IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received January 7, 2021, accepted January 26, 2021, date of publication February 1, 2021, date of current version February 9, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3056234

A Novel Meibomian Gland Morphology Analytic System Based on a Convolutional Neural Network

QI DAI[1,2], XINYI LIU[2], XIAOLEI LIN[3], YANA FU[2], CHAOQIAO CHEN[2], XINXIN YU[2], ZUHUI ZHANG[2], TIANKUN LI[2], MENGTING LIU[2], WEIHUA YANG[4], AND JUAN YE[1]

[1]Department of Ophthalmology, School of Medicine, Second Affiliated Hospital, Zhejiang University, Hangzhou 310009, China
[2]School of Ophthalmology and Optometry, Eye Hospital, Wenzhou Medical University, Wenzhou 325027, China
[3]Department of Ophthalmology and Visual Science, Eye, Ear, Nose, and Throat Hospital, Shanghai Medical College, Fudan University, Shanghai 200031, China
[4]Affiliated Eye Hospital, Nanjing Medical University, Nanjing 210029, China

Corresponding authors: Juan Ye (yejuan@zju.edu.cn) and Weihua Yang (benben0606@139.com)

This work was supported in part by the Projects of Medical and Health Technology Development Program in Zhejiang Province under Grant 2019PY049, in part by the Joint Funds of the Zhejiang Provincial Natural Science Foundation of China under Grant LSY19H120001, and in part by the Wenzhou Basic Scientific Research Project under Grant Y20180712.

ABSTRACT Meibomian glands dysfunction (MGD) is the main cause of dry eyes. Biological parameters of meibomian gland (MG) such as height, tortuosity and the degree of atrophy are closely related to its function. However, Thus, an effective quantitative diagnostic tool is needed for clinical diagnosis. Automatic quantification of MGs' morphological features could be a challenging task and play an important role in MGD diagnosis and classification. Our main objective is to develop an artificial intelligence (AI) system for evaluating MGs' morphology and explore the relationship between the morphological parameters and functions. We proposed a novel MGs extraction method based on convolutional neural network (CNN) with enhanced mini U-Net. A prospective study was conducted, 120 subjects were included and taken meibography. The training and validation sets encompassed 60 subjects; and the test set consisted of other 60 subjects with comprehensive examinations for ocular surface disease index questionnaire (OSDI), tear meniscus height (TMH), tear break-up time (TBUT), corneal fluorescein staining (CFS), lid margin score, and meibum expressibility score. The algorithm effectively extracted MGs from meibography even with this small training sample. As a result, while the intersection over union (IoU) achieved 0.9077, the repeatability was 100%. The processing time for each image was 100ms. Using this method, the investigators identified a significant and linear correlation between MG morphology and clinical parameters. This study provided a new method for quantification of MGs' morphological features obtained by meibography, which has advantages in reducing analysis time, improving diagnostic efficiency, and assisting ophthalmologists with limited clinical expertise.

INDEX TERMS Deep learning, convolutional neural network (CNN), meibomian gland dysfunction (MGD), meibography.

I. INTRODUCTION

Meibomian glands (MGs) are important ocular adnexa and the largest sebaceous glands in the human body. MGs are important in maintaining the health and stability of the ocular surface [1]. Meibomian glands dysfunction (MGD) is the main cause of dry eyes. Therefore, it is essential to evaluate the function of MG in dry eye patients. An existing study confirmed that most of patients with dry eye disease reported experiencing meibomian glands dysfunction (MGD) [2]. MG morphology is closely related to the severity of MGD. Several studies have suggested that the meibomian gland image index can indicate the health of MGs [3], [4]. The main parameter of the index is the degree of atrophy [5]. For instance, Pult *et al.* found that the width and bending of the upper eyelid's MGs are also related to noninvasive tear break-up time (NIBUT).

One problem with previously published methods is that most studies adopted a semiquantitative method such as ImageJ and relied on manually marking the meibomian glands. However, since the ImageJ software used by researchers incorporated semiquantitative means, the software could only measure the MG bending degree at a

The associate editor coordinating the review of this manuscript and approving it for publication was Dezhong Peng.

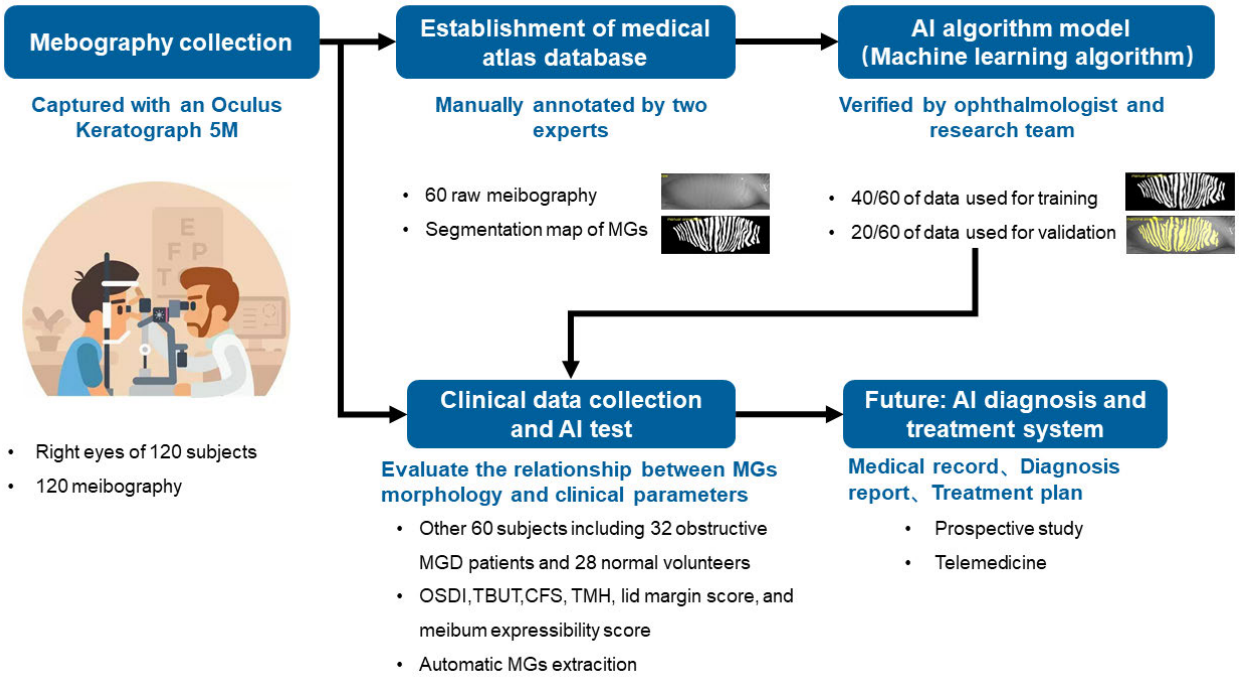


FIGURE 1. The flow chart of this study.

single point instead of measuring the overall gland distortion. Additionally, the semiquantitative assessment of meibomian glands remains subjective, and the efficiency is limited by manual calibration.

In subsequent research, image analysis systems have become popular research tools in analyzing ophthalmic images. Some methods have shown superiority in the analysis of MG morphology. For example, the traditional image algorithm (with image filtering, binarization, skeleton extraction, etc.) adopted by Arita *et al.* was able to analyze the morphology of meibomian glands and obtain relatively accurate results [6]. However, the algorithm had limitations, especially when used to assess single MG morphology such as MG tortuosity, height, and width [7]. Traditional image recognition methods based on image processing typically use algorithms to extract the features in the image, and such methods are typically manually set based on the subjective judgment of researchers.

Previous studies in artificial intelligence, especially deep learning, were proved effective in automatic evaluation of MG atrophy [8]–[10]. But these researches focused on MG atrophy ratio and classification, and did not extract comprehensive morphology parameters of each MG. The importance of morphological parameters, as well as the difficulty and time-consuming of the work, urgently requires an auxiliary approach to assist the clinical practice.

We have selected a convolutional neural network (CNN), a feedforward neural network, to manipulate MG images. CNNs have been incorporated with artificial intelligence (AI) to automatically learn the hidden features in images by integrating feature learning into the model building process, which can reduce the inaccuracy of a manual input design. Therefore, CNNs have been applied to the complex image analysis [11], [12] and have been proven to possess the ability to replace manual calibration. In fact, CNNs have been used in the study of ophthalmic diseases such as diabetic retinopathy [13] and glaucomatous optic neuropathy [14]. Additionally, in this research, we use U-Net, which is an efficient network and training strategy. The U-Net approach consists of two parts: a contracting path that can capture context and a symmetric expanding path that can precisely locate images. The architecture enables end-to-end training from very few images and outperforms a sliding-window convolutional network on the International Symposium on Biomedical Imaging (ISBI) challenge for the segmentation of neuronal structures in electron microscopic stacks [15].

This study developed an AI system based on a CNN for extracting every single MG and calculating morphology index accurately. Then, morphology of the total MGs was assessed. This AI system can quickly quantify the severity of MG atrophy and explore the relationship between MG morphology and MG functions, as shown in Fig 1.

II. METHODS

A. SUBJECT RECRUITMENT

A total of 120 subjects (from 8 to 51 years old) were recruited for our study from the affiliated eye hospital of Wenzhou Medical University. Only the right eyes of the

enrolled subjects were included. The exclusion criteria were as follows: (1) subjects with ocular diseases or conditions known to affect the anatomy of the anterior segment, such as ocular inflammation, a history of ocular surgery, contact lens wear, and/or trauma; (2) subjects with a history of systemic medications that affect MG functions; and (3) subjects with any other ocular or systemic diseases known to affect the tear film.

Among the 120 subjects, 60 subjects were recruited for AI system training and validation. After the training was finished, 32 subjects (16 males and 16 females, with a mean age of 33.28 ± 9.28 years) with obstructive MGD and 28 normal volunteers (11 males and 17 females, with a mean age of 25.25 ± 11.19 years) were included for AI automatic analysis test. All subjects were diagnosed with obstructive MGD by two experienced ophthalmologists based on ocular symptoms, lid margin abnormalities, and meiboscores. Subjects were diagnosed with obstructive MGD if any two of the three following criteria were met: I) ocular symptom score ≥ 3; II) lid margin score ≥ 2; and/or III) meiboscore ≥ 3 [16]. Only right eye measurements of the enrolled subjects were included. The study was conducted in accordance with the *Declaration of Helsinki* and was approved by the Research Ethics Committee of the Eye Hospital, Wenzhou Medical University. Informed consent to publish was obtained from all participants before their inclusion in the study.

B. AI TRAINING AND TESTING

1) DATA COLLECTION

Meibography images of the upper and lower eyelids for only the right eye were captured with an Oculus Keratograph 5M [17], and were further divided into 40 training images and 20 validation images in the present study. The MGs in the training images were manually annotated by two experts (Dr. Qi Dai and Dr. Yana Fu) with VIA 3 software [18]. Before training, the research team preprocessed the 40 training images with the following transformations:

2) IMAGE OPTIMIZATION

2.1) The image was converted to grayscale with the formation below:

$$I_{gray} = I_r \times 0.299 + I_g \times 0.587 + I_b \times 0.114$$

I_r is the value of the R channel of the RGB image, I_g the G channel of the RGB image, and I_b the B channel of the RGB image,

2.2) Standardization: The image was then standardized by the following formula:

$$X_{std} = \frac{X_{gray} - \mu}{\sigma}$$

where μ is the pixel average and σ is the pixel standard deviation.

2.3) Normalization: The image was then normalized by the following formula:

$$X_{norm} = \frac{X_{std} - \min(X_{std})}{\max(X_{std}) - \min(X_{std})}$$

2.4) The histogram was adjusted by Contrast Limited Adaptive Histogram Equalization (CLAHE): The MG images were inhomogeneous; some suffered low contrast, while others were overexposed. Therefore, the CLAHE algorithm was used to improve the global contrast and prevent over amplifying noise in relatively homogeneous image regions.

2.5) Gamma adjustment

The brightness of each pixel was adjusted by the following formula.

$$Y = \sqrt[0.8]{x}$$

x (range from [0, 1]) is the original pixel value in the image, and Y is the adjusted pixel value.

3) PROCESSING OF THE SAMPLE SIZE

Since the training and validation set had only 60 images and would be too small to train a deep convolutional neural network, we need to use some data augmentation methods to expand the training set. A common method is to randomly select sub-images from each training image. We obtained 10,000 patches with 64 pixels ×64 pixels from each preprocessed training image by randomly selecting the center of the training image. Then, we used 80% of the total patches for training and the remaining 20% for validation. Fig 2(a) presents the 40 images obtained from the patches, and Fig 2(b) shows the manual annotation corresponding to the sub-images.

This data augmentation method is simple to implement, but it has not achieved satisfactory results. The data augmentation method divides the original picture into multiple sub-images. In the process of dividing, the positional relationship between the sub-images is lost, as shown in Fig 3.

In order to solve this problem, we introduced a powerful data augmentation library (https://github.com/aleju/imgaug#citation) to augment training images in multiple dimensions. These augmentation methods include rotation, scaling, adding noise, contrast adjustment, distortion, blurring, sharpening, cropping, fliplr, affine etc. During the training process, the original image will randomly select the augmentation method, and perform augmentation operations with random parameters. Fig 4 shows some images generated after doing 11 randomly image augmentation. The one in the upper left corner is the original image.

4) ALGORITHM DESIGN AND TRAINING

The CNN structure was derived from the original U-Net architecture. This study simplified the U-Net architecture by decreasing the filter number of each convolutional layer and adding only two max pooling layers and two up-sampling layers, which contrasted with the traditional four layers in a conventional U-Net architecture. The numbers of filters for the

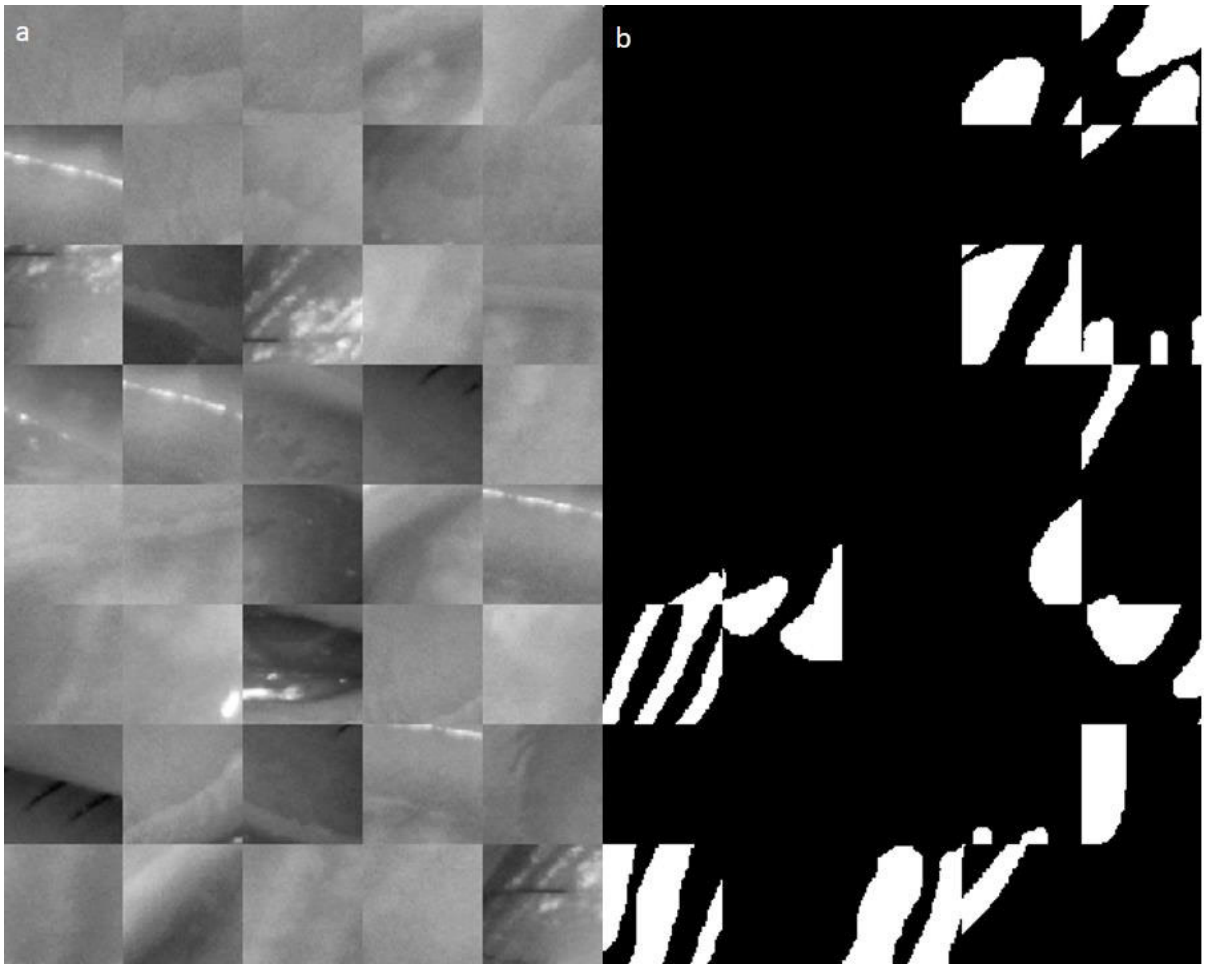


FIGURE 2. Sample of sub-images: 40 patches with 64 pixels ×64 pixels randomly selected from each preprocessed training image (a). Manual annotation of the sub-images corresponding to the same sub-images (b).

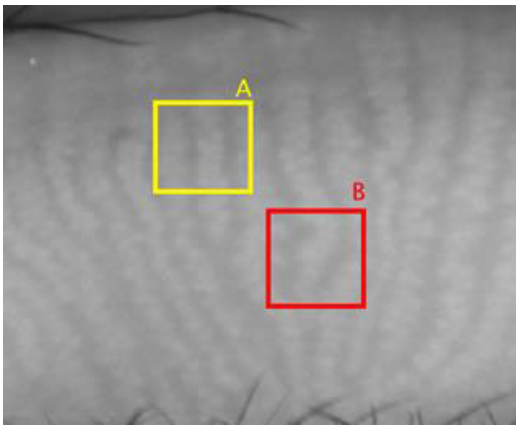


FIGURE 3. When the sub-images A and B are split from the original picture, we will never know A is in the upper left corner of B.

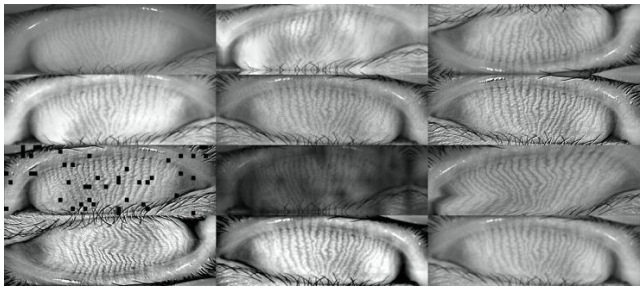


FIGURE 4. The original image (the upper left corner one) and 11 randomly image augmentation.

convolutional layers were 32, 64, 128, 64, and 32 [16]. In the following we refer to this network as mini U-Net. Cross-entropy was used for the loss function, and Adam was used for learning rate optimization. The rectifier linear unit was used as the activation function for each convolutional layer, and a dropout rate of 0.2 was used between two consecutive convolution layers to prevent the model from overfitting.

The model was implemented by Keras [19] and TensorFlow [20]. We divided the 60 sample images into 40 training images and 20 validating images. The training of the mini U-Net lasted for approximately 15 hours after the investigators performed 120 epochs on a GTX 1070 8G GPU with a mini batch of 4 patches. Finally, an Intersection of Union (IoU) of 0.895 was obtained in the validation set, indicating that the model had been trained to extract MGs from images.

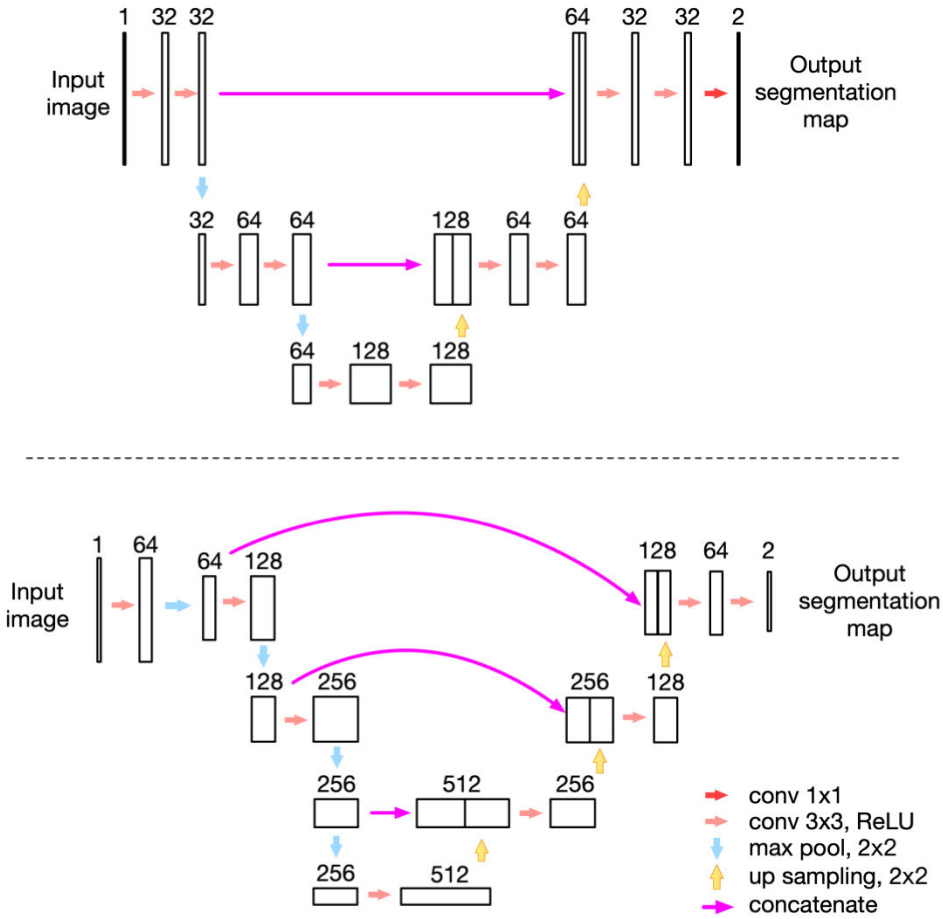


FIGURE 5. Network structure of the mini U-Net (image above) and enhanced mini U-Net CNN model (image blow).

In order to improve the accuracy of the mini U-Net, we have enhanced mini U-Net by adding two extra max pooling layers and two up sampling layers, the numbers of filters for the extra two convolutional layers were 256 and 512.When we retrained the enhanced model on the same machine, we got better results of 0.9077 on the validation Set. The network structure is shown in Fig 5.

During the calculation of the model, all input pictures will be scaled to a size of 288 pixels × 896 pixels and input into the model, the final predicted pictures will be restored to the size of the original pictures. Therefore, if the size of the original picture is much larger than 288 pixels × 896 pixels, the glands in the predicted result picture may be jagged. We need to perform a Gaussian filter on the result image to obtain smoother glands. The filtered result predicted by the enhanced mini U-Net of the image in Fig 6(a) is shown in Fig 6(c). After obtaining the prediction result of the picture, the researchers then extracted each MG in the prediction picture through traditional image processing methods provided by OpenCV. The extraction steps are delineated as follows:

4.1) Apply median filtering on the image;

4.2) Binarize the image;

4.3) Find the contour of the MG in the image. If the vertical distance between the two contours is small, this observation can be used to indicate that the contours belong to the same MG and are connected.

The extracted MGs are shown in green in Fig 6(d). And Fig 7 is the comparison samples of raw, manual annotation and machine annotation MGs

C. AI SYSTEM AND ITS APPLICATION ON MEIBOGRAPHY ANALYSIS CORRELATION ANALYSIS

1) CLINICAL PARAMETERS

Clinical assessments were performed sequentially as follows: symptom questionnaires, tear meniscus height (TMH), tear break-up time (TBUT), corneal fluorescein staining (CFS), lid margin abnormality, MG expressibility, and meibography. All of the subjects completed the Ocular Surface Disease Index (OSDI) questionnaire, and they were asked whether they had any of the 14 MGD-related ocular symptoms (symptom score) [17]. The Keratograph 5M was used to measure TMH and to perform the meibography scans. TMH was measured 5 seconds after blinking, and we measured the central TMH of the lower eyelid. TBUT was measured

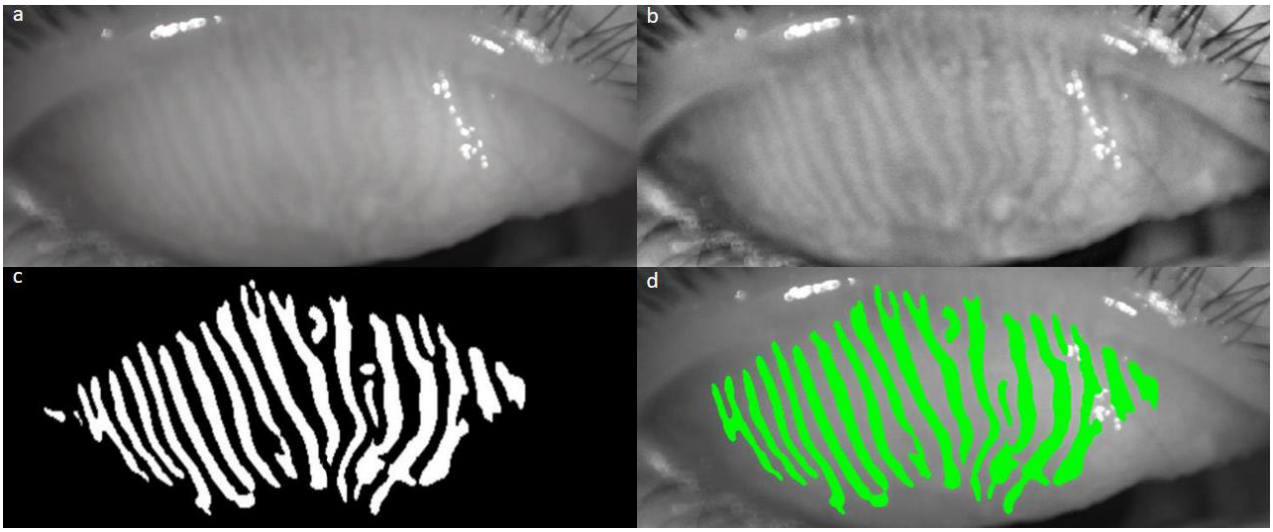


FIGURE 6. The raw image (a) and the preprocessed image after optimization (b). The prediction result of same image(c). The extracted MGs in mebograph are shown in green (d).

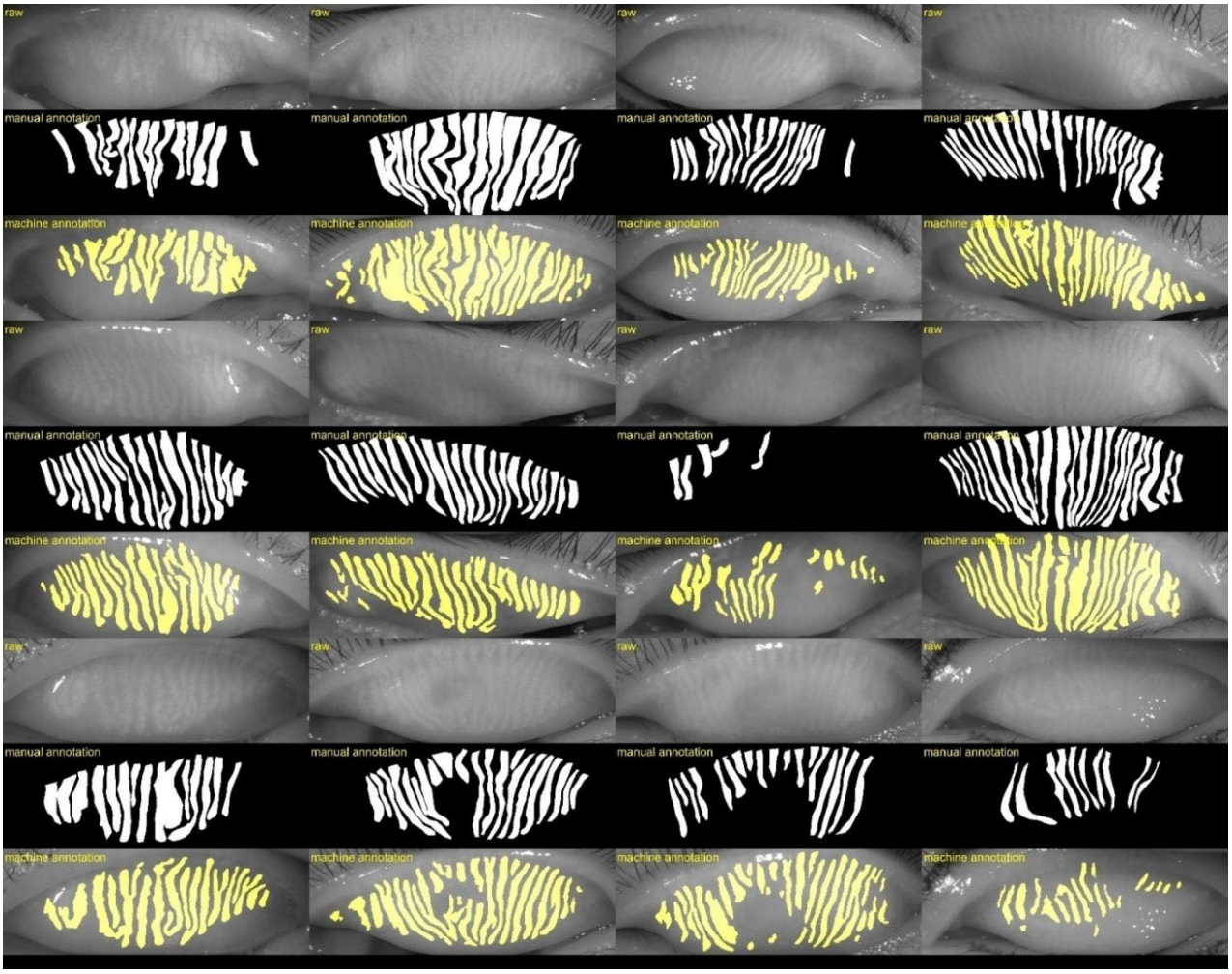


FIGURE 7. Comparison of raw, manual annotation and machine annotation MGs.

and CFS was performed after the instillation of fluorescein. TBUT was measured three times, and the mean value was recorded. CFS was graded using the Baylor grading scheme from0 to 4[21]. Lid margin abnormalities were scored from

0 through 4 according to the following four parameters: anterior or posterior displacement of the mucocutaneous junction, vascular engorgement, plugged meibomian gland orifices, and irregularity of the lid margin. We assessed the meibum quality and quantity of the 15 glands on each lower eyelid. The MG expressibility score ranged from 0 to 45[22]. Images of both the upper and lower MGs were captured by the K5M. To assess the degree of MG atrophy, we used the method described by Arita *et al.* [23] to calculate the meiboscore: 0, no atrophy; 1, atrophy of $< 1/3$ of the total lid area; 2, atrophy of 1/3 to 2/3 of the total lid area; and 3, atrophy of $> 2/3$ of the total lid area. The total meiboscore of the upper and lower eyelids results ranged from 0 to 6.

2) MORPHOLOGY PARAMETERS

After extracting the MGs, the investigators calculated the following parameters for the MGs:

2.1) Area ratio: Calculate the area of each MG (Smg) and the area of tarsus (St) in pixels. Then the area ratio is calculated by the formula below:

$$Area\ ratio = \frac{\Sigma Smg_i}{St}$$

2.2) Height and width of each MG: The height is the vertical difference between the top pixel and the bottom pixel of the MG, and the width is the area divided by the height.

2.3) Circumference of each MG: The circumference is the number of pixels on the edge of the MG.

2.4) Tortuosity of each MG: The tortuosity is calculated by the formula below:

$$Tortuosity = \frac{Circumference}{2 \times hight} - 1$$

2.5) Mean width of all MGs.

2.6) Mean width of the center 8 MGs.

D. STATISTICAL ANALYSIS

The normality of all data sets was evaluated by the Kolmogorov–Smirnov test. Either the independent samples t-test or the Mann–Whitney U test was used to compare differences between MGD subjects and normal control subjects. Because the age difference between the MGD subjects and normal controls was significant, an analysis of covariance was used to adjust for age when comparing the two groups. The correlations between various MG morphological factors (i.e., average height, average width, MGs area ratio, tortuosity of all MGs and central 8 MGs of the right upper or lower eyelid) and MG function parameters (i.e., OSDI, TBUT, CFS, lid margin score, meiboscore, and meibum expressibility score) were determined using Pearson's or Spearman's correlation analysis. The $\chi 2$ test was used to compare the sex ratios between the two groups. Student's t-test or the Mann–Whitney U test was utilized for MG tortuosity comparisons, depending on the data distribution. The correlations were further analyzed by Spearman's rank order correlation test. A two-sided $p < 0.05$ was considered statistically significant. All statistical analyses were performed using SPSS Statistics 21.0 (IBM, Armonk, NY).

TABLE 1. Clinical parameters of the 2 groups in the study population.

Parameter	Normal (n=28)	MGD (n=32)	*P*	*P**
Age (years, mean ± SD)	25.25 ± 11.19	33.28 ± 9.28	0.004	-
Sex (n, male/female)	11/17	16/16	0.409	-
OSDI (0-100)	6.89 ± 10.27	23.66± 18.94	< 0.001	0.001
Symptom score (0-14)	1.74 ± 1.74	4.80 ± 2.93	0.002	0.007
TBUT (seconds)	7.14 ± 4.08	2.88 ± 1.99	< 0.001	< 0.001
CFS (0-20)	0.11± 0.42	0.75 ± 1.52	0.028	0.150
TMH (mm)	0.19 ± 0.06	0.21 ± 0.06	0.201	0.371
Lid margin score (0-4)	0.46 ± 0.74	2.59 ± 1.50	< 0.001	< 0.001
Meiboscore (0-6)	1.46 ± 0.74	2.78 ± 1.24	< 0.001	< 0.001
Meibum expressibility score (0-45)	38.29 ± 7.99	17.19 ± 15.32	< 0.001	0.035

MGD= meibomian gland dysfunction
OSDI=Ocular Surface Disease Index
TBUT=tear break-up time
CFS=corneal fluorescein staining
TMH=tear meniscus height
* *P* values adjusted for age by analysis of covariance

III. RESULTS

The AI system training lasted for approximately 30 hours. During the training, the IoU of the training set continued to rise as the loss decreased. When the investigators identified the point where the IoU of the validation set stopped increasing and the loss stopped decreasing, the training would be stopped to prevent overfitting. We eventually obtained an IoU value of 0.9077 in the validation set. Additionally, with the GTX 1070 8G GPU, an MG image was analyzed in 100 milliseconds, and the repeatability reached 100%.

Then, a total of 60 eyes of 60 subjects were included in the AI automatic analysis study. The mean age of the subjects in the MGD group was 33.28 ± 9.28 years old, and that of the normal controls was 25.25 ± 11.19 years old. No significant difference in sex was observed between the MGD subjects and the normal controls. The baseline characteristics of the study groups are summarized in Table 1.

The average MGs height and area ratio in the upper eyelid were significantly correlated with OSDI ($r_{height} = -0.354$, $p_{height} = 0.005$) ($r_{area\ ratio} = -0.258$, $p_{arearatio} = 0.046$), TBUT ($r_{height} = 0.393$, $r_{arearatio} = 0.387$), lid margin score ($r_{height} = -0.336$, $r_{arearatio} = -0.258$), and meiboscore ($r_{height} = -0.562$, $r_{arearatio} = -0.589$) ($p < 0.05$). The results did not indicate significant correlations in the lower eyelid.

The average tortuosity of MGs in the lower eyelid was significantly correlated with OSDI ($r = 0.520$), TBUT ($r = -0.633$), meibum expressibility score ($r = -0.442$), meiboscore ($r = 0.335$), and lid margin score ($r = 0.568$) ($p < 0.05$). However, the average MGs tortuosity in the upper

TABLE 2. Correlations between MG tortuosity with tear film functions and MG status.

		OSDI	TBUT	CFS	TMH	Lid margin score	Meiboscore	Meibum expressibility score
Upper eyelid	average height of all MGs	-0.354†	0.393†	-0.158	0.105	-0.336*	-0.562‡	0.492‡
	average height of central 8	-0.387†	0.375†	-0.132	0.147	-0.302*	-0.628‡	0.421†
	average width of all MGs	-0.08	0.026	-0.108	0.166	0.012	0.136	0.026
	average width of central 8	0.021	0.046	-0.033	0.068	-0.013	-0.032	0.013
	Total MG area ratio	-0.258*	0.387†	-0.175	0.019	0.258*	-0.589‡	0.371†
	average tortuosity of all MGs	0.038	-0.170	0.077	-0.004	0.144	0.076	-0.383†
	average tortuosity of central 8	0.067	-0.030	0.167	-0.053	0.151	0.092	-0.442‡
Lower eyelid	average height of all MGs	-0.181	0.399†	-0.142	0.011	-0.398†	-0.276*	0.370†
	average height of central 8	-0.163	0.404†	-0.072	-0.016	-0.406†	-0.294*	0.389†
	average width of all MGs	0.091	0.009	-0.141	0.166	-0.018	-0.104	0.160
	average width of central 8	0.02	0.076	-0.017	-0.091	-0.061	-0.094	0.198
	Total MG area ratio	-0.053	0.309*	-0.132	0.093	0.224	-0.384†	0.349†
	average tortuosity of all MGs	0.141	-0.235	0.028	-0.006	0.270*	0.047	-0.445‡
	average tortuosity of central 8	0.520‡	-0.633‡	0.186	-0.094	0.568‡	0.335†	-0.418†
Total eyelid	average height of all MGs	-0.325*	0.374	-0.167	-0.068	-0.301*	-0.559‡	0.448‡
	average height of central 8 MGs	-0.390*	0.455†	-0.148	-0.099	-0.328*	-0.596‡	-0.479‡
	average width of all MGs	-0.039	0.039	-0.167	0.153	0.258*	0.099	0.042
	Total MG area ratio	-0.179	0.454‡	-0.191	0.069	-0.253*	-0.628‡	0.396†
	average tortuosity of all MGs	0.054	-0.280*	-0.165	-0.013	0.139	0.140	-0.421†

Spearman's rank correlation coefficient test.
*p<0.05
†p<0.005
‡p<0.001

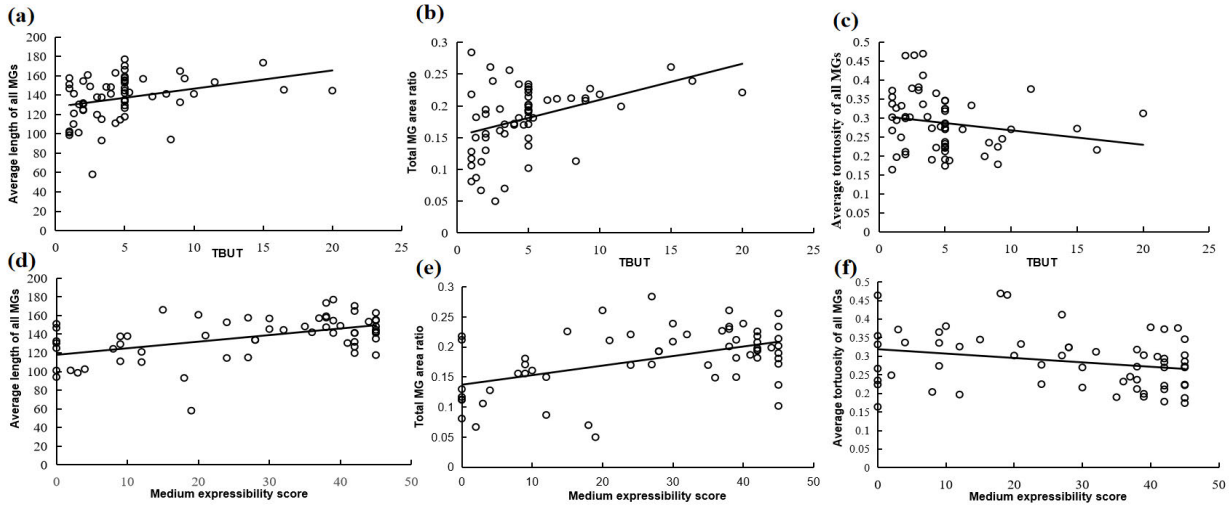


FIGURE 8. The TBUT was positively correlated with the average length of all MGs (a) and the total MG area ratio (b). The TBUT was negatively correlated with the average tortuosity of all MGs(c). The meibum expressibility score was positively correlated with the average length of all MGs (d) and the total MG area ratio (e). The meibum expressibility score was negatively correlated with the average tortuosity of all MGs (f).

eyelids had a significant correlation with the meibum expressibility score ($r = 0.383$) ($p < 0.001$).

The results are shown in Table 2 and Fig 8.

IV. DISCUSSION

Most AI aided MG morphology studies, however, have focused on MG atrophy; few of them has discussed the

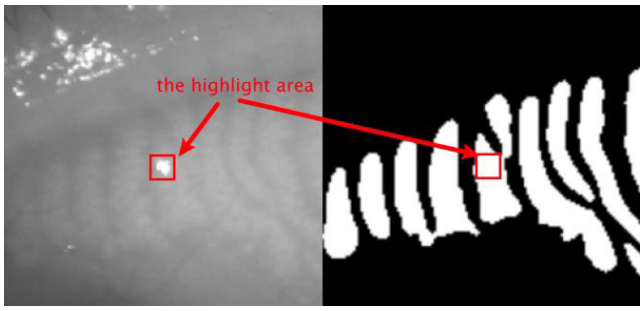


FIGURE 9. **The CNN model can correctly handle the highlight areas and recognize the highlight area as part of the gland.**

characteristics of MG morphology, such as tortuosity. The present study introduced an AI system based on a CNN, a relatively new technology in the computer vision field [15], [24], to automatically analyze MG morphology, including the proportion of the meibomian gland area along with the height, width, circumference and tortuosity of the MG. It was challenging to compute morphological features based on the already obtained gland lines. Previous studies set strict requirements for the input images, so the detection results were not satisfactory when eyelid folds appeared in the eyelid area and caused the eyelid gland to distort from its regular strip to a flake or a scattered shape [6], [7]. To solve this problem, this study's algorithms optimized image recognition, and the algorithms were combined with traditional digital image processing algorithms, such as contrast adjustment, low-pass filtering, corrosion, and expansion, to improve the recognition of meibomian glands in the background. Meanwhile, during the image shooting process, due to the inconsistency of light intensity, the captured images will also have inconsistencies in brightness. And due to the existence of the flash, it will cause the part of the captured image to have very strong highlights. Traditional MGD extraction methods based on image processing are based on the color values of pixels for gland extraction, as a result these methods are easily affected by light intensity. Using the same set of preprocessing parameters to process a large number of sample images with different light intensity is difficult to obtain a good recognition effect. At the same time, these methods cannot comprehensively consider a certain area of the gland and its adjacent area. If a certain part of the gland is just in the highlight area of the image, these methods cannot accurately extract it. Fig 9 shows that the CNN model can automatically connect the highlighted areas in the glands. The CNN-based model is trained in sample images under different light intensities. After much iteration, the model can automatically exclude the influence of different light intensities on gland extraction. As well as, CNN can identify glands with the same characteristics at different locations in the image through convolution operation; through down-sampling operation, it can correctly process the relative relationship of the glands, and the identified glands have better continuity.

Furthermore, comparing our two models of mini U-Net and the enhanced one, we find that the parameters of mini U-Net is only one-tenth of that of the enhanced one (mini U-Net has about 470,000 parameters, and the enhanced one has nearly 4.5 million parameters). The number of parameters represents the complexity of the model and also determines the amounts of calculations. For the complex models, including enhanced U-net and original U-net, the amount of calculation is greatly increased, but the accuracy on the same verification set is only improved by about 1%. On the other hand, we only have 60 training and validation images, which is relatively a small number for training a complex model, although we have done a lot of data augmentation algorithms [25], [26] to expand the training set, but training a complex model on small samples is much more likely to get overfitting. For the same reason, we did not use the more complex original U-Net for training. We are preparing larger training dataset to support more complex models such as ResNet50 [27], ResNet101 [27], FCN32 [28], SegNet [29] and PSPNet[30] in the future work. In general, in spite of a small number of subjects in this study was small, the accuracy of detecting MG morphology was very high (IoU = 90.77, repeatability = 100%).

A detailed analysis of MG morphology is an important index used to determine the extent and severity of MGD. Similar to previous studies [3], [31], [32], the AI system in the present study reveals that the MG's ductal height and area ratio are significantly correlated with symptoms, tear film stability, lid margin abnormality, and MG expressibility. Different from previous studies [9]–[11], [23], [33]–[35], the present study, with a high-performance and highly accurate AI system, can extract every single gland and analyze its height, width, tortuosity, and so on. The area proportion of MGs was calculated after each MG was extracted and accumulated. This reveals that the AI approach can obtain the lost area of the MG accurately and reduce the inconsistency and variability of the grading system caused by errors near the grading transition area. This MGs analysis method creates too much human workload to be used as a regular evaluation index, but such a task is appropriate for an AI system. With the technological advancements, the improvement in computational speed allows tens of thousands of images to be analyzed simultaneously and completely within milliseconds. As a result, the accuracy and efficiency of analysis can be greatly enhanced, the cost of manual analysis can be reduced, and the errors between different experience markers can be avoided.

Asymptomatic MGD has been documented as a preclinical stage of MGD with clinical changes but without lid-related symptoms [36]. Using this MGs analysis method, our team found in earlier study that MG tortuosity was significantly higher in asymptomatic MGD subjects than in normal controls and that MG tortuosity showed a significant correlation with TBUT, lid margin score, meiboscore, and meibum expressibility score [37]. Based on these findings, MG tortuosity may serve as a good index for diagnosing MGD. However, manual labeling can be time consuming and erroneous because of the impossibility of controlling

variables, making the popularization of this index difficult. The proposed AI system can help overcome these problems by measuring MG tortuosity quickly and accurately. While the team's previous study [37] had to analyze the reliability of parameters such as within-subject SD (SW), within-subject coefficient of variation (CVw), and intraclass correlation coefficient (ICC), the AI system prevents these manual errors. The present study, when repeated, reached perfect reliability (100%). Similar to our previous manual marking study [37], significant correlations between the average MG tortuosity in both eyelids and MGD parameters (i.e., TBUT, MG expressibility score, $p < 0.05$) were detected. These results suggest that MG tortuosity can be an applicable diagnostic method for MGD. When compared to the lower eyelid, the present study highlights that the changes in MG morphology (in terms of the height, the ratio of atrophy area, and so on) in the upper eyelid have more significant correlations with MGD symptoms, tear film stability, and lid margin abnormalities. These findings are consistent with studies that adopt a manual marking method [3], [5]. Furthermore, the results indicated that the lid margin score and the meiboscore of the upper eyelid were higher than those of the lower eyelid. This suggests that the analysis of MG morphology in the upper eyelid is sufficient to diagnose and treat subjects with MGD.

The present study also has limitations. The algorithm used in the study is a universal algorithm. If the algorithm is developed further into a special algorithm for MG images, MG morphology recognition is expected to have greater accuracy. Additionally, the sample size of the study is considered small. For future research, more subjects should be recruited for training and testing the AI system.

V. CONCLUSION

We proposed a novel MGs extraction method based on convolutional neural network (CNN) with enhanced mini U-Net for the diagnosis of MGD. The advantages of this study are as follows: small sample expansion, automatic gland morphology description/calculation, improvement of detection rate of early micro-lesions, and simplified CNN can reduce calculation time. This system increased the sensitivity and precision of MGD diagnosis, reduced high intra-observer variability and medical professionals' workloads, reduce analysis time, and assisting ophthalmologists with limited expertise. A CNN-based AI system can be used to efficiently and effectively analyze MG morphologic characteristics to help improve the efficiency of MGD treatment and patient management.

ACKNOWLEDGMENT

(Qi Dai and Xinyi Liu are co-first authors.) The authors thank Jing Ye for technological guidance. They also thank Xuewen Chen, Lu Li, Chaoqun Zhang, and Yingyu Mao, who work in the department of dry eye, for their help in collecting data.

REFERENCES

[1] J. P. Craig, K. Blades, and S. Patel, "Tear lipid layer structure and stability following expression of the meibomian glands," *Ophthalmic Physiol. Opt.*, vol. 15, no. 6, pp. 569–574,1995.

[2] C. Baudouin, E. M. Messmer, P. Aragona, G. Geerling, Y. A. Akova, J. Benítez-del-Castillo, K. G. Boboridis, J. Merayo-Lloves, M. Rolando, and M. and Labetoulle, "Revisiting the vicious circle of dry eye disease: A focus on the pathophysiology of meibomian gland dysfunction," *Br. J. Ophthalmol.*, vol. 100, no. 3, pp. 300–306,2016.

[3] G. Giannaccare, L. Vigo, M. Pellegrini, S. Sebastiani, and F. Carones, "Ocular surface workup with automated noninvasive measurements for the diagnosis of meibomian gland dysfunction," *Cornea*, vol. 37, no. 6, pp. 740–745, Jun. 2018.

[4] Y. Ban, S. Shimazaki-Den, K. Tsubota, and J. Shimazaki, "Morphological evaluation of meibomian glands using noncontact infrared meibography," *Ocular Surf.*, vol. 11, no. 1, pp. 47–53, Jan. 2013.

[5] H. Pult, B. H. Riede-Pult, and J. J. Nichols, "Relation between upper and lower lids' meibomian gland morphology, tear film, and dry eye," *Optom. Vis.*, vol. 89, no. 3, pp. 310–315, 2012.

[6] R. Arita, J. Suehiro, T. Haraguchi, R. Shirakawa, H. Tokoro, and S. Amano, "Objective image analysis of the meibomian gland area," *Br. J. Ophthalmol.*, vol. 98, no. 6, pp. 746–755, 2014.

[7] Y. W. Koh, "Detection of meibomian glands and classification of meibography images," *J. Biomed. Opt.*, vol. 17, no. 8, Aug. 2012, Art. no. 086008.

[8] J. Wang, T. N. Yeh, R. Chakraborty, S. X. Yu, and M. C. Lin, "A deep learning approach for meibomian gland atrophy evaluation in meibography images," *Transl. Vis. Sci. Technol.*, vol. 8, no. 6, p. 37, Dec. 2019.

[9] Y. W. Zhou, Y. Yu, Y. B. Zhou, Y. J. Tan, L. L. Wu, Y. Q. Xing, and Y. N. Yang, "An advanced imaging method for measuring and assessing meibomian glands based on deep learning," *Zhonghua Zhi.*, vol. 56, no. 10, pp. 774–779,2020.

[10] S. Maruoka, H. Tabuchi, D. Nagasato, H. Masumoto, T. Chikama, A. Kawai, N. Oishi, T. Maruyama, Y. Kato, T. Hayashi, and C. Katakami, "Deep neural network-based method for detecting obstructive meibomian gland dysfunction with *in vivo* laser confocal microscopy," *Cornea*, vol. 39, no. 6, pp. 720–725, Jun. 2020.

[11] A. Rajkomar, S. Lingam, A. G. Taylor, M. Blum, and J. Mongan, "High-throughput classification of radiographs using deep convolutional neural networks," *J. Digit. Imag.*, vol. 30, no. 1, pp. 95–101, Feb. 2017.

[12] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018.

[13] R. Gargeya and T. Leng, "Automated identification of diabetic retinopathy using deep learning," *Ophthalmology*, vol. 124, no. 7, pp. 962–969, Jul. 2017.

[14] S. Keel, J. Wu, P.Y. Lee, J. Scheetz, and M.G. He, "Visualizing deep learning models for the detection of referable diabetic retinopathy and glaucoma," *JAMA Ophthalmol.*, vol. 137, no. 3, pp. 288–292, Mar. 2019.

[15] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *Proc. Int. Conf. Med. Image Comput.-Assist. Intervent. (MICCAI)*, 2015, pp. 234–241.

[16] R. Arita, K. Itoh, S. Maed, K. Maeda, A. Furuta, S. Fukuoka, A. Tomidokor, and S. Amano, "Proposed diagnostic criteria for obstructive meibomian gland dysfunction," *Ophthalmology*, vol. 116, no. 11, pp. 2058–2063, Nov. 2009.

[17] M. Markoulli, T. B. Duong, M. Lin, and E. Papas, "Imaging the tear film: A comparison between the subjective keeler tearscope-plus and the objective Oculus keratograph 5M and LipiView interferometer," *Current Eye Res.*, vol. 43, no. 2, pp. 155–162, Feb. 2018.

[18] A. Dutta and A. Zisserman, "The VIA annotation software for images, audio and video," in *Proc. 27th ACM Int. Conf. Multimedia*, Oct. 2019, pp. 2276–2279.

[19] F. Chollet, *Keras*. San Francisco, CA, USA: Google, 2015. [Online]. Available: https://keras.io/

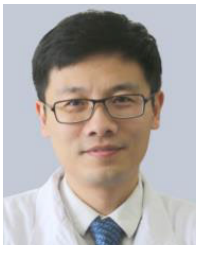
[20] M. Abadi, P. Barham, J. Chen, Z. Chen, and A. Davis, "TensorFlow: A system for large-scale machine learning," in *Proc. OSDI*, vol. 16, 2016, pp. 265–283.

[21] C. Sade De Paiva and S. C. Pflugfelder, "Corneal epitheliopathy of dry eye induces hyperesthesia to mechanical air jet stimulation," *Amer. J. Ophthalmol.*, vol. 137, no. 1, pp. 109–115, Jan. 2004.

[22] B. Cochener, A. Cassan, and L. Omiel, "Prevalence of meibomian gland dysfunction at the time of cataract surgery," *J. Cataract Refractive Surg.*, vol. 44, no. 2, pp. 144–148, Feb. 2018.

[23] R. Arita, K. Itoh, K. Inoue, and S. Amano, "Noncontact infrared meibography to document age-related changes of the meibomian glands in a normal population," *Ophthalmology*, vol. 115, no. 5, pp. 911–915, May 2008.

[24] R. Hemelings, B. Elen, J. Barbosa-Breda, S. Lemmens, M. Meire, S. Pourjavan, E. Vandewalle, S. Van De Veire, M. B. Blaschko, P. De Boever, and I. Stalmans, "Accurate prediction of glaucoma from colour fundus images with a convolutional neural network that relies on active and transfer learning," *Acta Ophthalmol.*, vol. 98, no. 1, pp. e94–e100, Feb. 2020.
[25] S. C. Wong, A. Gatt, V. Stamatescu, and M. D. McDonnell, "Understanding data augmentation for classification: When to warp?" in *Proc. Int. Conf. Digit. Image Comput.*, Nov. 2016, pp. 1–6, doi: 10.1109/DICTA.2016.7797091.
[26] A. Miko-Ajczyk and M. Grochowski, "Data augmentation for improving deep learning in image classification problem," in *Proc. Int. Interdiscipl. PhD Workshop (IIPhDW)*, May 2018, pp. 117–122, doi: 10.1109/IIPHDW.2018.8388338.
[27] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.
[28] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Apr. 2017.
[29] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.
[30] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2881–2890, doi: 10.1109/CVPR.2017.660.
[31] K. B. Green-Church, I. Butovich, M. Willcox, D. Borchman, F. Paulsen, S. Barabino, and B. J. Glasgow, "The international workshop on meibomian gland dysfunction: Report of the subcommittee on tear film lipids and lipid-protein interactions in health and disease," *Invest. Ophthalmol. Vis. Sci.*, vol. 52, no. 4, pp. 1979–1993, Mar. 2011.
[32] J. D. Nelson, J. Shimazaki, J. M. Benitez-del-Castillo, J. P. Craig, J. P. McCulley, S. Den, and G. N. Foulks, "The international workshop on meibomian gland dysfunction: Report of the definition and classification subcommittee," *Invest. Ophthalmol. Vis. Sci.*, vol. 52, no. 4, pp. 1930–1937, Mar. 2011.
[33] S. C. Pflugfelder, S. C. Tseng, O. Sanabria, H. Kell, C. G. Garcia, C. Felix, W. Feuer, and B. L. Reis, "Evaluation of subjective assessments and objective diagnostic tests for diagnosing tear-film disorders known to cause ocular irritation," *Cornea*, vol. 17, no. 1, pp. 38–56, Jan. 1998.
[34] J. J. Nichols, D. A. Berntsen, G. L. Mitchell, and K. K. Nichols, "An assessment of grading scales for meibography images," *Cornea*, vol. 24, no. 4, pp. 382–388, May 2005.
[35] H. Pult and J. J. Nichols, "A review of meibography," *Optometry Vis. Sci.*, vol. 89, no. 5, pp. E760–E769, May 2012.
[36] P. A. Asbell, F. J. Stapleton, K. Wickström, E. K. Akpek, P. Aragona, R. Dana, M. A. Lemp, and K. K. Nichols, "The international workshop on meibomian gland dysfunction: Report of the clinical trials subcommittee," *Invest. Ophthalmol. Vis. Sci.*, vol. 52, no. 4, pp. 2065–2085, Mar. 2011.
[37] X. Lin, Y. Fu, L. Li, C. Chen, X. Chen, Y. Mao, H. Lian, W. Yang, and Q. Dai, "A novel quantitative index of meibomian gland dysfunction, the meibomian gland tortuosity," *Transl. Vis. Sci. Technol.*, vol. 9, no. 9, p. 34, Aug. 2020.



QI DAI received the M.D. degree from Zhejiang University, China. He is currently the Deputy Director of the Corneal Disease Department, Zhejiang Eye Hospital, Wenzhou Medical University, Hangzhou. His current research interests include treatment of corneal disease, including dry eye, keratoplasty, and AI for eye diseases.

XINYI LIU received the M.D. degree from Wenzhou Medical University, China. She is currently pursuing the master's degree with the Eye Hospital, Wenzhou Medical University.

XIAOLEI LIN received the M.D. degree from the Zhongshan Ophthalmic Center, Sun Yat-sen University, China. Her research interests include dry eye, meibomian gland dysfunction, and cataract.

YANA FU received the M.D. degree from Zhejiang University, China. She is currently the Deputy Director of the Dry eye Department, Zhejiang Eye Hospital, Wenzhou Medical University, Hangzhou. Her current research interests include treatment of cornea disease and cataract, including dry eye and cataract.

CHAOQIAO CHEN received the M.D. degree from Zhejiang University, China. She is currently an Anesthetist with the Zhejiang Eye Hospital, Wenzhou Medical University, Hangzhou. Her current research interests include outpatient department of ophthalmology for children sedative anesthesia, and general anesthesia for infants.

XINXIN YU received the M.D. and Ph.D. degrees from Wenzhou Medical University, China. She is currently a Resident with the Eye Hospital, Wenzhou Medical University.

ZUHUI ZHANG received the M.D. degree from Wenzhou Medical University, China. He is currently a Resident with the Eye Hospital, Wenzhou Medical University.

TIANKUN LI received the M.D. degree from Wenzhou Medical University, China. He is currently pursuing the master's degree with the Eye Hospital, Wenzhou Medical University.

MENGTING LIU is currently a Student with Wenzhou Medical University.

WEIHUA YANG graduated from Soochow University. He is currently a Medical Doctor in ophthalmic artificial intelligence. He is also an Associate Professor of ophthalmology and the Director of the Laboratory of Ophthalmology Artificial Intelligence and Big Data, Affiliated Eye Hospital.

JUAN YE received the M.D. degree from Zhejiang University, China, and the Ph.D. degree from Yonsei University, South Korea. She is currently a Professor with the Department of Ophthalmology, Second Affiliated Hospital, Zhejiang University, and the Deputy Director of the Chinese Society of Oculoplastic Surgery and Orbital Disease. Her current research interests include treatment of eyelid deformity, orbital reconstruction, and epidemiology of age-related eye disease.