A Fully Automatic Postoperative Appearance Prediction System for Blepharoptosis Surgery with Image-based Deep Learning

Yiming Sun, MD,[1,*] Xingru Huang, PhD,[2,*] Qianni Zhang, PhD,[2] Sang Yeul Lee, MD, PhD,[3] Yaqi Wang, PhD,[4] Kai Jin, MD, PhD,[1] Lixia Lou, MD, PhD,[1] Juan Ye, MD, PhD[1]

Purpose: To automatically predict the postoperative appearance of blepharoptosis surgeries and evaluate the generated images both objectively and subjectively in a clinical setting.

Design: Cross-sectional study.

Participants: This study involved 970 pairs of images of 450 eyes from 362 patients undergoing blepharoptosis surgeries at our oculoplastic clinic between June 2016 and April 2021.

Methods: Preoperative and postoperative facial images were used to train and test the deep learning–based postoperative appearance prediction system (POAP) consisting of 4 modules, including the data processing module (P), ocular detection module (O), analyzing module (A), and prediction module (P).

Main Outcome Measures: The overall and local performance of the system were automatically quantified by the overlap ratio of eyes and by lid contour analysis using midpupil lid distances (MPLDs). Four ophthalmologists and 6 patients were invited to complete a satisfaction scale and a similarity survey with the test set of 75 pairs of images on each scale.

Results: The overall performance (mean overlap ratio) was 0.858 ± 0.082. The corresponding multiple radial MPLDs showed no significant differences between the predictive results and the real samples at any angle ($P > 0.05$). The absolute error between the predicted marginal reflex distance-1 (MRD1) and the actual postoperative MRD1 ranged from 0.013 mm to 1.900 mm (95% within 1 mm, 80% within 0.75 mm). The participating experts and patients were "satisfied" with 268 pairs (35.7%) and "highly satisfied" with most of the outcomes (420 pairs, 56.0%). The similarity score was 9.43 ± 0.79.

Conclusions: The fully automatic deep learning–based method can predict postoperative appearance for blepharoptosis surgery with high accuracy and satisfaction, thus offering the patients with blepharoptosis an opportunity to understand the expected change more clearly and to relieve anxiety. In addition, this system could be used to assist patients in selecting surgeons and the recovery phase of daily living, which may offer guidance for inexperienced surgeons as well. *Ophthalmology Science 2022;2:100169 © 2022 by the American Academy of Ophthalmology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).*

Blepharoptosis, either congenital or acquired, is broadly regarded as the most common disorder of eyelids in the clinic. This drooping of the upper eyelid can have a significant impact on patients not only functionally but also psychologically, including reduced independence and increased appearance-related anxiety and depression.[1-4] Surgery intervention provides an effective treatment option for ptosis patients.[5] Nevertheless, the lack of expected surgical outcome on a patient's own face can increase the rates of anxiety and depression before surgeries or during the recovery, which can even become a barrier to patients' decision-making. This kind of perioperative anxiety and depression can severely affect blepharoptosis patients on the stress response level, pain perception, compliance rate, and postoperative recovery.[6] Thus, there is an urgent need for a method to automatically predict and visualize the expected postoperative appearance to assist surgeons and patients before undergoing blepharoptosis surgery.

Several attempts to simulate postoperative appearance for ptosis patients have been reported. Mawatari and Fukushima[7] used the image processing software (Adobe Photoshop) to prepare the predictive images and surveyed their patients in 2016. However, this method was complicated, time-consuming, and not suitable for elderly patients, and tended to narrow eyelid fissures, possibly resulting from an enlargement of the cornea during image processing. In 2021, they developed a new method that included the utility of a curved hook for 1 eye and the mirror image processing software for the other eye.[8] This approach is simpler but still subjective and time-consuming and only applied to patients with aponeurotic blepharoptosis with fair-to-good levator function.

Recently, deep learning, as a subfield of artificial intelligence, has been increasingly applied in ophthalmology.[9] Several attempts in evaluating eyelid contours have been reported.[10-13] Our previous study has proposed a deep learning−based image analysis to automatically compare preoperative eyelid morphology and postoperative results after blepharoptosis surgery.[14] Remarkably, deep learning neural networks have led to exciting prospects in synthetic images.[15-20] This approach has been applied in orbital decompression for thyroid-associated ophthalmopathy to synthesize the postoperative appearance. However, the synthesized images were of relatively low quality and could not be evaluated objectively. The data are not collected from the clinic and required manually cropping.[21]

In this study, we aim to develop a fully automatic postoperative appearance prediction system (POAP) for patients with varying degrees of blepharoptosis, including 4 modules: the data processing module (P), ocular detection module (O), analyzing module (A), and prediction module (P). Also, we introduce the concept of overlap ratio of eyes for a general assessment of system accuracy and conducted surveys for subjective evaluation in a clinical setting.

Methods

Subjects

This study involved patients undergoing varying types of blepharoptosis operations at our oculoplastic clinic between June 2016 and April 2021. The patients were in stable recovery, with a mean follow-up period of 115.7 days (range, 6−1409 days), whose eyes presented with a symmetric and natural eyelid contour without significant redness and swelling, and the upper eyelid position above the papillary margin. The surgeries were performed by a senior surgeon (J.Y.) who had more than 15 years of experience in oculoplastic surgery. Cases with previous lid surgery, orbital surgery, or any abnormality or surgery that could affect eyelid shape and function were excluded. Preoperative and postoperative images were taken in the primary position of gaze using a digital camera (Canon 500D with 100-mm macro lens, Canon Corporation). A 10-mm diameter round sticker was attached to the middle of the forehead for distance reference.

The Ethics Committee of the Second Affiliated Hospital of Zhejiang University, College of Medicine, approved this study. Informed consent was obtained from patients aged ≥ 18 years and guardians of those aged <18 years. All methods adhered to the tenets of the Declaration of Helsinki.

Development of the Automatic Postoperative Appearance Prediction System

We aimed to develop a fully automatic system appliable in clinical practice that is able to return instantly predicted outcomes based on an input image of a patient's face. The automatic prediction system was designed by integrating 4 main functional modules: the ocular detection module, analyzing module, data processing module, and prediction module (Fig 1). We set aside 75 pairs of images (8%) as a test set among a total of 970 pairs of preoperative and postoperative images collected from 450 eyes of 362 patients, and the remaining 895 pairs were used for training and validation of the model.

Step 1: The first step involved automated recognition of the eyelid and corneal limbus. A preprocessed pair of images was input into the ocular detection module (O) that was developed in our previous work (Fig 1, the ocular detection module).[14,22,23] In the first part, the eyes and periocular area were located, and in the second part, the located eyelid and corneal limbus were segmented. Then, the segmented masks were output for the next step.

Step 2a: Next, the dimensions of the eyelid were automatically measured based on the reference sticker. These masks were sent into the analyzing module (A) developed in our previous work (Fig 1, the analyzing module).[14,22,23] First, the eyes were rotated parallel, including locating and connecting the center of the pupil and rotating the image to make the connected line horizontal. Next, the 10-mm diameter round sticker attached to the middle of the forehead was detected and segmented. Then, to determine the actual size of the image, the numbers of pixels in the horizontal and vertical directions of the sticker region were measured, and the lengths (mm) of a single horizontal and vertical pixel were obtained, respectively. Next, the eyelids in the masks were quantified according to the method demonstrated afterward.

Step 2b: At the same time, the eyelid and surrounding region were extracted from the full original image based on the landmarks. The masks in step 1 were sent into the data processing module (P) where they were replaced and located back on the full original image (Fig 1, the data processing module). Then, the medial canthi were detected and anchored. The full original image with masks and anchor points was cropped into a piece of strip image where the anchor points stayed horizontally at the trisection and vertically at the lower trisection of each strip image.

Step 3: Next, the predicted eyelid was generated. Strip images were obtained after paired preoperative and postoperative training data received the previous process. The strip images were sent into the prediction module (P) (Fig 1, the prediction module). We used a Pix2Pix conditional generative adversarial network (GAN) architecture[16] that included 2 main structures: a generator for producing predicted images from original images, with R2Unet as the backbone[24] and a discriminator that aimed to distinguish generated images from real images. The goal of training is to lead the generator to generate convincing forecasting images as close to their real counterparts as possible.

In the training process, we used the Adam optimizer, with an initial learning rate of 0.0001. The number of training iterations (epoch) was set to 200. The upper and lower areas of the image were complemented into 512×512. The overall loss of training is $\lambda \mathscr{L}_{Generator} + \mathscr{L}_{discriminator}$. For the generator, the $L1$ loss is used for the sake of clarity. We set $\lambda = 100$ and used the discriminator structure in the original article.[16] During training, data augmentation methods were applied to mitigate the impact of insufficient data, including adding random noise, random scaling, and elastic transformation.

The testing images were sent into steps 4a to 4c as follows:

Step 4a: The test preoperative images were input into the described 3 steps to generate the predicted eyelid. The test images were fed into the modules in steps 1 and 2 to obtain cropped strip images. The cropped test strip images were then sent to the trained GAN model in step 3 to generate predicted pieces (Figs 1 and 2).

Step 4b: The dimensions of the predicted eyelid were automatically measured through step 2a. The generated pieces were passed to the ocular detection module and analyzing module (steps 1 and 2a) again to complete the automatic measurement of the predicted eyes.

Step 4c: The full-face predicted postoperative appearance was achieved in the last step. The generated pieces in step 3 were eventually pasted back to the original preoperative images and underwent color calibration for the final output.
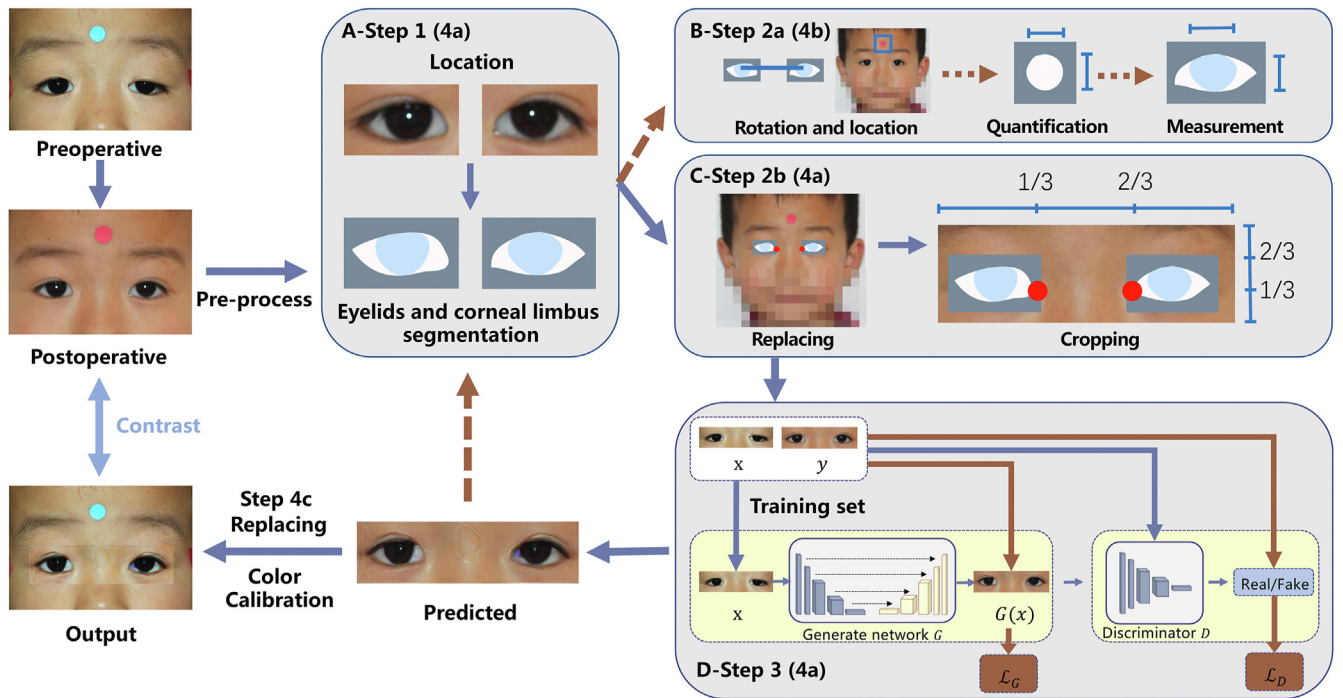
Figure 1. Flowchart of the fully automatic postoperative appearance prediction system (POAP) for ptosis. **A,** Ocular detection module (**steps 1, 4a**): A pair of preoperative and postoperative training images were input into an ocular detection module developed in our previous work. **B,** Analyzing module (**steps 2a, 4b**): The mask output in the ocular detection module was used to define the rotation angles to make the eyes parallel, and then the round sticker was detected and segmented in pixels to quantify the eyelids. **C,** Data processing module (**steps 2b, 4a**): The mask output in the ocular detection module was replaced back on the original images, and the medial canthi were detected and anchored. Then the original images were cropped into strip images where the anchor points stayed horizontally at the trisection and vertically at the lower trisection on each image. **D,** Prediction module (**steps 3, 4a**): The preprocessed strip images went through a generator and a discriminator as the training process. Then the test images, having gone through the ocular detection module and the analyzing module, were fed into the trained prediction model and output as predicted pieces. These generated pieces were then passed to the ocular detection module and the analyzing module again to complete the automatic measurement of the predicted eyes, while at the same time, the generated pieces were eventually pasted back to the original preoperative images and underwent color calibration for the final output (**step 4c**).

Evaluation of the Predicted Outcome

Assessment of the Overall Prediction Performance: Overlap Ratio. To evaluate the overall performance, we introduced the overlap ratio. The overlap ratio, ranging from 0 to 1, was to evaluate the overlap area between the predictive eyes and the real postoperative eyes, specifically defined as the ratio of the intersecting area and the union area between the predicted eyes and the postoperative ground truth. Thus, the higher the overlap ratio is, the more similar the images are and the more realistic the prediction is. The overlap ratio for all eyes was automatically measured (Fig 3C). The edges of the eyelids were automatically
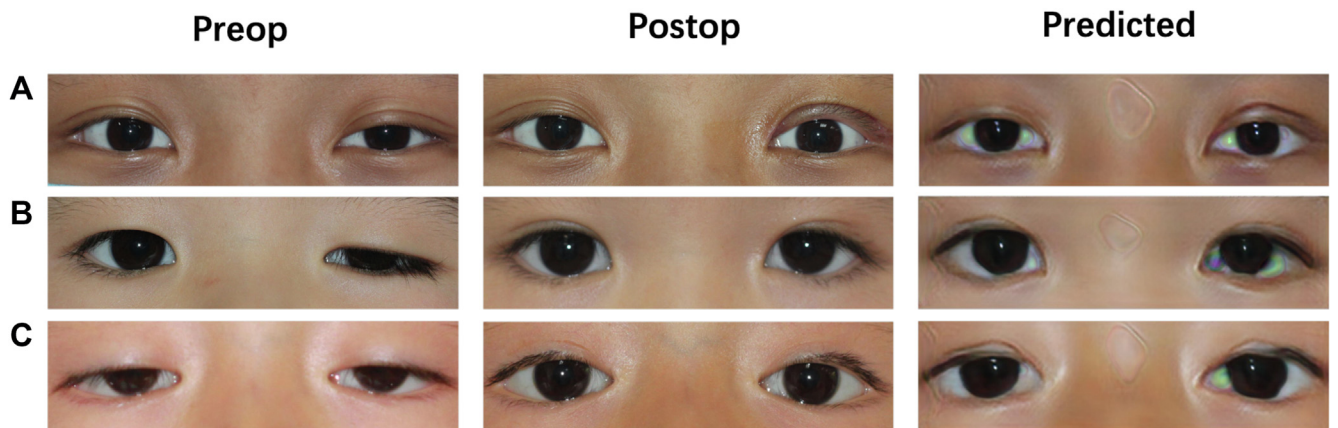

Figure 2. Samples of input preoperative images, ground truth postoperative images, and generated images. **A,** Sample of an 11-year-old patient with unilateral blepharoptosis undergoing levator resection. **B,** Sample of a 7-year-old patient with unilateral blepharoptosis undergoing frontalis suspension using autogenous fascia lata. **C,** Sample of a 4-year-old patient with bilateral blepharoptosis undergoing frontalis suspension using a silicon rod.
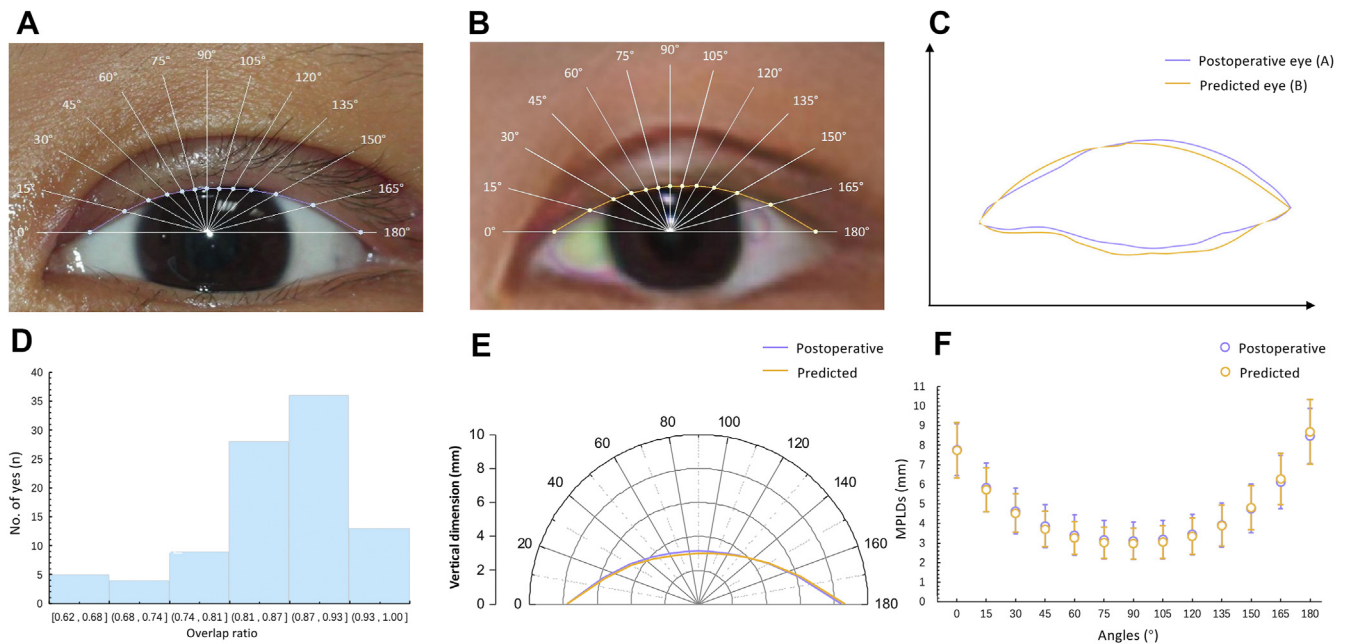
Figure 3. Demonstration of the objective assessment of the predicted performance. **A, B,** Example of a pair of postoperative and predicted images. **C,** Demonstration of the general performance of **A, B** (overlap ratio): the ratio of the intersecting area over the union area of the predicted eye region (orange) and the real postoperative eye region (purple). The overlap ratio of the sample was 0.95. **D,** The distribution of overlap ratio in the test set. **E,** Demonstration of the mean local performance in the test set (midpupil lid distances [MPLDs]): the distances between the corneal light reflex and the upper eyelid margin at different angles. **F,** The mean MPLDs of postoperative eyes and predicted eyes in the test set.

extracted and placed in the same coordinate system, with an alignment of the medial canthi. Then, the overlap ratio was automatically calculated.

Assessment of Local Prediction Performance: Lid Contour Analysis. The lid contour analysis was applied to automatically evaluate and quantify local prediction performance (Fig 3E), which mainly involved the multiple radial midpupil lid distances (MPLDs), defined as the distance between the corneal light reflex and the upper eyelid margin as the primary clinical measurement.[25] The algorithm was developed on the basis of our former work,[14,22,23] referring to the technique described by Milbratz et al[13] and other previous studies,[26-28] and consisted of the following steps. The pupil center was detected and set as the origin, and then 13 radial lines at 15° intervals from 0° to 180° were set. The intersections of the lines on the lid margin edge were marked automatically, and MPLDs (in pixels) at different angles were measured. Then, the MPLDs in pixels were converted into millimeters according to the horizontal and vertical ratio with the 10-mm diameter round sticker. A pairwise t test of 13 pairs MPLDs (in millimeters) at each angle was used for comparisons between predicted eyes and ground truth eyes, as well as predicted eyes and preoperative eyes. All P values presented were 2 sided and statistically significant if <0.05.

Subjective Evaluation of the Predicted Images: A Satisfaction and Similarity Survey. To evaluate the predicted performance subjectively, we invited 4 experts specializing in ophthalmology and 6 patients who had undergone blepharoptosis surgery to complete satisfaction and similarity surveys. In the satisfaction survey, ophthalmologists and patients were asked to simulate clinical situations where the preoperative image was input into the system, and the predicted image was output as a result. They rated their overall satisfaction of the predicted outcome of 75 pairs of preoperative images and corresponding predicted images on a 5-point scale response of "1 = highly not satisfied," "2 = not satisfied,"

"3 = neutral," "4 = satisfied," and "5 = highly satisfied." In the similarity survey, ophthalmologists and patients were asked to compare the predicted image with the actual postoperative image. They rated the visual similarity of 75 paired postoperative images and predicted images ranging from "0 = highly not similar" to "10 = highly similar." At the end of the survey, several additional questions (5-point) were asked: "To what extent are you (ophthalmologists) willing to apply this technique in the clinic?" "To what extent are you (patients) anxious or hesitated due to the unknown postoperative appearance?" "To what extent are you (patients) relieved with the help of the technique?"

Results

Patient Characteristics

A total of 1157 images (566 preoperative, 691 postoperative) from 450 eyes (362 patients, aged 0−77 years) were collected during the study period ranging from 2016 to 2021. Cases included 213 eyes (47.3%) undergoing frontalis suspension using a silicon rod, 121 eyes (26.9%) undergoing frontalis suspension using autogenous fascia lata, 108 eyes (24.0%) undergoing levator resection, and 8 eyes (1.8%) undergoing levator aponeurosis repair. The median number of days between the preoperative and postoperative photographs was 16 (interquartile range, 13−135 days). The test sample consisted of 150 images from 95 eyes (75 patients) with a median number of follow-ups of 16 days (interquartile range, 13−134 days). Table 1 shows the clinical characteristics of the eyes included in the study.

Table 1. Clinical Characteristics for Eyes and Patients Included in the Study

	Training	Test	Total
No. of patients	287	75	362
No. of eyes	355	95	450
No. of pairs of photographs	895	75	970
Age (yrs)			
Mean ± SD	8 ± 11.8	7.2 ± 7.2	7.9 ± 11
Range	0−77	1−45	0−77
Female gender (%)	36.2	33.3	35.6
Surgery type (No. of eyes)			
Frontalis suspension using a silicon rod	182	31	213
Frontalis suspension using autogenous Fascia lata	84	37	121
levator resection	81	27	108
Levator aponeurosis repair	8	0	8
Follow-up (days)			
Mean	119	102.9	115.7
Range	6−1409	6−1248	6−1409

SD = standard deviation.

Evaluation of the Predicted Results

Assessment of the Overall Prediction Performance: Overlap Ratio. The mean overlap ratio of all the predictive images and the corresponding ground truth was 0.858 ± 0.082. Figure 3 demonstrates an illustration of overlap ratio and the numerical distribution of the overlap ratio. The overlap ratio ranged from 0.64 to 0.97, with most in the region of 0.87−0.93 (36 eyes of 95, 37.9%) and 28 eyes in the region of 0.87−0.93 (29.5%) (Fig 3D). Figure 2 shows 3 pairs of samples, including the input and output and the real images.

Assessment of Local Prediction Performance: Lid Contour Analysis. Table 2 lists the automatically acquired MPLDs from 0° to 180° in all images (postoperative and predicted). From the 90° midline, the MPLDs of the generated eyelid and the actual eyelid showed a tendency to increase in both the nasal and temporal sectors (Fig 3F). There were no significant differences in any corresponding MPLDs between the generated postoperative eyelid and the real postoperative eyelid at any angle ($P > 0.05$) (Table 2). The average MPLD at 90°, or the marginal reflex distance-1 (MRD1), was 2.973 ± 0.802 mm for generated postoperative eyes, 3.117 ± 0.962 mm ($P = 0.128$) for ground truth eyes, and 0.157 ± 1.280 mm ($P<0.001$) for preoperative eyes. Figure 3E and F show the mean postoperative and predicted MPLDs at different angles, respectively, in the polar coordinate and rectangular coordinate. Furthermore, the absolute error between the predicted MRD1 and the actual postoperative MRD1 ranged from 0.013 mm to 1.900 mm. More specifically, there were 95% (90 pairs of 95) of eyes with an absolute error within 1 mm and 80% (76 pairs of 95) within 0.75 mm.

Subjective Evaluation of the Predicted Images: A Satisfaction and Similarity Survey. Ophthalmologists and patients rated their overall satisfaction as "highly satisfied"

in more than half of the cases (420 pairs of 750, 56.0%), and 268 pairs (35.7%) of images were rated as "satisfied" (Table 3). The lower-rated "neutral" and "not satisfied" cases (60 pairs, 8.0%; 2 pairs, 0.3%) may be due to the comment that the edge of the predicted eyelid was not natural enough compared with the real postoperative image, according to the free-text comments provided by some patients (n = 2). Two ophthalmologists involved were highly willing to apply this technique to the clinic, and the other 2 ophthalmologists rated "willing to apply." All the patients surveyed were highly anxious or hesitated because of the unknown postoperative appearance and felt relieved with the help of the technique ("highly relieved," n = 4; "relieved," n = 2). The mean score of the 10-point similarity scale is 9.43, ranging from 7 to 10, with a standard deviation of 0.79.

Discussion

The study presents a deep learning method designed to automatically generate postoperative images based on preoperative images before blepharoptosis surgeries with factual subjective evaluation and an objective evaluation method to quantify and assess the predicted results. In particular, the synthesized images highly resembled the real images. To our knowledge, this is a system applicable in clinical settings, which could automatically predict postoperative appearance for patients with ptosis.

The GAN technique, a new type of deep learning technique, has been gaining more attention in the medical field. In previous studies, there has been a prediction for a soft tissue deformity after an osteotomy,[29,30] a probabilistic finite element model for the prediction of postoperative facial images for orthognathic surgery,[31] and a similar technique for orbital plastic surgery.[32] These studies require the acquisition of whole facial data collected through computed tomography or magnetic resonance imaging. Abe et al[33] introduced an image-based method

Table 2. Assessment of Local Prediction Performance: Comparison of the Midpupil Lid Distances from 0°to 180°

Angle (°)	Postoperative Ground Truth Mean ± SD (mm)	Predicted Mean ± SD (mm)	P Value
0	7.769 ± 1.323	7.734 ± 1.405	0.772
15	5.836 ± 1.237	5.722 ± 1.115	0.331
30	4.638 ± 1.176	4.532 ± 0.992	0.343
45	3.858 ± 1.099	3.723 ± 0.904	0.198
60	3.407 ± 1.025	3.256 ± 0.839	0.126
75	3.166 ± 0.978	3.018 ± 0.806	0.118
90	3.117 ± 0.962	2.973 ± 0.802	0.128
105	3.180 ± 0.978	3.053 ± 0.844	0.183
120	3.442 ± 1.034	3.352 ± 0.925	0.371
135	3.923 ± 1.125	3.894 ± 1.036	0.783
150	4.766 ± 1.244	4.807 ± 1.134	0.717
165	6.124 ± 1.354	6.272 ± 1.304	0.232
180	8.468 ± 1.416	8.682 ± 1.638	0.125

SD = standard deviation.

Table 3. Five-Point Satisfaction Scale (75 Pairs) and 10-Point Similarity Survey (75 Pairs)

	Highly Unsatisfied n (%)	Unsatisfied n (%)	Neutral n (%)	Satisfied n (%)	Highly Satisfied n (%)	Similarity (10-Point) Mean ± SD
Subjects						
Ophthalmologist 1	-	-	-	20 (26.7)	55 (73.3)	9.45 ± 0.57
Ophthalmologist 2	-	-	20 (26.7)	29 (38.7)	26 (34.7)	8.76 ± 0.86
Ophthalmologist 3	-	1 (1.3)	15 (20.0)	42 (56.0)	17 (22.7)	8.31 ± 0.98
Ophthalmologist 4	-	1 (1.3)	5 (6.7)	40 (53.3)	29 (38.7)	9.12 ± 0.57
Patient 1	-	-	3 (4.0)	5 (6.7)	67 (89.3)	9.99 ± 0.11
Patient 2	-	-	-	41 (54.7)	34 (45.3)	9.49 ± 0.57
Patient 3	-	-	10 (13.3)	-	65 (86.7)	9.73 ± 0.44
Patient 4	-	-	7 (9.3)	33 (44.0)	35 (46.7)	9.97 ± 0.23
Patient 5	-	-	-	22 (29.3)	53 (70.7)	9.80 ± 0.40
Patient 6	-	-	-	36 (48.0)	39 (52.0)	9.86 ± 0.66
Ophthalmologists (n=4)	-	2 (0.7)	40 (13.3)	131 (43.7)	127 (42.3)	8.91 ± 0.88
Patients (n=6)	-	-	20 (4.4)	137 (30.4)	293 (65.1)	9.78 ± 0.48
Total (n=10)	-	2 (0.3)	60 (8.0)	268 (35.7)	420 (56.0)	9.43 ± 0.79

SD = standard deviation.

for postoperative appearance prediction. Their method required an examiner to manually select racial feature points. Recently, a postoperative prediction method for thyroid eye disease was introduced, which was automated during training and generating processes, but still needs manual cropping for the region of interest. In that study, images were collected from Google and the synthesized images were 64 × 64 pixels, which is a relatively low resolution. By contrast, our system was fully automatic, which was trained on real clinical images with higher resolution.

Crucial to the assessment of any predicting methods is the realness of the synthesized image. In the previous studies, Mawatari and Fukushima[7] surveyed their patients using questionnaires in 2016[7] and then measured MRD1, eyebrow height, and pretarsal skin height on predictive and real images in 2021.[8] However, the measurement was manually obtained and based on the assumption that the corneal diameter was 11 mm, which can be relatively subjective. Yoo et al[21] validated their model using a VGG-16 classifier, which was more related to the variability of spatial and structural features than to the realism of the synthesized images. In our work, we introduced several objective methods to automatically quantify and evaluate the predictive outcome both in general and in detail, including using a 10-mm diameter round sticker for horizontal and vertical distance reference. We also conducted a satisfaction and similarity survey among doctors and patients to subjectively evaluate the predicted outcome.

In general, the predictive images were highly consistent with actual surgical outcomes, according to the overlap ratio we introduced. Our results showed that the mean overlap ratio was 0.858, with most samples in the 0.87−0.93 region (36 eyes of 95, 37.9%), demonstrating that the synthesized postoperative images were highly consistent with the ground truth images.

We analyzed the lid contour to quantify the local performance in practice, which identified with the general performance. We did not find any quantifiable difference in eyelid contour between generated eyelids and real postoperative eyelids ($P>0.05$), which indicates the high prediction accuracy of the method in terms of local performance. Also, the absolute error of the predicted eyelid and the ground truth eyelid presented a high probability of the accurate prediction (95% within 1 mm; 80% within 0.75 mm).

The satisfaction and similarity surveys showed high satisfaction and visual similarity (9.84 of 10) according to the ophthalmologists and patients involved in the study. Nevertheless, some of the predicted images were slightly unnatural around the eyelid crease as reported by some participants, which can be improved in the future. In general, everyone included in the survey supported the technique to be applied in the clinic.

The promising result was partially due to the effort (Fig 1, the data preprocessing module) before the prediction module. Blepharoptosis is the downward displacement of the upper eyelid; thus, the main changes after ptosis surgery focus on the upper eyelid. To ensure the deep learning model in the prediction module mainly learns in the focus area, we anchored the medial canthi and cropped the initial image into the piece as we described previously, which may improve the prediction performance.

The fully automatic method is a step toward clinical implementation that can predict postoperative appearance using images. Compared with previous methods for predicting ptosis surgery outcome,[7,8] our method is free from the use of a curved hook and manual operations, which makes it simpler to apply in clinical practice. This automatic method allows us to include broader kinds of blepharoptosis patients. In this study, we included patients of different levator functions and ages who had undergone varying types of blepharoptosis surgery. In the future, we intend to predict appearance during the different recovery phases to promote doctor−patient communication and alleviate anxiety. Furthermore, trained on various postoperative results (including suboptimal results), the system may assist surgeons in detecting patients who might not be good candidates for repair. Also, our method has the

potential to guide inexperienced surgeons and to help them set up their own prediction style, which may enable patients to choose their favorite.

Study Limitations

There are several limitations in this study. First, this study obtained all images using high-resolution digital cameras. Out-of-focus images were excluded in these photographs, and this may result in less accuracy when poorer-quality photographs are used. This could be remediable by retraining the model on lower-quality images. Second, the prediction module using pix2pix GAN may not be readily interpretable. For this reason, the predictive outcome should be further reviewed by medical professionals. Third, the lid crease prediction should be optimized in the future, which may lead to quantification of the predicted lid crease. Fourth, the system was developed using one surgeon's panel, which would be trained and tested for different surgeons in our future plans. Finally, our model will further benefit from external validation in various populations, because the racial difference in eyelid anatomy exists.

Conclusions

In this work, we developed a fully automatic deep learning system (POAP) to predict the postoperative appearance for blepharoptosis surgery that performed well both objectively and subjectively. Our system offers patients with blepharoptosis an opportunity to understand the expected change more clearly and to help them relieve anxiety. In addition, this system shows the potential to be further developed to help patients select surgeons and to alleviate anxiety in different recovering phases, and can be used as an adjunctive tool to guide inexperienced oculoplastic surgeons in a clinical setting.

Footnotes and Disclosures

[1] Department of Ophthalmology, The Second Affiliated Hospital of Zhejiang University, School of Medicine, Hangzhou, China.

[2] School of Electronic Engineering and Computer Science, Queen Mary University of London, London, United Kingdom.

[3] Yonsei University, Seoul, Korea.

[4] College of Media Engineering, Communication University of Zhejiang, Hangzhou, China.

*Y.S. and X.H. contributed equally to this work.

Disclosure(s):

All authors have completed and submitted the ICMJE disclosures form.

The author(s) have no proprietary or commercial interest in any materials discussed in this article.

HUMAN SUBJECTS: Human subjects were included in this study. The Ethics Committee of the Second Affiliated Hospital of Zhejiang University, College of Medicine approved this study. Informed consent was obtained from patients aged $\geq$ 18 years and guardians of those aged < 18 years. All methods adhered to the tenets of the Declaration of Helsinki.

No animal subjects were used in this study.

Author Contributions:

Conception and design: Sun, Huang, Lou, Ye

Data collection: Sun, Huang, Zhang, Lee, Wang, Jin, Lou, Ye

Analysis and interpretation: Sun, Lou, Ye

Obtained funding: N/A; No additional funding was provided.

Overall responsibility: Sun, Huang, Zhang, Lee, Wang, Jin, Lou, Ye

Abbreviations and Acronyms:
GAN = generative adversarial network; **MPLD** = midpupil lid distance; **MRD1** = marginal reflex distance-1; **POAP** = postoperative appearance prediction system.

Keywords:
Blepharoptosis, Deep learning, Postoperative prediction.

Correspondence:
Juan Ye, MD, PhD, Department of Ophthalmology, The Second Affiliated Hospital of Zhejiang University, School of Medicine, Jiefang Road 88, Hangzhou, 310009, China. E-mail: yejuan@zju.edu.cn; and Lixia Lou, MD, PhD, Department of Ophthalmology, The Second Affiliated Hospital of Zhejiang University, School of Medicine, Jiefang Road 88, Hangzhou, 310009, China. E-mail: loulixia110@zju.edu.cn.

References

1. Alniemi ST, Pang NK, Woog JJ, Bradley EA. Comparison of automated and manual perimetry in patients with blepharoptosis. *Ophthalmic Plast Reconstr Surg.* 2013;29:361−363.
2. Bacharach J, Lee WW, Harrison AR, Freddo TF. A review of acquired blepharoptosis: prevalence, diagnosis, and current treatment options. *Eye (Lond).* 2021;35:2468−2481.
3. Richards HS, Jenkinson E, Rumsey N, et al. The psychological well-being and appearance concerns of patients presenting with ptosis. *Eye (Lond).* 2014;28:296−302.
4. James H, Jenkinson E, Harrad R, et al. Members of Appearance Research Collaboration. Appearance concerns in ophthalmic patients. *Eye (Lond).* 2011;25:1039−1044.
5. Richards HS, Jenkinson E, Rumsey N, Harrad RA. Pre-operative experiences and post-operative benefits of ptosis surgery: a qualitative study. *Orbit.* 2017;36:147−153.
6. Cai H, Xu S. A randomized trial of psychological intervention on perioperative anxiety and depression of patients with severe blepharoptosis undergoing autologous fascia lata

frontal muscle suspension. *Ann Palliat Med.* 2021;10:3185−3193.

7. Mawatari Y, Fukushima M. Predictive images of postoperative levator resection outcome using image processing software. *Clin Ophthalmol.* 2016;10:1877−1881.

8. Mawatari Y, Kawaji T, Kakizaki H, et al. Usefulness of mirror image processing software for creating images of expected appearance after blepharoptosis surgery. *Int Ophthalmol.* 2021;41:1151−1156.

9. Ting DSW, Lin H, Ruamviboonsuk P, et al. Artificial intelligence, the internet of things, and virtual clinics: ophthalmology at the digital translation forefront. *Lancet Digit Health.* 2020;2:e8−e9.

10. Danesh J, Ugradar S, Goldberg R, Rootman DB. A novel technique for the measurement of eyelid contour to compare outcomes following Muller's muscle-conjunctival resection and external levator resection surgery. *Eye.* 2018;32:1493−1497.

11. Chun YS, Park HH, Park IK, et al. Topographic analysis of eyelid position using digital image processing software. *Acta Ophthalmol.* 2017;95:E625−E632.

12. Mocan MC, Ilhan H, Gurcay H, et al. The expression and comparison of healthy and ptotic upper eyelid contours using a polynomial mathematical function. *Curr Eye Res.* 2014;39:553−560.

13. Milbratz GH, Garcia DM, Guimaraes FC, Cruz AA. Multiple radial midpupil lid distances: a simple method for lid contour analysis. *Ophthalmology.* 2012;119:625−628.

14. Lou L, Cao J, Wang Y, et al. Deep learning-based image analysis for automated measurement of eyelid morphology before and after blepharoptosis surgery. *Ann Med.* 2021;53:2278−2285.

15. Iqbal T, Ali H. Generative Adversarial Network for Medical Images (MI-GAN). *J Med Syst.* 2018;42:231.

16. Isola P, Zhu JY, Zhou TH, Efros AA. Image-to-image translation with conditional adversarial networks. *Proc Cvpr Ieee.* 2017:5967−5976.

17. Yang QS, Yan PK, Zhang YB, et al. Low-dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss. *Ieee T Med Imaging.* 2018;37:1348−1357.

18. Yi X, Babyn P. Sharpness-aware low-dose CT denoising using conditional generative adversarial network. *J Digit Imaging.* 2018;31:655−669.

19. You CY, Yang QS, Shan HM, et al. Structurally-sensitive multi-scale deep neural network for low-dose CT denoising. *Ieee Access.* 2018;6:41839−41855.

20. Abdelmotaal H, Abdou AA, Omar AF, et al. Pix2pix conditional generative adversarial networks for Scheimpflug camera color-coded corneal tomography image generation. *Transl Vis Sci Techn.* 2021;10:21.

21. Yoo TK, Choi JY, Kim HK. A generative adversarial network approach to predicting postoperative appearance after orbital decompression surgery for thyroid eye disease. *Comput Biol Med.* 2020;118:103628.

22. Lou L, Yang L, Ye X, et al. A novel approach for automated eyelid measurements in blepharoptosis using digital image analysis. *Curr Eye Res.* 2019;44:1075−1079.

23. Cao J, Lou L, You K, et al. A novel automatic morphologic analysis of eyelids based on deep learning methods. *Curr Eye Res.* 2021;46:1495−1502.

24. Alom MZ, Yakopcic C, Taha TM, Asari VK. Nuclei segmentation with recurrent residual convolutional neural networks based U-Net (R2U-Net). *Proc Naecon Ieee Nat.* 2018:228−233.

25. Morris CL, Morris WR, Fleming JC. A histological analysis of the Mullerectomy: redefining its mechanism in ptosis repair. *Plast Reconstr Surg.* 2011;127:2333−2341.

26. Akaishi P, Galindo-Ferreiro A, Cruz AAV. Symmetry of upper eyelid contour after unilateral blepharoptosis repair with a single-strip frontalis suspension technique. *Ophthalmic Plast Reconstr Surg.* 2018;34:436−439.

27. Choudhary MM, Chundury R, McNutt SA, Perry JD. Eyelid contour following conjunctival Mullerectomy with or without tarsectomy blepharoptosis repair. *Ophthalmic Plast Reconstr Surg.* 2016;32:361−365.

28. Ghassabeh YA. A sufficient condition for the convergence of the mean shift algorithm with Gaussian kernel. *J Multivariate Anal.* 2015;135:1−10.

29. Pan B, Xia JJ, Yuan P, et al. Incremental kernel ridge regression for the prediction of soft tissue deformations. *Med Image Comput Comput Assist Interv.* 2012;15(Pt 1):99−106.

30. Pan B, Zhang G, Xia JJ, et al. Prediction of soft tissue deformations after CMF surgery with incremental kernel ridge regression. *Comput Biol Med.* 2016;75:1−9.

31. Knoops PGM, Borghi A, Ruggiero F, et al. A novel soft tissue prediction methodology for orthognathic surgery based on probabilistic finite element modelling. *PLoS One.* 2018;13:e0197209.

32. Luboz V, Chabanas M, Swider P, Payan Y. Orbital and maxillofacial computer aided surgery: patient-specific finite element models to predict surgical outcomes. *Comput Methods Biomech Biomed Engin.* 2005;8:259−265.

33. Abe N, Kuroda S, Furutani M, Tanaka E. Data-based prediction of soft tissue changes after orthognathic surgery: clinical assessment of new simulation software. *Int J Oral Maxillofac Surg.* 2015;44:90−96.