experimental results show that our network can accurately grade the different severity of the disease, and also has excellent

performance in the screening of KC. We use the Score-CAM to visualize the region of interest of the proposed network, and the results demonstrate that it focuses on the region similar to the one relied on by ophthalmologists.

Compared with other KC detection and grading methods based on machine learning [22,36–44], our method has distinct advantages. First of all, our method is a truly automated end-to-end KC diagnosis method. From report input to result output, the whole process is automatic, and doctors don't need to analyze complex parameter indicators, which will effectively relieve their work pressure. Secondly, compared with other KC classification methods, our method has better innovations in feature extraction and fusion, which can obviously improve the performance while significantly reducing the amount of model parameters, and it will be conducive to the application of our method in ophthalmic detection devices.

The current gold standard for KC grading relies on human interpretation of biomicroscopic features and tomography scans, and has previously been shown to be limited by poor reproducibility [66]. Although it is easy to illustrate the corneal shape using digital indices, they can't represent the spatial gradient and distribution of corneal curvature, elevation, diopter and thickness. With the advent of advanced corneal imaging devices, multiple objective metrics obtained have been proposed for the diagnosis and severity assessment of KC. Nevertheless, these highly complex and numerous parameters may pose a clinical challenge to the ophthalmologist. Moreover, the metrics obtained by different devices may be based on different reference standards, so the algorithms trained based on metrics from one device are not universally applicable. To avoid analysis and processing of complex parameters, we use color-coded maps to provide the CNNs with more complete information about the overall state of the cornea. The heat maps in Fig. 5 and Fig. 6 reveal that our network can locate the key regions of the corneal maps. With the features in these regions, it is easier to distinguish KC and its various stages. In addition, more and more studies have shown that data from the posterior surface of the cornea are also important references in the diagnosis of KC [60–63], which is also demonstrated by the results of our experiments in Table 2 and Table 5. Such consistency between our results and previous studies further illustrates the feasibility of using CNNs for KC detection.

Although the proposed method has achieved good performance on the test data, there are still some limitations. First, all comparison and evaluation experiments are based on limited data acquired from one device, and the sample size in each category is unbalanced. Therefore, we should collect more data from multiple imaging devices to further validate the robustness of the method, and improve the loss function to cope with the category imbalance. Second, some key diagnostic metrics can be used to aid network classification, and we consider to implement it in the subsequent work. In the future, we will collect more data and use the improved method for the screening of early-stage KC, so as to better assist ophthalmologists in the diagnosis and treatment of the disease.

**Funding.** National Natural Science Foundation of China (U20A20170, 62271337); National Key Research and Development Program of China (2018YFA0701700).

**Disclosures.** The authors declare that there are no conflicts of interest related to this article.

**Data availability.** Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

## References

1. Y. S. Rabinowitz, "Keratoconus," *Surv. Ophthalmol.* **42**(4), 297–319 (1998).
2. J. Nottingham, "Practical observations on conical cornea, and on the short sight, and other defects of vision connected with it," in *Mr. Churchill's publications in medicine* <https://doi.org/10.5962/bhl.title.160281> (1854).
3. E. Arnal, C. Peris-Martínez, J. L. Menezo, S. Johnsen-Soriano, and F. J. Romero, "Oxidative stress in keratoconus?" *Invest. Ophthalmol. Visual Sci.* **52**(12), 8592–8597 (2011).
4. J. H. Krachmer, R. S. Feder, and M. W. Belin, "Keratoconus and related noninflammatory corneal thinning disorders," *Surv. Ophthalmol.* **28**(4), 293–322 (1984).

5. R. H. Kennedy, W. M. Bourne, and J. A. Dyer, "A 48-year clinical and epidemiologic study of keratoconus," *Am. J. Ophthalmol.* **101**(3), 267–273 (1986).
6. A. Pearson, B. Soneji, N. Sarvananthan, and J. Sandford-Smith, "Does ethnic origin influence the incidence or severity of keratoconus?" *Eye* **14**(4), 625–628 (2000).
7. M. Edwards, C. N. McGhee, and S. Dean, "The genetics of keratoconus," *Clin. Exp. Ophthalmol.* **29**(6), 345–351 (2001).
8. Y. S. Rabinowitz, "The genetics of keratoconus," *Ophthalmol. Clin. North America* **16**(4), 607–620 (2003).
9. Y. Bykhovskaya, B. Margines, and Y. S. Rabinowitz, "Genetics in Keratoconus: where are we?" *Eye and Vis.* **3**(1), 16 (2016).
10. W. E. Smiddy, T. R. Hamburg, G. P. Kracher, and W. J. Stark, "Keratoconus: contact lens or keratoplasty?" *Ophthalmology* **95**(4), 487–492 (1988).
11. V. Jhanji, N. Sharma, and R. B. Vajpayee, "Management of keratoconus: current scenario," *Br. J. Ophthalmol.* **95**(8), 1044–1050 (2011).
12. V. M. Rathi, P. S. Mandathara, and S. Dumpati, "Contact lens in keratoconus," *Indian J. Ophthalmol.* **61**(8), 410–415 (2013).
13. J. Colin, B. Cochener, G. Savary, and F. Malet, "Correcting keratoconus with intracorneal rings," *J. Cataract Refractive Surg.* **26**(8), 1117–1122 (2000).
14. E. Coskunseven, G. D. Kymionis, N. S. Tsiklis, S. Atun, E. Arslan, M. R. Jankov, and I. G. Pallikaris, "One-year results of intrastromal corneal ring segment implantation (KeraRing) using femtosecond laser in patients with keratoconus," *Am. J. Ophthalmol.* **145**(5), 775–779.e1 (2008).
15. G. Wollensak, "Crosslinking treatment of progressive keratoconus: new hope," *Current Opinion Ophthalmol.* **17**(4), 356–360 (2006).
16. F. Raiskup-Wolf, A. Hoyer, E. Spoerl, and L. E. Pillunat, "Collagen crosslinking with riboflavin and ultraviolet-A light in keratoconus: long-term results," *J. Cataract Refractive Surg.* **34**(5), 796–801 (2008).
17. I. Bahar, S. Levinger, and I. Kremer, "Wavefront-supported photorefractive keratectomy with the Bausch & Lomb Zyoptix in patients with myopic astigmatism and suspected keratoconus," *J. Refract. Surg.* **22**, 533–538 (2006).
18. M. D. Wagoner, S. D. Smith, W. J. Rademaker, and M. A. Mahmood, "Penetrating keratoplasty vs. epikeratoplasty for the surgical treatment of keratoconus," *J. Refract. Surg.* **17**, 138–146 (2001).
19. J. Colin and S. Velou, "Implantation of Intacs and a refractive intraocular lens to correct keratoconus," *J. Cataract Refractive Surg.* **29**(4), 832–834 (2003).
20. T. M. El-Raggal and A. A. A. Fattah, "Sequential Intacs and Verisyse phakic intraocular lens for refractive improvement in keratoconic eyes," *J. Cataract Refractive Surg.* **33**(6), 966–970 (2007).
21. M. Romero-Jiménez, J. Santodomingo-Rubido, and J. S. Wolffsohn, "Keratoconus: a review," *Contact Lens Anterior Eye* **33**(4), 157–166 (2010).
22. M. C. Arbelaez, F. Versaci, G. Vestri, P. Barboni, and G. Savini, "Use of a support vector machine for keratoconus and subclinical keratoconus detection by topographic and tomographic data," *Ophthalmology* **119**(11), 2231–2238 (2012).
23. M. Amsler, "Kératocône classique et kératocône fruste; arguments unitaires," *Ophthalmologica* **111**(2-3), 96–101 (1946).
24. J. L. Alió and M. H. Shabayek, "Corneal higher order aberrations: a method to grade keratoconus," *J. Refract. Surg.* **22**, 539–545 (2006).
25. T. T. McMahon, L. Szczołka-Flynn, J. T. Barr, R. J. Anderson, M. E. Slaughter, J. H. Lass, S. K. Iyengar, and C. S. Group, "A new method for grading the severity of keratoconus: the Keratoconus Severity Score (KSS)," *Cornea* **25**(7), 794–800 (2006).
26. J. Duncan and J. A. Gomes, "A new tomographic method of staging/classifying keratoconus: the ABCD grading system," *International J. Keratoconus and Ectatic Corneal Diseases* **4**(3), 85–93 (2015).
27. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Commun. ACM* **60**(6), 84–90 (2017).
28. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv*, preprint arXiv:1409.1556 (2014).
29. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition* (2015), pp. 1–9.
30. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition* (2016), pp. 770–778.
31. G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition* (2017), pp. 4700–4708.
32. I. P. Okuwobi, Z. Ji, W. Fan, S. Yuan, L. Bekalo, and Q. Chen, "Automated quantification of hyperreflective foci in SD-OCT with diabetic retinopathy," *IEEE J. Biomed. Health Inform.* **24**(4), 1125–1136 (2020).
33. M. T. Islam, S. A. Imran, A. Arefeen, M. Hasan, and C. Shahnaz, "Source and camera independent ophthalmic disease recognition from fundus image using neural network," in *2019 IEEE International Conference on Signal Processing, Information, Communication & Systems (SPICSCON)* (IEEE, 2019), pp. 59–63.

34. D. E. Worrall, C. M. Wilson, and G. J. Brostow, "Automated retinopathy of prematurity case detection with convolutional neural networks," in *Deep Learning and Data Labeling for Medical Applications* (Springer, 2016), pp. 68–76.
35. D. S. J. Ting, V. H. Foo, L. W. Y. Yang, J. T. Sia, M. Ang, H. Lin, J. Chodosh, J. S. Mehta, and D. S. W. Ting, "Artificial intelligence for anterior segment diseases: Emerging applications in ophthalmology," *Br. J. Ophthalmol.* **105**(2), 158–168 (2021).
36. N. Maeda, S. D. Klyce, and M. K. Smolek, "Neural network classification of corneal topography. Preliminary demonstration," *Invest. Ophthalmol. Visual Sci.* **36**(7), 1327–1335 (1995).
37. P. A. Accardo and S. Pensiero, "Neural network-based system for early keratoconus detection from corneal topography," *J. Biomed. Inf.* **35**(3), 151–159 (2002).
38. M. B. Souza, F. W. Medeiros, D. B. Souza, R. Garcia, and M. R. Alves, "Evaluation of machine learning classifiers in keratoconus detection from orbscan II examinations," *Clinics* **65**(12), 1223–1228 (2010).
39. D. Smadja, D. Touboul, A. Cohen, E. Doveh, M. R. Santhiago, G. R. Mello, R. R. Krueger, and J. Colin, "Detection of subclinical keratoconus using an automated decision tree classification," *Am. J. Ophthalmol.* **156**(2), 237–246.e1 (2013).
40. A. Lavric and P. Valentin, "KeratoDetect: keratoconus detection algorithm using convolutional neural networks," *Computational Intelligence and Neuroscience* **2019**, 8162567 (2019).
41. P. Zéboulon, G. Debellemanière, M. Bouvet, and D. Gatinel, "Corneal topography raw data classification using a convolutional neural network," *Am. J. Ophthalmol.* **219**, 33–39 (2020).
42. B.-I. Kuo, W.-Y. Chang, T.-S. Liao, F.-Y. Liu, H.-Y. Liu, H.-S. Chu, W.-L. Chen, F.-R. Hu, J.-Y. Yen, and I.-J. Wang, "Keratoconus screening based on deep learning approach of corneal topography," *Trans. Vis. Sci. Tech.* **9**(2), 53 (2020).
43. R. Feng, Z. Xu, X. Zheng, H. Hu, X. Jin, D. Z. Chen, K. Yao, and J. Wu, "KerNet: a novel deep learning approach for keratoconus and sub-clinical keratoconus detection based on raw data of the Pentacam HR system," *IEEE J. Biomed. Health Inform.* **25**(10), 3898–3910 (2021).
44. K. Kamiya, Y. Ayatsuka, Y. Kato, F. Fujimura, M. Takahashi, N. Shoji, Y. Mori, and K. Miyata, "Keratoconus detection using deep learning of colour-coded maps with anterior segment optical coherence tomography: a diagnostic accuracy study," *BMJ open* **9**(9), e031313 (2019).
45. A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv*, preprint arXiv:1704.04861 (2017).
46. M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2018), pp. 4510–4520.
47. A. Howard, M. Sandler, G. Chu, L.-C. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, and V. Vasudevan, "Searching for mobilenetv3," in *Proceedings of the IEEE/CVF International Conference on Computer Vision* (2019), pp. 1314–1324.
48. V. Mnih, N. Heess, and A. Graves, "Recurrent models of visual attention," *Advances in Neural Information Processing Systems* 27 (NIPS 2014) 27 (2014).
49. J. Fu, J. Liu, H. Tian, Y. Li, Y. Bao, Z. Fang, and H. Lu, "Dual attention network for scene segmentation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision* (2019), pp. 3146–3154.
50. J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2018), pp. 7132–7141.
51. S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International Conference on Machine Learning* (PMLR, 2015), pp. 448–456.
52. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in Neural Information Processing Systems* 30 (NIPS 2017) (2017).
53. X. Pan, C. Ge, R. Lu, S. Song, G. Chen, Z. Huang, and G. Huang, "On the integration of self-attention and convolution," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2022), pp. 815–825.
54. K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *European Conference on Computer Vision* (Springer, 2016), pp. 630–645.
55. T. Xiao, Y. Xu, K. Yang, J. Zhang, Y. Peng, and Z. Zhang, "The application of two-level attention models in deep convolutional neural network for fine-grained image classification," in *Proceedings of the IEEE conference on computer vision and pattern recognition* (2015), pp. 842–850.
56. S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)* (2018), pp. 3–19.
57. P. Shaw, J. Uszkoreit, and A. Vaswani, "Self-attention with relative position representations," *arXiv*, preprint arXiv:1803.02155 (2018).
58. M. L. McHugh, "Interrater reliability: the kappa statistic," *Biochemia. Medica.* **22**(3), 276–282 (2012).
59. S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition* (2017), pp. 1492–1500.
60. K. Baker, Y.-L. Chen, L. Shi, J. Lewis, L. Kugler, and M. Wang, "Does the Posterior Corneal Elevation Provide the First Indication of Keratoconus?" *Invest. Ophthalmol. Visual Sci.* **51**(13), 4963 (2010).

61. Z. Schlegel, T. Hoang-Xuan, and D. Gatinel, "Comparison of and correlation between anterior and posterior corneal elevation maps in normal eyes and keratoconus-suspect eyes," *J. Cataract Refractive Surg.* **34**(5), 789–795 (2008).
62. ÖÖ Uçakhan, V. Çetinkor, M. Özkan, and A. Kanpolat, "Evaluation of Scheimpflug imaging parameters in subclinical keratoconus, keratoconus, and normal eyes," *J. Cataract Refractive Surg.* **37**(6), 1116–1124 (2011).
63. A. Saad and D. Gatinel, "Topographic and tomographic properties of forme fruste keratoconus corneas," *Invest. Ophthalmol. Visual Sci.* **51**(11), 5546–5555 (2010).
64. H. Wang, Z. Wang, M. Du, F. Yang, Z. Zhang, S. Ding, P. Mardziel, and X. Hu, "Score-CAM: Score-weighted visual explanations for convolutional neural networks," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops* (2020), pp. 24–25.
65. L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M. A. Fadhel, M. Al-Amidie, and L. Farhan, "Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions," *J. Big Data* **8**(1), 53–74 (2021).
66. X. Chen, J. Zhao, K. C. Iselin, D. Borroni, D. Romano, A. Gokul, C. N. McGhee, Y. Zhao, M.-R. Sedaghat, and H. Momeni-Moghaddam, "Keratoconus detection of changes using deep learning of colour-coded maps," *BMJ Open Ophthalmol.* **6**(1), e000824 (2021).