



Deep learning-based image analysis of eyelid morphology in thyroid-associated ophthalmopathy

Ji Shao^{1#}, Xingru Huang^{2#}, Tao Gao¹, Jing Cao¹, Yaqi Wang³, Qianni Zhang², Lixia Lou¹, Juan Ye¹

¹Department of Ophthalmology, The Second Affiliated Hospital of Zhejiang University, School of Medicine, Hangzhou, China; ²School of Electronic Engineering and Computer Science, Queen Mary University of London, London, UK; ³College of Media Engineering, Communication University of Zhejiang, Hangzhou, China

Contributions: (I) Conception and design: J Ye, L Lou, J Shao, X Huang; (II) Administrative support: J Ye; (III) Provision of study materials or patients: J Ye, T Gao; (IV) Collection and assembly of data: J Shao, X Huang, J Cao; (V) Data analysis and interpretation: J Shao, X Huang, Q Zhang, Y Wang; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Juan Ye; Lixia Lou. Department of Ophthalmology, The Second Affiliated Hospital of Zhejiang University, School of Medicine, No. 88 Jiefang Road, Hangzhou 310009, China. Email: yejuan@zju.edu.cn; loulixia110@zju.edu.cn.

Background: We aimed to propose a deep learning-based approach to automatically measure eyelid morphology in patients with thyroid-associated ophthalmopathy (TAO).

Methods: This prospective study consecutively included 74 eyes of patients with TAO and 74 eyes of healthy volunteers visiting the ophthalmology department in a tertiary hospital. Patients diagnosed as TAO and healthy volunteers who were age- and gender-matched met the eligibility criteria for recruitment. Facial images were taken under the same light conditions. Comprehensive eyelid morphological parameters, such as palpebral fissure (PF) length, margin reflex distance (MRD), eyelid retraction distance, eyelid length, scleral area, and mid-pupil lid distance (MPLD), were automatically calculated using our deep learning-based analysis system. MRD1 and 2 were manually measured. Bland-Altman plots and intraclass correlation coefficients (ICCs) were performed to assess the agreement between automatic and manual measurements of MRDs. The asymmetry of the eyelid contour was analyzed using the temporal: nasal ratio of the MPLD. All eyelid features were compared between TAO eyes and control eyes using the independent samples *t*-test.

Results: A strong agreement between automatic and manual measurement was indicated. Biases of MRDs in TAO eyes and control eyes ranged from -0.01 mm [95% limits of agreement (LoA): -0.64 to 0.63 mm] to 0.09 mm (LoA: -0.46 to 0.63 mm). ICCs ranged from 0.932 to 0.980 ($P < 0.001$). Eyelid features were significantly different in TAO eyes and control eyes, including MRD1 (4.82 ± 1.59 vs. 2.99 ± 0.81 mm; $P < 0.001$), MRD2 (5.89 ± 1.16 vs. 5.47 ± 0.73 mm; $P = 0.009$), upper eyelid length (UEL) (27.73 ± 4.49 vs. 25.42 ± 4.35 mm; $P = 0.002$), lower eyelid length (LEL) (31.51 ± 4.59 vs. 26.34 ± 4.72 mm; $P < 0.001$), and total scleral area ($S_{A_{TOTAL}}$) (96.14 ± 34.38 vs. 56.91 ± 14.97 mm²; $P < 0.001$). The MPLDs at all angles showed significant differences in the 2 groups of eyes ($P = 0.008$ at temporal 180°; $P < 0.001$ at other angles). The greatest temporal-nasal asymmetry appeared at 75° apart from the midline in TAO eyes.

Conclusions: Our proposed system allowed automatic, comprehensive, and objective measurement of eyelid morphology by only using facial images, which has potential application prospects in TAO. Future work with a large sample of patients that contains different TAO subsets is warranted.

Keywords: Thyroid-associated ophthalmopathy (TAO); eyelid morphology; automatic measurement; facial images; deep learning

Submitted Jun 03, 2022. Accepted for publication Nov 25, 2022. Published online Jan 03, 2023.

doi: 10.21037/qims-22-551

View this article at: <https://dx.doi.org/10.21037/qims-22-551>

Introduction

Thyroid-associated ophthalmopathy (TAO) occurs mainly in patients with hyperthyroidism and sometimes in patients with hypothyroidism, Hashimoto thyroiditis, or euthyroidism (1). The prevalence of TAO is higher in females than in males (1-4). However, more severe cases tend to be more frequent in males and patients at a more advanced age (5). As an autoimmune disease, TAO mainly affects eyelids, extraocular muscles, and adipose tissue in the orbit. The clinical signs and symptoms are broad, with ocular features varying from eyelid abnormality, exophthalmos, diplopia, restrictive ocular motility, to optic nerve dysfunction (1). Eyelid abnormality, presented mainly as a retraction of the upper and lower eyelid, is one of the most common signs of TAO (6). Exposure of the cornea followed by lid retraction results in dry eye, keratitis, and strong concerns about appearance (7,8).

Accurate measurement of eyelid characteristics is essential for the diagnosis, severity grading, surgery design, and evaluation of the treatment effectiveness of TAO. Traditional assessment of eyelid position in TAO is obtained by clinicians' manual measurement with a ruler; therefore, it mainly focuses on 1-dimensional features, such as palpebral fissure (PF) length and margin reflex distance (MRD) (9,10). Acquiring accurate eyelid parameters using manual measurements requires clinicians to be highly experienced and needs the cooperation of patients. It is also challenging to obtain continuous and stable measurements due to the interobserver difference during the follow-up period. Thus, attaining standardized and detailed measurements of eyelid features is of crucial importance for improving the diagnosis and treatment of TAO.

A few studies have focused on analyzing digital face images to reliably measure eyelid characteristics. Edwards *et al.* (11) confirmed the reliability of computer-based image measurement of eyelid position with excellent intraobserver and interobserver agreement. However, the clinicians' workloads were still heavy since the eyelid heights were measured by placing calipers directly on the screen. In recent years, several studies attempted to calculate the eyelid parameters of patients with TAO using various types of software in order to quantitatively analyze eyelid

abnormalities and evaluate the effectiveness of different surgical methods (12-17). Nevertheless, subjective factors were present because manual measurement processes were still involved. Deep learning, a representative branch of artificial intelligence, has shown advanced performance in automatic segmentation using medical images (18,19). Our team proposed an automatic system for eyelid measurement and applied it to patients with blepharoptosis before and after surgery (20). An objective and fully automatic computer-based assessment system for patients with TAO is critically needed and would play an important role in the clinic.

Therefore, we hypothesized that an automatic eyelid analysis system based on deep learning could be established to accurately measure comprehensive eyelid features in patients with TAO and healthy controls. We present the following article in accordance with the STARD reporting checklist (available at <https://qims.amegroups.com/article/view/10.21037/qims-22-551/rc>).

Methods

Participants

This study followed the principles of the Declaration of Helsinki (as revised in 2013) and was approved by the Ethics Committees of the Second Affiliated Hospital of Zhejiang University, School of Medicine (Approval No. 2020-583). Informed consent was obtained from all adult participants and the guardians of minors. This study was registered with ClinicalTrials.gov (No. NCT04921020).

This prospective study consecutively included patients with TAO in the ophthalmology department of the Second Affiliated Hospital of Zhejiang University, School of Medicine between November 2020 and November 2021. Age- and gender-matched healthy volunteers who visited the hospital for vision screening were recruited as the control group. The diagnosis of TAO was based on the Bartley criteria (21). One eye was randomly selected when the patient was diagnosed with bilateral TAO, and the ipsilateral eye of the age- and gender-matched healthy control was included in the study. Participants with strabismus, coexisting eyelid diseases (e.g., ptosis,

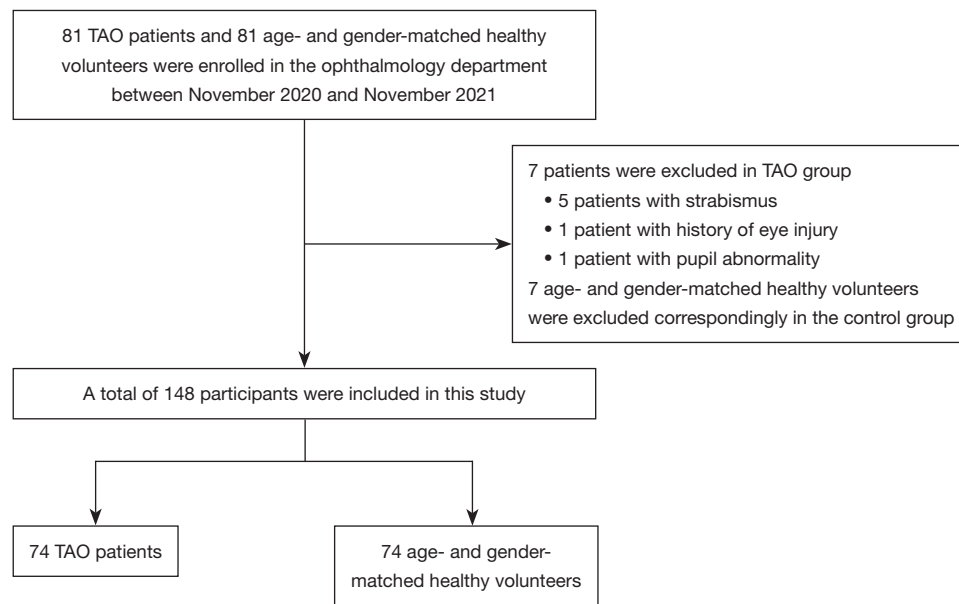


Figure 1 Flowchart of the study population. TAO, thyroid-associated ophthalmopathy.

Table 1 Demographic information of participants included in this study

Characteristics	TAO group	Control group	P value
Age (years) ^a	43.76±13.69	43.28±12.84	0.663
Sex ^b , n (%)			1.000
Female	57 (77.03)	57 (77.03)	
Male	17 (22.97)	17 (22.97)	
Number of participants ^b , n (%)	74 (100.00)	74 (100.00)	
Right eye included	37 (50.00)	37 (50.00)	
Left eye included	37 (50.00)	37 (50.00)	

^a, data are presented as mean ± standard deviation. ^b, data are presented as number (frequency). TAO, thyroid-associated ophthalmopathy.

blepharospasm), a history of eye injury, or abnormalities of the cornea and pupil were excluded.

Ultimately, 148 eyes of 148 participants, including 74 eyes of 74 patients with TAO (mean ± SD of age: 43.76±13.69 years old) and 74 eyes of 74 healthy volunteers (43.28±12.84 years old) were recruited for this study (Figure 1). In the TAO group, 38 patients (51.35%) were diagnosed with bilateral TAO, and 36 patients were diagnosed (48.65%) with unilateral TAO. The TAO group consisted of 67 patients with hyperthyroidism, 4 patients with euthyroidism, 2 patients with Hashimoto thyroiditis, and 1 patient with primary hypothyroidism.

The demographic characteristics of 148 participants are shown in Table 1. MRD1 and MRD2 in 74 TAO and 74 control eyes were manually measured by an experienced ophthalmologist blinded to the automatic measurement results. The eyes of the examiner and the patient were on the same level, and the examiner held a penlight that was directed at the patient's eyes. The MRD1 was defined as the vertical distance between the upper eyelid margin and the cornea reflection of the penlight, with the patient in the primary gaze. The MRD2 was defined as the vertical distance between the lower eyelid margin and the reflection on the cornea (9). No adverse events occurred during this study.

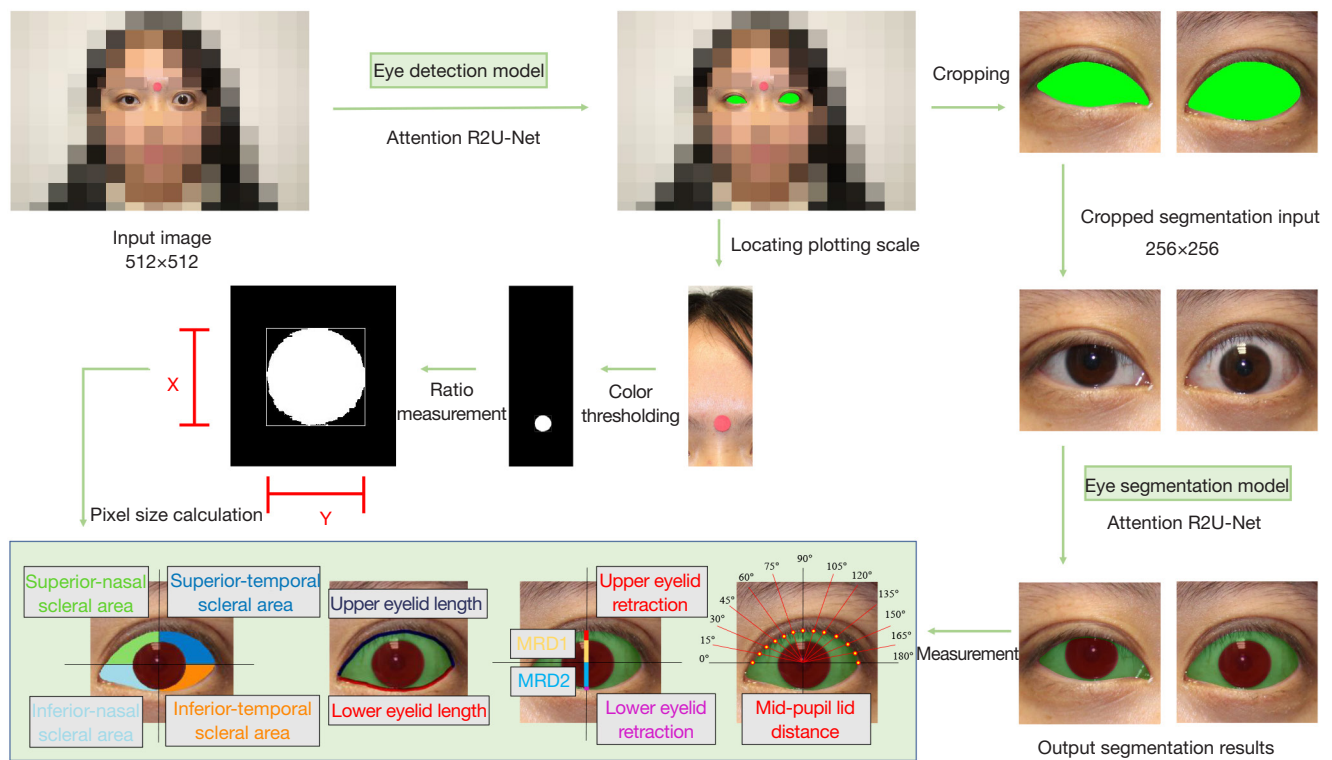


Figure 2 Workflow of the automatic eyelid analysis system. Recurrent residual convolutional neural networks with an attention gate connection based on U-Net (Attention R2U-Net) were used in the first stage for the detection of the eye and in the second stage for the segmentation of the eyelid and cornea contour. After the circle marker that was attached to the forehead of participants was located for pixel calculation, the eyelid parameters were transformed into the actual distances. This image has been published with the participant's consent. MRD, margin reflex distance.

Image collection

Facial images were obtained by a digital camera (Canon 500D with a 100 mm macro lens; Canon Corporation, Tokyo, Japan) when participants gazed in the primary position and kept their eyes open naturally. A red circular marker was attached to their forehead as a reference for measuring the distance in reality. All the images were taken under the same lighting conditions, with the camera fixed by a tripod and set at 1 meter in front of the patient at eye level.

Automatic image analyses by deep learning

The automatic eyelid analysis system used deep learning networks to locate the eye region and segment the eyelid margin and the corneal limbus. After locating the plotting scale, pixel values of measured eyelid parameters were

converted into the distances in reality. The workflow is shown in *Figure 2*.

The proportion of human eyes in the facial images was relatively small. To ensure the accuracy of eyelid and cornea segmentation, we trained the eye detection model and the eye segmentation model separately. Recurrent residual convolutional neural networks with attention gate connection based on U-Net (Attention R2U-Net) were used in our study (22). R2U-Net replaced the traditional convolutional block in each layer with a recurrent residual convolutional unit. This enabled the network to capture low-layer features and generate segmentation with higher accuracy in challenging areas than that with traditional convolutional block. The structure of the Attention R2U-Net is shown in *Figure 3*. The distribution of images in the training set, validation set, and test set is shown in *Table 2*.

In step 1, we trained the eye detection model. A total of 30,000 facial images with landmark locations of

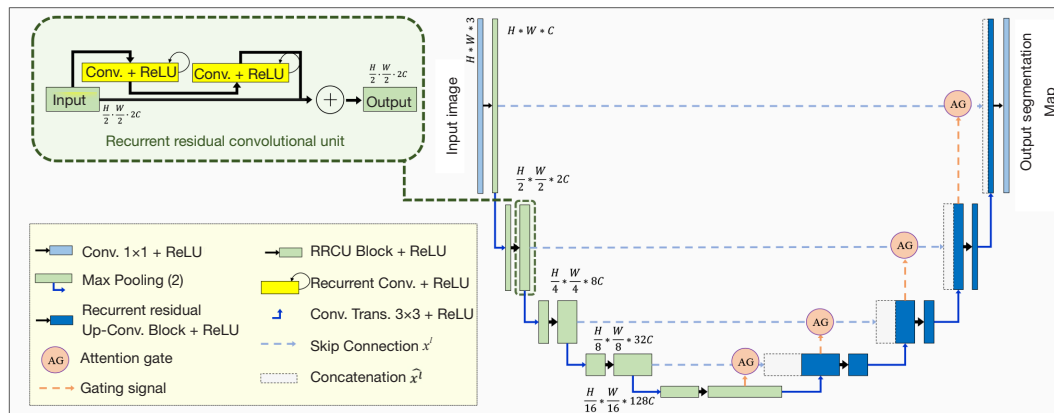


Figure 3 The architecture of the recurrent residual convolutional neural networks with an attention gate connection based on U-Net (Attention R2U-Net). Conv., convolution; ReLU, rectified linear unit; RRCU, recurrent residual convolutional unit; H, height; W, weight; C, channel; AG, attention gate.

Table 2 The distribution of images in the eye detection and segmentation model

Distribution	Training	Validation	Test
Eye detection			
Number of facial images	18,000	3,000	9,000 and 148
Eye segmentation			
Number of participants	1,726	136	148
Number of eye images	3,452	272	148

the eye were extracted from the CelebFaces Attributes Dataset (23). These images were used to train the eye localization network via the first-stage eye detection model, and the Attention R2U-Net was employed as the backbone [logistic loss function: binary cross entropy loss; optimizer: Adam (lr =0.00001); input image size =512×512 pixels; epoch =200; batch size =4].

In step 2, we trained the eye segmentation model. A total of 1,862 facial images of healthy volunteers were collected from the Second Affiliated Hospital of Zhejiang University, School of Medicine. Two clinicians were asked to outline the eyelid margin and the corneal limbus. These labeled images were used to train an eye segmentation model, again with the Attention R2U-Net as the backbone [logistic loss function: L1 loss; optimizer: Adam (lr =0.00001); input image size =256×256 pixels; epoch =200; batch size =4].

In step 3, new images were predicted. Facial images of 148 participants, including 74 TAO patients and 74 healthy volunteers, were used as the test set. In this set, the

cornea and eyelid margin of 74 TAO eyes and 74 control eyes were delineated manually to evaluate the accuracy of automatic segmentation. In the image postprocessing stage, several versions of each test image were acquired through elastic transformation, random rotation, and random multiscaling. Then, segmentation predictions were obtained on the different transformed versions of the image, and the segmentation boundaries on all versions were fused to generate an overall boundary. The resulting fused segmentation boundary in the eye region is believed to be more robust to instance variation factors. The final output boundary was smoothed to obtain the eyelid mask and corneal limbus mask.

In step 4, pixel size was calculated. Threshold segmentation of the circular marker (10 mm in diameter) on the forehead was performed. Then, the millimeter:pixel ratio (R) was calculated so that the size in pixels could be transformed into the distance in real space.

In step 5, features were measured. The measurements

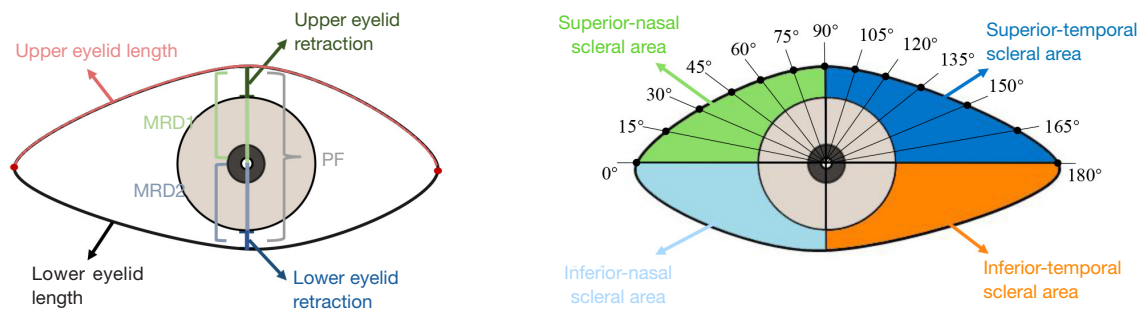


Figure 4 A schematic diagram of the eyelid morphological parameters. MRD1 and 2 refer to the vertical distances from the pupil center to the upper and lower eyelids, respectively. PF is the sum of MRD1 and MRD2. The upper and lower eyelids were separated according to the inner and outer canthus. The scleral area was divided into 4 parts centered on the pupil. A radial line was drawn every 15° to calculate mid-pupil lid distances. MRD, margin reflex distance; PF, palpebral fissure.

of PF, MRD1, MRD2, eyelid length [including upper eyelid length (UEL) and lower eyelid length (LEL)], eyelid retraction distance [including upper eyelid retraction (UER) and lower eyelid retraction (LER)], total scleral area [SA_{TOTAL} ; including superior-nasal (SN), superior-temporal (ST), inferior-nasal (IN), and inferior-temporal (IT) scleral area], and mid-pupil lid distance (MPLD) of 74 TAO eyes and 74 control eyes were automatically conducted based on the masked images (Figure 4).

Three points of the corneal limbus were randomly selected to fit the circle, and the center of this circle was defined as the pupil center. Since the cornea is not a perfect circle (24), this process was repeated 2,000 times, and the pupil center was finally located using mean shift with a Gaussian kernel (25). Then, the numbers of vertical pixels from the upper eyelid to the pupil center and from the pupil center to the lower eyelid were defined as PN_{MRD1} and PN_{MRD2} , respectively. A vertical line was drawn across the pupil center, which intersected the upper eyelid and upper corneal limbus. The pixel number between these 2 intersections was noted as PN_{UER} . Similarly, the pixel number between the lower corneal limbus and lower eyelid in vertical orientation was recorded as PN_{LER} . When the cornea was covered by the eyelid, the value of the eyelid retraction distance was denoted as zero. We defined the outmost pixels in the eyelid border as the outer and inner canthus to differentiate the upper and lower eyelid. The pixel numbers of the UEL and LEL were calculated as PN_{UEL} and PN_{LEL} , respectively. In addition, a horizontal line and a vertical line across the pupil center were automatically drawn to separate the SN, ST, IN, and IT scleral areas. Twelve radial lines originated from the pupil

center intersected the nasal sector (0°, 15°, 30°, 45°, 60°, and 75°) and temporal sector (105°, 120°, 135°, 150°, 165°, and 180°) of the upper eyelid on each 15 degree. To assess the symmetry of the upper eyelid, the pixel numbers from the pupil center to each intersection on the upper eyelid were noted as $PN_{N/T}$. Finally, the measurements of all eyelid morphological parameters were converted into mm or mm^2 as follows (Figure S1):

$$MRD1 = PN_{MRD1} \times R \quad [1]$$

$$MRD2 = PN_{MRD2} \times R \quad [2]$$

$$PF = MRD1 + MRD2 \quad [3]$$

$$UER / LER = PN_{UER/LER} \times R \quad [4]$$

$$UEL / LEL = PN_{UEL/LEL} \times R \quad [5]$$

$$SN / ST / IN / IT = PN_{SN/ST/IN/IT} \times R^2 \quad [6]$$

$$SA_{TOTAL} = SN + ST + IN + IT \quad [7]$$

$$MPLD_{N0/15/30/45/60/75} = PN_{N0/15/30/45/60/75} \times R \quad [8]$$

$$MPLD_{T105/120/135/150/165/180} = PN_{T105/120/135/150/165/180} \times R \quad [9]$$

Statistical analyses

The accuracy of eye segmentation tasks was assessed by Dice coefficients and Intersection over Union (IoU), which are statistical tools to calculate the similarity of manual

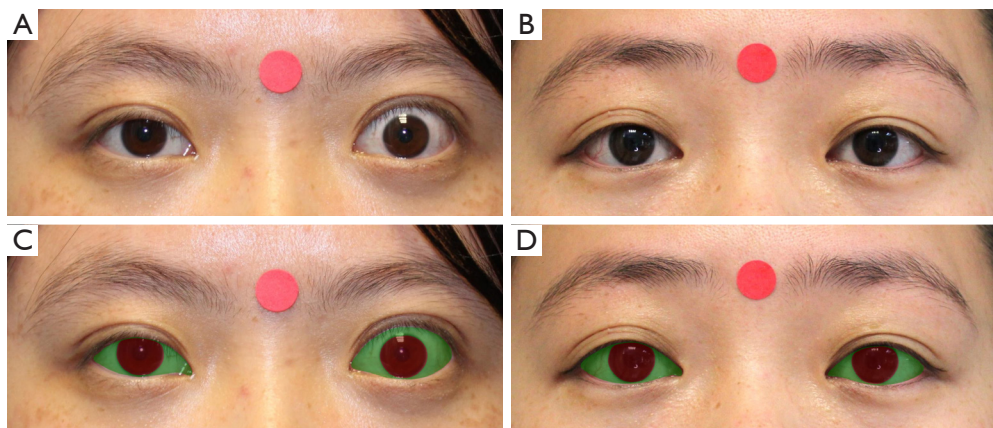


Figure 5 Representative results of the automatic eyelid and cornea segmentation based on deep learning. (A,B) The original images of a TAO patient and an age- and gender-matched healthy volunteer. (C,D) The automatically segmented images of the eyelid and cornea. (C,D) The cornea is marked red, and the sclera is marked green. This has been published with the participants' consent. TAO, thyroid-associated ophthalmopathy.

annotation and automatic segmentation. The intraclass correlation coefficients (ICCs) of MRD1 and MRD2 were calculated to evaluate the agreement between automatic and manual measurements, as well as the agreement between 2 repeated automatic MRD measurements. Higher ICC values implied greater agreement, with $0.41 < \text{ICC} \leq 0.60$ indicating moderate agreement, $0.60 < \text{ICC} \leq 0.80$ indicating substantial agreement, and $0.80 < \text{ICC} \leq 1.00$ indicating excellent agreement (26). To visualize the difference between MRDs, Bland-Altman plots were also drawn. In these scatterplots, the Y-axis represented the difference between 2 measurements, and the X-axis represented the average. Lower mean bias values and more clustered points between confidence limits implied higher agreement between the 2 measurements (manual *vs.* automatic; 2 repeated automatic measurements). All eyelid morphological parameters were compared in TAO eyes and control eyes with the independent samples *t*-test. The upper eyelid contours of the TAO eyes and control eyes were drawn based on the mean value of MPLD at each angle, and the difference in MPLDs at each angle between the 2 group eyes was evaluated. A P value less than 0.05 was considered statistically significant. All statistical analyses were performed with SPSS 25.0 (IBM Corporation, Armonk, NY, USA).

The minimum sample size was calculated using PASS (version 2021; NCSS, Kaysville, UT, USA). The alpha was prespecified as 0.05, and the width of the confidence interval was prespecified to be 0.10. The minimum number

of 62 eyes was required to meet the smallest possible value of 0.90 for the ICC (27). This study was conducted from November 2020 to November 2021 to include the required number of eyes.

Results

Model performance

The eye detection model reached an accuracy of 0.996 on the CelebFaces Attributes Dataset and 0.985 on the dataset consisting of 148 participants (Table S1). The Dice coefficients for eye segmentation tasks in the test set were 0.947 for the eyelid and 0.952 for the cornea. The IoU value was 0.903 and 0.912 for eyelid and cornea segmentation, respectively. The 4-fold cross-validation was performed in the dataset of 1,862 participants. The mean IoU was 0.896 for the eyelid and 0.908 for the cornea, which validated the robustness of the automatic model (Table S2). Figure 5 and Figure S2 exhibit the performance of segmentation of eyelid and cornea in the TAO group and control group.

Measurement agreement

Table 3 displays the repeated automatic and manual measurements of MRD1 and MRD2 in both groups. The ICC was 0.980 [95% confidence interval (CI): 0.969–0.988; $P < 0.001$] for MRD1 and 0.964 (95% CI: 0.943–0.977; $P < 0.001$) for MRD2 in TAO eyes, and 0.967 (95% CI:

0.949–0.979; $P < 0.001$) for MRD1 and 0.932 (95% CI: 0.888–0.958; $P < 0.001$) for MRD2 in control eyes, which indicated excellent agreement between automatic and manual measurement (Table 4). The ICCs between repeated automatic measurements of MRDs were up to 0.998 ($P < 0.001$), showing the high repeatability of the automatic system.

Bland-Altman plots (Figure 6; Figure S3) also confirmed the consistency between any 2 measurements. The bias [95% limits of agreement (LoA)] between automatic and manual measurements was 0.06 (–0.54 to 0.67) mm for MRD1 and –0.01 (–0.64 to 0.63) mm for MRD2 in TAO eyes. The bias (95% LoA) between automatic and manual measurements was 0.01 (–0.39 to 0.41) mm for MRD1 and 0.09 (–0.46 to 0.63) mm for MRD2 in control eyes.

Table 3 Manual and 2 repeated automatic measurements of MRD1 and MRD2 in 148 participants

Measurements	MRD1 (mm)	MRD2 (mm)
TAO		
Manual	4.76±1.55	5.90±1.23
First automatic	4.82±1.59	5.89±1.16
Second automatic	4.82±1.59	5.89±1.16
Control		
Manual	2.98±0.77	5.38±0.81
First automatic	2.99±0.81	5.47±0.73
Second automatic	2.99±0.81	5.46±0.74

Data are presented as mean ± standard deviation. MRD, margin reflex distance; TAO, thyroid-associated ophthalmopathy.

Table 4 Intraclass correlation coefficients between 2 measurements of MRD1 and MRD2 in 148 participants

Measurements	MRD1	MRD2
TAO		
Manual and automatic	0.980 (0.969–0.988)***	0.964 (0.943–0.977)***
Automatic first and second	0.999 (0.999–1.000)***	0.999 (0.999–0.999)***
Control		
Manual and automatic	0.967 (0.949–0.979)***	0.932 (0.888–0.958)***
Automatic first and second	0.999 (0.998–0.999)***	0.998 (0.997–0.999)***

Data are presented as intraclass correlation (95% confidence interval). ***, $P < 0.001$. MRD, margin reflex distance; TAO, thyroid-associated ophthalmopathy.

Comparison of eye characteristics in TAO and control eyes

Evaluation of eyelid morphological parameters

The automatic measurements of eyelid morphological features of TAO eyes and control eyes are listed in Table 5. The independent samples *t*-test revealed significantly greater PF, MRD1, and MRD2 in TAO eyes. Notably, MRD1 was 4.82±1.59 mm in TAO eyes vs. 2.99±0.81 mm in control eyes ($P < 0.001$). TAO eyes also had longer eyelid lengths, with values of 27.73±4.49 mm in the upper eyelid and 31.51±4.59 mm in the lower eyelid. The difference in UER between TAO eyes and control eyes was 0.52 mm, and the difference in LER was 0.40 mm, indicating a stronger effect of thyroid diseases on the upper eyelid than the lower eyelid.

In both groups, ST scleral areas were significantly larger than SN scleral areas ($P < 0.001$; Table 5). The values of SN, ST, IN, and IT scleral areas in TAO eyes were 13.39±10.05, 23.92±15.35, 31.64±13.18, and 27.19±13.27 mm², respectively, which were significantly greater compared with those in control eyes. This result implied that the scleral in TAO eyes was more exposed than it was in normal eyes.

Comparison of eyelid contour and symmetry

Figure 7A shows the MPLDs at different angles (from 0° to 180°) in TAO and control eyes. There were significant differences in MPLD at each angle between the 2 groups ($P = 0.008$ at temporal 180°; $P < 0.001$ at other angles). In control eyes, the maximum temporal: nasal ratio of MPLD was present in the horizontal direction, and this ratio went down gradually toward the vertical direction. However, in TAO eyes, the greatest temporal-nasal asymmetry appeared

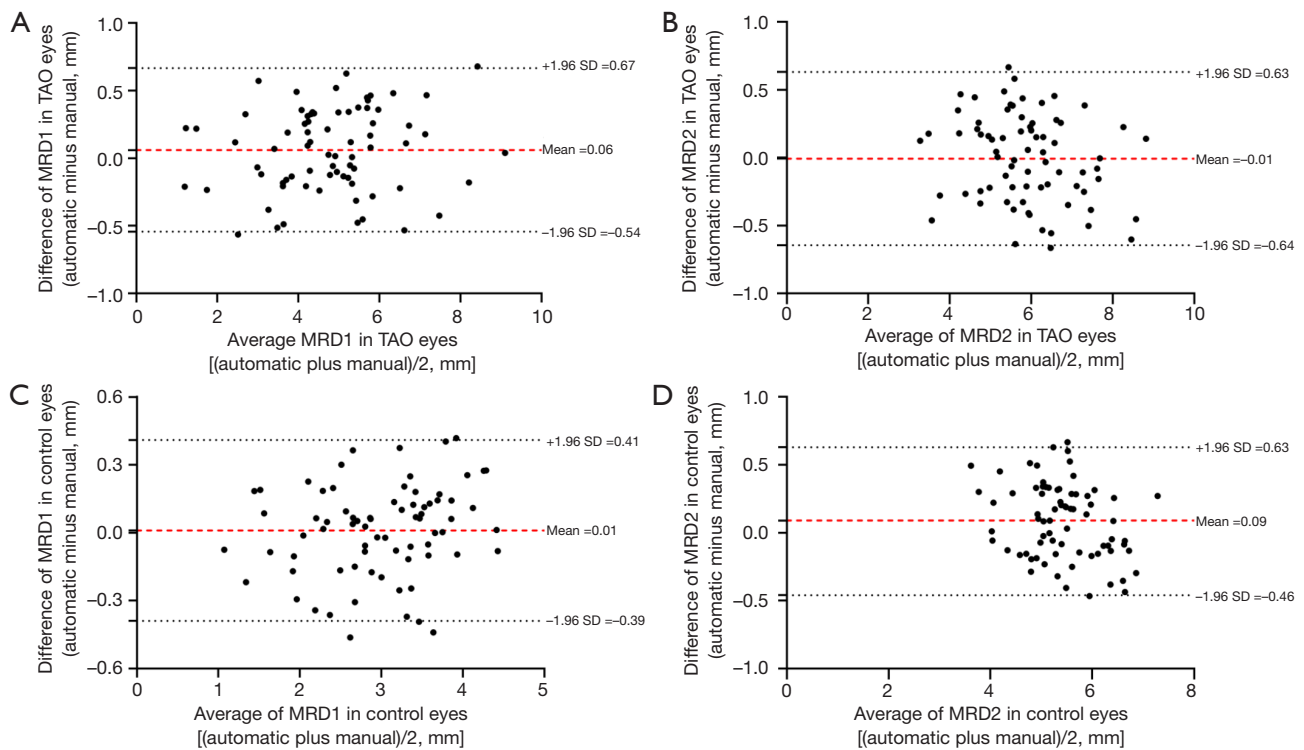


Figure 6 Bland-Altman plots demonstrating excellent agreement between automatic and manual measurements in MRD1 and MRD2. (A) The difference of MRD1 in TAO eyes. (B) The difference of MRD2 in TAO eyes. (C) The difference of MRD1 in control eyes. (D) The difference of MRD2 in control eyes. TAO, thyroid-associated ophthalmopathy; MRD, margin reflex distance; SD, standard deviation.

Table 5 Comparison of eyelid morphological parameters in TAO eyes and control eyes

Parameters	TAO eyes	Control eyes
Palpebral fissure length (mm)	10.72±1.76	8.46±1.03***
MRD1 (mm)	4.82±1.59	2.99±0.81***
MRD2 (mm)	5.89±1.16	5.47±0.73**
Eyelid length (mm)	59.25±5.20	51.77±7.12***
Upper eyelid length (mm)	27.73±4.49	25.42±4.35**
Lower eyelid length (mm)	31.51±4.59	26.34±4.72***
Eyelid retraction distance (mm)		
Upper eyelid retraction distance (mm)	0.52±0.79	0
Lower eyelid retraction distance (mm)	0.46±0.78	0.06±0.21***
Total scleral area (mm ²)	96.14±34.38	56.91±14.97***
Superior-nasal scleral area (mm ²)	13.39±10.05	3.58±2.42***
Superior-temporal scleral area (mm ²)	23.92±15.35	8.44±5.26***
Inferior-nasal scleral area (mm ²)	31.64±13.18	21.82±6.28***
Inferior-temporal scleral area (mm ²)	27.19±13.27	23.07±8.16*

Data are presented as mean ± standard deviation. ***, P<0.001; **, P<0.01; *, P<0.05. TAO, thyroid-associated ophthalmopathy; MRD, margin reflex distance.

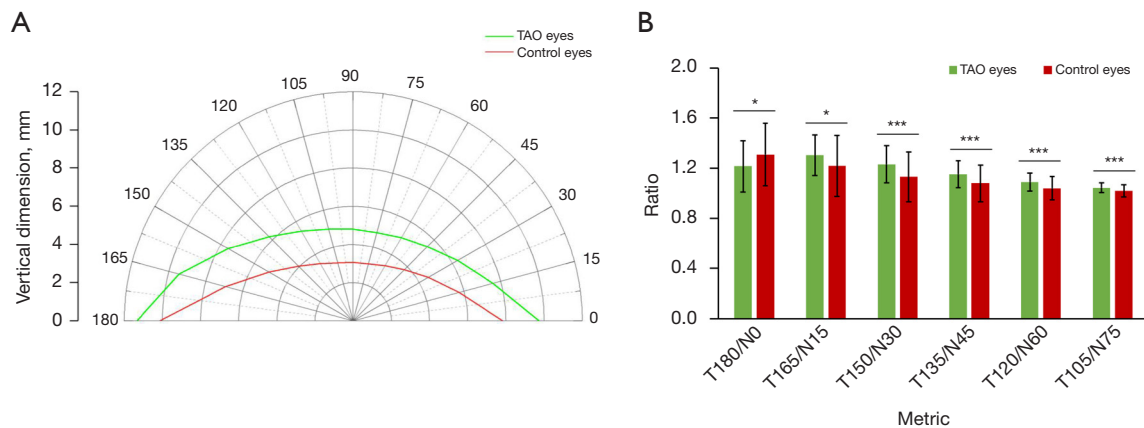


Figure 7 Comparison of the eyelid contour and symmetry in TAO eyes and control eyes. (A) A polar plot showing the eyelid contour of TAO eyes and control eyes according to MPLD. (B) A bar graph displaying the eyelid symmetry of TAO eyes and control eyes according to the temporal: nasal ratio of MPLDs. *, $P < 0.05$; ***, $P < 0.001$. TAO, thyroid-associated ophthalmopathy; MPLD, mid-pupil lid distance.

at 75° from the vertical line, as shown in *Figure 7B*. All temporal: nasal ratios of MPLD were significantly greater in TAO eyes compared with those in control eyes.

Discussion

In this study, we applied an automatic deep learning-based system to measure eyelid features in TAO and control eyes. Eyelid parameters, including PF, MRD1, MRD2, eyelid length (including UEL and LEL), eyelid retraction length (including UER and LER), and scleral areas were all significantly greater in TAO eyes. The findings suggested obvious eyelid retraction in TAO eyes. There were also significant differences in temporal: nasal ratios of MPLD, which indicated a so-called lateral flare sign in TAO.

Measuring eyelid parameters precisely is critical for TAO diagnosis and treatment evaluation. UER is considered an important diagnostic criterion in thyroid eye diseases (6). Its characteristic sign of lateral flare was first described in the early 1950s, which has been explained by different hypotheses, including the lateral extensions of Muller's muscle (28), the stronger lateral horn of the levator palpebrae superioris muscle compared to the medial horn (28,29), and the fibrotic process of the intermuscular septum (30). Although the lateral flare sign has been frequently described in previous studies, the most retracted part of the eyelid varies among individuals. The maximum retraction predominantly appears at 30° to 75° from the midline (12,13,31,32). Traditionally, quantitative analysis of TAO eyelids has been mainly performed according to the

MRD and PF, which reflect the severity of eyelid retraction. However, measuring these 2 features in the midline manually is not sufficient for treatment evaluation. The management of lateral flare and restoration of normal eyelid contour are also key factors in surgical evaluation (33-36).

Some researchers have attempted to quantitatively describe the eyelid contour on digital face images (11-16,31,32,37-42). Cruz *et al.* (37) measured the upper eyelid contour in ptosis and Graves disease and fitted the contours with second-degree polynomial functions. However, the mathematic functions were not clinically useful to characterize lid abnormality. Cruz *et al.* (15) continued their efforts to quantify UER induced voluntarily and by Graves orbitopathy using MRD and nasal and temporal areas. Nevertheless, this quantification method failed to precisely identify the specific eyelid deformation. A few studies applied MPLD to analyze eyelid malposition, but the task was time-consuming since the intersections of radial lines and eyelids still needed to be marked manually (12-14,17,31). The Bezier line with manual adjustment has been applied to quantify the contour of the upper or lower eyelid in TAO eyes (16,32,39), but manual marks or adjustments increase the workload for clinicians. Unlike previous studies (12-17,31,32,37-39), the analysis system proposed in this paper was fully automatic and capable of providing important eyelid features for the clinic, including PF, MRD1, MRD2, and MPLD. When large volumes of facial images needed to be analyzed, this automatic system could measure eyelid parameters without human involvement, saving time and effort.

Boboridis *et al.* (43) reported that the difference between manually measured MRDs could be up to 0.5 mm among clinicians with different levels of experience. In our study, an experienced ophthalmologist measured the MRD to minimize manual measurement error. The ICC was 0.980 for MRD1 and 0.964 for MRD2 in TAO eyes, and 0.967 for MRD1 and 0.932 for MRD2 in control eyes, indicating a high agreement between automatic and manual measurement. In addition, the maximal bias of the 2 repeated automatic measurements for MRDs was 0.003, which showed the high repeatability of the automatic method.

Another strength of our study is the potential application of digital healthcare for patients with TAO. Eyelid change in these patients at an early stage is a dynamic process, and the variations need to be recorded to assess the TAO condition. There are also a proportion of patients who are reluctant to undergo an operation and thus choose botulinum toxin therapy as conservative treatment. Since drug sensitivity varies among individuals, the length of lowering of the eyelid and its duration differ between patients. Real-time measurement of the TAO eyelid is required to determine the number and frequency of drug injections and to adjust the dose. Digital facial images are easy to obtain and preserve at different times. Digital images can also be conveniently transmitted between different medical institutions, which removes the restriction on patients attending the same hospital during follow-up. Finally, the proposed deep learning-based system in our study requires only 3 seconds to present comprehensive and quantitative results. Its efficient and stable performance holds promise for longitudinal clinical evaluation.

There were several limitations to this study. First, the pupil center was located based on the assumption that the cornea and pupil boundaries are perfect circles sharing the same centre. Second, to be processed by our system, the facial images had to show the whole face with the forehead and the chin included. Third, our analysis system could only be applied to 2-dimensional images, and the area measurement did not fully quantify the real value without considering the anteroposterior dimension. Fourth, the sample size was limited in the present study. We are planning to recruit a large sample of patients with TAO that will contain different subsets. Furthermore, more patients with TAO and follow-up records should be included to explore the changes in eyelid contour during TAO development and treatment. Fifth, the automatic eyelid system is a proof of concept, and future work is required

to validate the effectiveness of the proposed method when the lighting conditions, facial pose, skin color, and device manufacturer change.

Conclusions

We proposed an automatic system to quantitatively measure TAO eyelid parameters and compared the results with control eyes. This system allowed quick, comprehensive, and objective measurement of eyelids in facial images, which has potential application prospects in TAO.

Acknowledgments

Funding: This work was jointly supported by grants from the National Natural Science Foundation Regional Innovation and Development Joint Fund (No. U20A20386), the National Natural Science Foundation of China (No. 82000948), the National Key Research and Development Program of China (No. 2019YFC0118400), the Zhejiang Provincial Key Research and Development Plan (No. 2019C03020), the National Natural Science Foundation of China (No. 81870635), and the Clinical Medical Research Center for Eye Diseases of Zhejiang Province (No. 2021E50007).

Footnote

Reporting Checklist: The authors have completed the STARD reporting checklist. Available at <https://qims.amegroups.com/article/view/10.21037/qims-22-551/rc>

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <https://qims.amegroups.com/article/view/10.21037/qims-22-551/coif>). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by the Ethics Committees of the Second Affiliated Hospital of Zhejiang University, School of Medicine (No. 2020-583) and registered with ClinicalTrials.gov (No. NCT04921020). Informed consent was obtained from all adult participants and the guardians of minors.

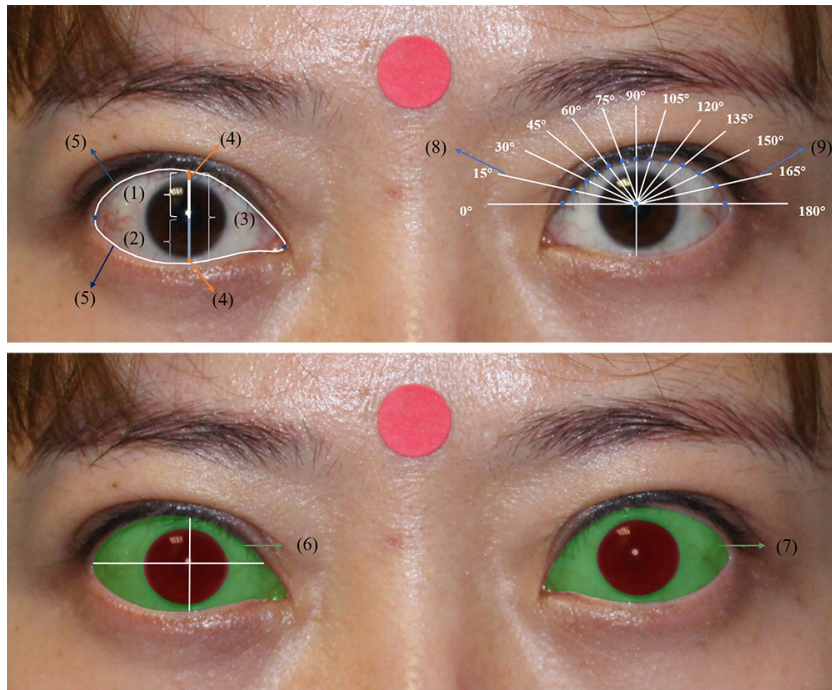
Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

- Bartley GB. The epidemiologic characteristics and clinical course of ophthalmopathy associated with autoimmune thyroid disease in Olmsted County, Minnesota. *Trans Am Ophthalmol Soc* 1994;92:477-588.
- Abraham-Nordling M, Byström K, Törning O, Lantz M, Berg G, Calissendorff J, Nyström HF, Jansson S, Jörneskog G, Karlsson FA, Nyström E, Ohrling H, Orn T, Hallengren B, Wallin G. Incidence of hyperthyroidism in Sweden. *Eur J Endocrinol* 2011;165:899-905.
- Kozaki A, Inoue R, Komoto N, Maeda T, Inoue Y, Inoue T, Ayaki M. Proptosis in dysthyroid ophthalmopathy: a case series of 10,931 Japanese cases. *Optom Vis Sci* 2010;87:200-4.
- Marcocci C, Bartalena L, Bogazzi F, Panicucci M, Pinchera A. Studies on the occurrence of ophthalmopathy in Graves' disease. *Acta Endocrinol (Copenh)* 1989;120:473-8.
- Manji N, Carr-Smith JD, Boelaert K, Allahabadi A, Armitage M, Chatterjee VK, Lazarus JH, Pearce SH, Vaidya B, Gough SC, Franklyn JA. Influences of age, gender, smoking, and family history on autoimmune thyroid disease phenotype. *J Clin Endocrinol Metab* 2006;91:4873-80.
- Day RM. Ocular Manifestations of Thyroid Disease: Current Concepts. *Trans Am Ophthalmol Soc* 1959;57:572-601.
- Versura P, Campos EC. The ocular surface in thyroid diseases. *Curr Opin Allergy Clin Immunol* 2010;10:486-92.
- Coulter I, Frewin S, Krassas GE, Perros P. Psychological implications of Graves' orbitopathy. *Eur J Endocrinol* 2007;157:127-31.
- Putterman AM. Margin reflex distance (MRD) 1, 2, and 3. *Ophthalmic Plast Reconstr Surg* 2012;28:308-11.
- Gorman CA. The measurement of change in Graves' ophthalmopathy. *Thyroid* 1998;8:539-43.
- Edwards DT, Bartley GB, Hodge DO, Gorman CA, Bradley EA. Eyelid position measurement in Graves' ophthalmopathy: reliability of a photographic technique and comparison with a clinical technique. *Ophthalmology* 2004;111:1029-34.
- Ribeiro SF, Milbratz GH, Garcia DM, Fernandes VL, Rocha-Sousa A, Falcão-Reis FM, Cruz AA. Lateral and medial upper eyelid contour abnormalities in graves orbitopathy: the influence of the degree of retraction. *Ophthalmic Plast Reconstr Surg* 2013;29:40-3.
- Kang D, Lee J, Park J, Lee H, Park M, Baek S. Analysis of Lid Contour in Thyroid Eye Disease With Upper and Lower Eyelid Retraction Using Multiple Radial Midpupil Lid Distances. *J Craniofac Surg* 2016;27:134-6.
- Ribeiro SF, Milbratz GH, Garcia DM, Devoto M, Guilherme Neto H, Mörschbacher R, Pereira FJ, Cruz AA. Pre- and postoperative quantitative analysis of contour abnormalities in Graves upper eyelid retraction. *Ophthalmic Plast Reconstr Surg* 2012;28:429-33.
- Cruz AA, Akaishi PM, Coelho RP. Quantitative comparison between upper eyelid retraction induced voluntarily and by Graves orbitopathy. *Ophthalmic Plast Reconstr Surg* 2003;19:212-5.
- Garcia DM, Cruz AAV, Espírito Santo RO, Milbratz GH, Ribeiro SFT. Lower Eyelid Contour in Graves Orbitopathy. *Curr Eye Res* 2019;44:1216-9.
- Gonçalves ACP, Nogueira T, Gonçalves ACA, Silva LD, Matayoshi S, Monteiro MLR. A Comparative Study of Full-Thickness Blepharotomy Versus Transconjunctival Eyelid Lengthening in the Correction of Upper Eyelid Retraction in Graves' Orbitopathy. *Aesthetic Plast Surg* 2018;42:215-23.
- Lin H, Xiao H, Dong L, Teo KB, Zou W, Cai J, Li T. Deep learning for automatic target volume segmentation in radiation therapy: a review. *Quant Imaging Med Surg* 2021;11:4847-58.
- Esteva A, Chou K, Yeung S, Naik N, Madani A, Mottaghi A, Liu Y, Topol E, Dean J, Socher R. Deep learning-enabled medical computer vision. *NPJ Digit Med* 2021;4:5.
- Lou L, Cao J, Wang Y, Gao Z, Jin K, Xu Z, Zhang Q, Huang X, Ye J. Deep learning-based image analysis for automated measurement of eyelid morphology before and after blepharoptosis surgery. *Ann Med* 2021;53:2278-85.
- Bartley GB, Gorman CA. Diagnostic criteria for Graves' ophthalmopathy. *Am J Ophthalmol* 1995;119:792-5.
- Zuo Q, Chen S, Wang Z. R2au-net: Attention recurrent residual convolutional neural network for multimodal

- medical image segmentation. *Secur Commun Netw* 2021;2021:1-10.
23. Liu ZW, Luo P, Wang XG, Tang XO. Deep learning face attributes in the wild. *Proceedings of 2015 IEEE International Conference on Computer Vision; 2015 Dec 11-18; Santiago, CHILE. New York: IEEE; 2015.*
 24. Augusteyn RC, Nankivil D, Mohamed A, Maceo B, Pierre F, Parel JM. Human ocular biometry. *Exp Eye Res* 2012;102:70-5.
 25. Ghassabeh YA. A sufficient condition for the convergence of the mean shift algorithm with gaussian kernel. *J Multivar Anal* 2015;135:1-10.
 26. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-74.
 27. Bujang MA, Baharum N. A simplified guide to determination of sample size requirements for estimating the value of intraclass correlation coefficient: A review. *Arch Orofac Sci* 2017;12:1-11.
 28. Morton AD, Elner VM, Lemke BN, White VA. Lateral extensions of the Müller muscle. *Arch Ophthalmol* 1996;114:1486-8.
 29. Anderson RL, Beard C. The levator aponeurosis. Attachments and their clinical significance. *Arch Ophthalmol* 1977;95:1437-41.
 30. Goodall KL, Jackson A, Leatherbarrow B, Whitehouse RW. Enlargement of the tensor intermuscularis muscle in Graves' ophthalmopathy. A computed tomographic and magnetic resonance imaging study. *Arch Ophthalmol* 1995;113:1286-9.
 31. Milbratz GH, Garcia DM, Guimarães FC, Cruz AA. Multiple radial midpupil lid distances: a simple method for lid contour analysis. *Ophthalmology* 2012;119:625-8.
 32. Equiterio BS, Garcia DM, Cruz AA, Rootman DB, Goldberg RA, Sales-Sanz M, Galindo-Ferreiro A, Diniz S. Lid Flare Measurement with Lateral Midpupil Distances. *Curr Eye Res* 2021;46:1309-13.
 33. Waller RR. Eyelid malpositions in Graves' ophthalmopathy. *Trans Am Ophthalmol Soc* 1982;80:855-930.
 34. MORAN RE. The correction of exophthalmos and levator spasm. *Plast Reconstr Surg* (1946) 1956;18:411-26.
 35. Putterman AM, Urist M. Surgical treatment of upper eyelid retraction. *Arch Ophthalmol* 1972;87:401-5.
 36. Buffam FV, Rootman J. Lid retraction--its diagnosis and treatment. *Int Ophthalmol Clin* 1978;18:75-86.
 37. Cruz AA, Coelho RP, Baccega A, Lucchezi MC, Souza AD, Ruiz EE. Digital image processing measurement of the upper eyelid contour in Graves disease and congenital blepharoptosis. *Ophthalmology* 1998;105:913-8.
 38. Chang EL, Bernardino CR, Rubin PA. Normalization of upper eyelid height and contour after bony decompression in thyroid-related ophthalmopathy: a digital image analysis. *Arch Ophthalmol* 2004;122:1882-5.
 39. Huelin FJ, Sales-Sanz M, Ye-Zhu C, Cruz AAV, Muñoz-Negrete FJ, Rebolleda G. Bézier curves as a total approach to measure the upper lid contour: redefining clinical outcomes in palpebral surgery. *Br J Ophthalmol* 2021. [Epub ahead of print]. pii: bjophthalmol-2021-319666. doi: 10.1136/bjophthalmol-2021-319666.
 40. Lou L, Yang L, Ye X, Zhu Y, Wang S, Sun L, Qian D, Ye J. A Novel Approach for Automated Eyelid Measurements in Blepharoptosis Using Digital Image Analysis. *Curr Eye Res* 2019;44:1075-9.
 41. Cao J, Lou L, You K, Gao Z, Jin K, Shao J, Ye J. A Novel Automatic Morphologic Analysis of Eyelids Based on Deep Learning Methods. *Curr Eye Res* 2021;46:1495-502.
 42. Sales-Sanz M, Huelin FJ, Ye-Zhu C, Cruz AAV, Muñoz-Negrete FJ, Rebolleda G. Müllerotomy with anterior graded Müller muscle disinsertion for Graves upper eyelid retraction: validation of surgical outcomes using Bézier curves. *Graefes Arch Clin Exp Ophthalmol* 2022. [Epub ahead of print]. doi: 10.1007/s00417-022-05853-w.
 43. Boboridis K, Assi A, Indar A, Bunce C, Tyers AG. Repeatability and reproducibility of upper eyelid measurements. *Br J Ophthalmol* 2001;85:99-101.

Cite this article as: Shao J, Huang X, Gao T, Cao J, Wang Y, Zhang Q, Lou L, Ye J. Deep learning-based image analysis of eyelid morphology in thyroid-associated ophthalmopathy. *Quant Imaging Med Surg* 2023;13(3):1592-1604. doi: 10.21037/qims-22-551



$$\begin{aligned}
 (1) \text{ MRD1} &= \text{PN}_{\text{MRD1}} \times R \\
 (2) \text{ MRD2} &= \text{PN}_{\text{MRD2}} \times R \\
 (3) \text{ PF} &= \text{MRD1} + \text{MRD2} \\
 (4) \text{ UER / LER} &= \text{PN}_{\text{UER / LER}} \times R \\
 (5) \text{ UEL / LEL} &= \text{PN}_{\text{UEL / LEL}} \times R \\
 (6) \text{ SN / ST / IN / IT} &= \text{PN}_{\text{SN / ST / IN / IT}} \times R^2 \\
 (7) \text{ SA}_{\text{TOTAL}} &= \text{SN} + \text{ST} + \text{IN} + \text{IT} \\
 (8) \text{ MPLD}_{\text{N0/15/30/45/60/75}} &= \text{PN}_{\text{N0/15/30/45/60/75}} \times R \\
 (9) \text{ MPLD}_{\text{T105/120/135/150/165/180}} &= \text{PN}_{\text{T105/120/135/150/165/180}} \times R
 \end{aligned}$$

Figure S1 Example image in which the measured pixel values were converted to the actual distances. This image has been published with the participant's consent. MRD, margin reflex distance; PF, palpebral fissure; UER, upper eyelid retraction; LER, lower eyelid retraction; UEL, upper eyelid length; LEL, lower eyelid length; SN, superior-nasal scleral area; ST, superior-temporal scleral area; IN, inferior-nasal scleral area; IT, inferior-temporal scleral area; SA_{TOTAL} , total scleral area; MPLD, mid-pupil lid distance; R, the millimeter: pixel ratio.

Table S1 Performance of the eye detection model on 2 different datasets

Parameters	CelebFaces attributes dataset	Dataset consisting of 148 participants
Recall	0.875	0.887
Specificity	0.999	0.992
Accuracy	0.996	0.985

Table S2 The results of 4-fold cross-validation in 1862 participants

Items	Eyelid		Cornea	
	Dice	IoU	Dice	IoU
Fold 1	0.936	0.884	0.946	0.900
Fold 2	0.947	0.903	0.951	0.910
Fold 3	0.946	0.901	0.951	0.910
Fold 4	0.943	0.895	0.953	0.912
Mean	0.943	0.896	0.950	0.908

IoU, Intersection over Union; Dice, dice coefficient.

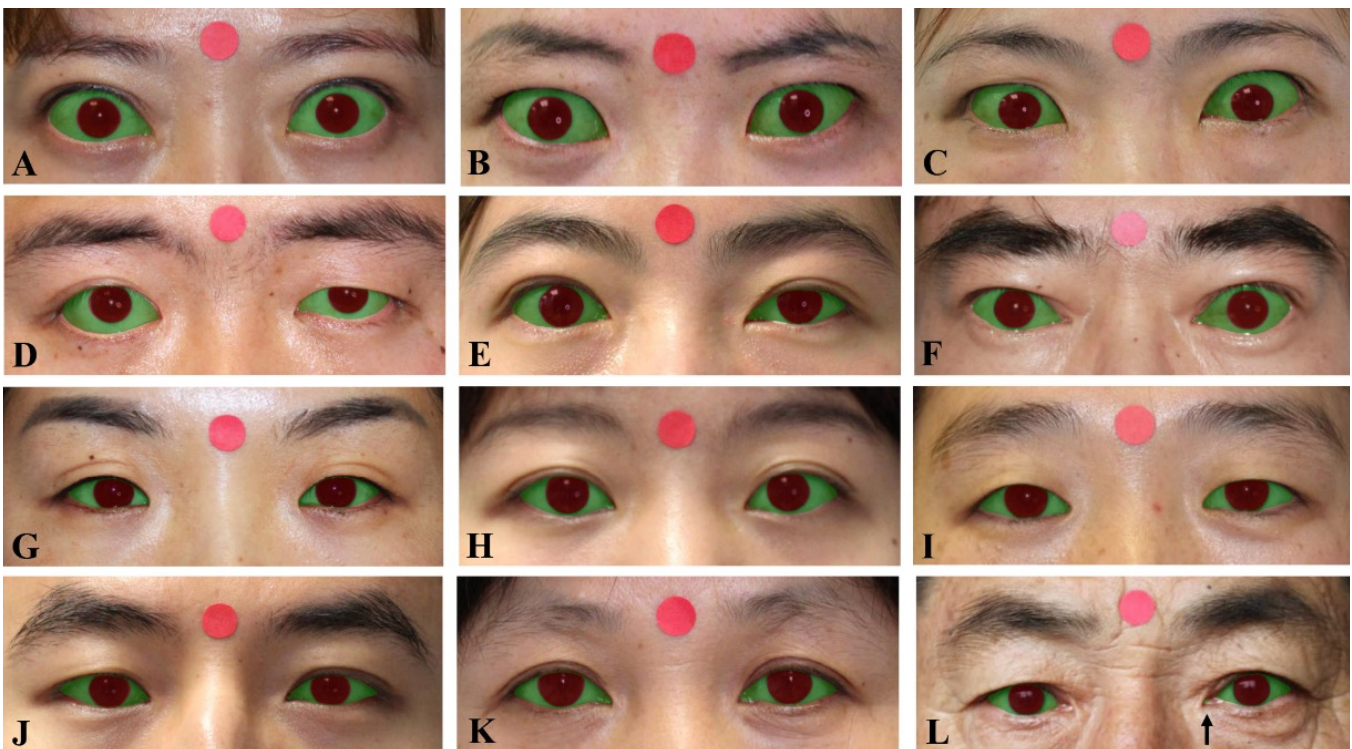


Figure S2 Representative results of segmentation for the eyelid and cornea in patients with TAO and healthy participants. (A-F) Patients with TAO. (G-L) Healthy participants. The black arrow in the last image (L) shows the incomplete segmentation of the eyelid contour in the inner canthus. These images are published with the participants' consent. TAO, thyroid-associated ophthalmopathy.

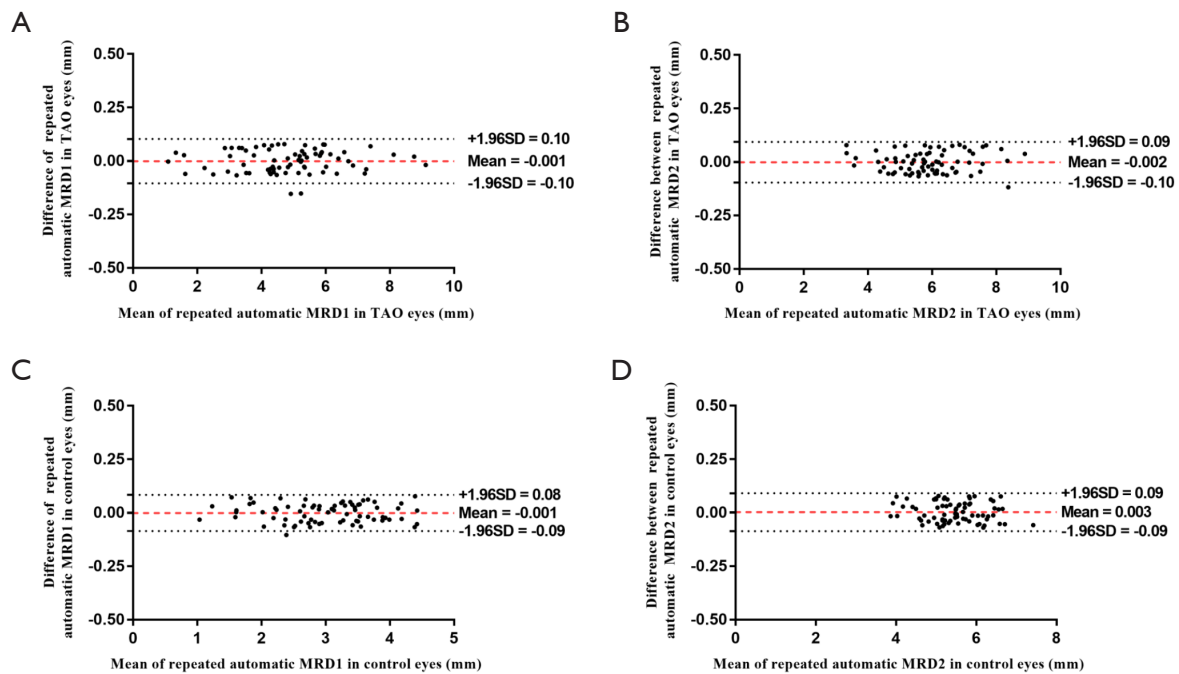


Figure S3 The Bland-Altman plots of 2 repeated automatic MRDs in patients with TAO and control participants. TAO, thyroid-associated ophthalmopathy; MRD, margin reflex distance.