Automatic identification of meibomian gland dysfunction with meibography images using deep learning

Yi Yu
Renmin Hospital of Wuhan University: Wuhan University Renmin Hospital
Yiwen Zhou
Wuhan University Renmin Hospital
Yanning Yang (ophyyn@163.com)
Renmin Hospital of Wuhan University: Wuhan University Renmin Hospital https://orcid.org/0000-0001-8419-6706
Yabiao Zhou
Wuhan University
Yuejiao Tan
Wuhan University
Lianlian Wu
Wuhan University Renmin Hospital Department of Gastroenterology
Shan Hu
Wuhan University

Research Article

Keywords: Deep learning, Artificial intelligence, Dry eye, Meibomian gland dysfunction, Meibography

Posted Date: February 23rd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-181617/v1

Abstract

Background

Artificial intelligence is developing rapidly, bringing increasing numbers of intelligent products into daily life. However, it has little progress in dry eye, which is a common disease and associated with meibomian gland dysfunction(MGD). Non-invasive infrared meibography, known as an effective diagnostic tool of MGD, allows for objective observation of meibomian glands. Thus, we discuss a deep learning method to measure and assess meibomian glands of meibography.

Methods

We used Mask R-CNN deep learning(DL) framework. A total of 1878 meibography images were collected and manually annotated by two licensed eyelid specialists with two classes: conjunctiva and meibomian glands. The annotated pictures were used to establish a DL model. An independent test dataset contained 58 images was used to compare the accuracy and efficiency of the deep learning model with specialists.

Results

The DL model calculated the ratio of meibomian gland loss with precise values by achieving high accuracy in the identification of conjunctiva (validation loss < 0.35, mAP > 0.976) and meibomian glands (validation loss < 1.0, mAP > 0.92). The comparison between specialists' annotation and the DL model evaluation showed that there is little difference between the gold standard and the model. Each image takes 480ms for the model to evaluate, almost 21 times faster than specialists.

Conclusions

The DL model can improve the accuracy of meibography image evaluation, help specialists to grade the meibomian glands and save their time to some extent.

Background

Dry eye (DE) is a multifactorial ocular surface disease which is generally classified as aqueous deficient dry eye (ADDE) and evaporative dry eye (EDE) [1]. The prevalence of DE is reported approximately 5–35% all over the world[2]. In the United States, the wide prevalence of dry eye imposes a substantial economic burden (an estimated $3.8 billion in health care expenditures annually). Each year, the societal costs (i.e., reduced productivity and indirect costs) associated with this chronic condition amount to approximately $55 billion in the United States[3, 4]. Meibomian gland dysfunction (MGD), which characterized by

functional abnormalities of the meibomian glands[5], is considered to be the leading cause of EDE and the primary determinant of dry eye symptoms[6]. Over 85% of patients who clinically diagnosed with DE have been reported to have co-morbid signs of MGD[7]. While different subtype of DE require diverse targeted treatments, the specific diagnosis of DE subtype and MGD is crucial for obtaining better improvements.

Previous studies have found that meibomian lipids secreted by meibomian glands (MGs) constitute the lipid layer and help preventing tears from excessive evaporation[8]. Thus, MGs play an essential role in the stability of tear film. It is important to evaluate both the morphology and function of the meibomian gland in order to understand the pathophysiology of MGD, make clinical diagnosis and facilitate targeted treatments[9]. Traditionally, MGD can be diagnosed with slit-lamp biomicroscopy through the observation of gland orifice obstruction. In spite of obstruction, MGD has additional clinical signs such as duct dilation, atrophic degeneration and gland loss[10]. Other important features include MG thickness, density, length, distortion, and interglandular space. Non-invasive infrared meibography, recently known as an effective tool, allows for real-time, objective observation of the morphology of MGs[11]. However, the assessment of meibography images are rough and subjective, which does not correspond to the trend towards individualizing treatments, personalized medicine and the following management of chronic disease. Therefore, it is important to develop a precise and objective method to evaluate meibography images.

The rapid development of deep learning may bring about a revolutionary change in the medical industry. In the field of ophthalmology, the diagnosis of most diseases is based on the recognition of multiple images, while image recognition happens to be a popular area of deep learning application. To our knowledge, deep learning has demonstrated great performance on the diagnosis of ophthalmic diseases such as diabetic retinopathy, age-related macular degeneration, glaucoma, and retinopathy of prematurity[12]. However, the automated recognition of DE, grading of DE severity and the DE subtype classification remain a challenge.

In this study, we established a deep learning method for automated recognition and assessment of meibography images, thus relieving the societal and medical burden at the same time.

Materials And Methods

Study design and participants

A total of 950 infrared meibography images of 475 patients were collected. All images were recorded as JPG format using the Oculus® Keratograph 5M (OCULUS, Germany) from January 2017 to January 2019 at Renmin Hospital of Wuhan University Eye Center. This study was approved by the institutional review board of Renmin Hospital of Wuhan University (ID: WDRY2019-K010) and the research was conducted in accordance with the tenets of the Declaration of Helsinki. Because of the retrospective nature and completely anonymized use of images in this study, informed consent was not required. We deleted all

patients' sensitive information before viewing images to ensure that their personal information remained anonymous and confidential. Eyelid specialists that participated in this study were under informed consent.

Pre-processing

The pre-processing includes image cropping and screening (Figure 1).

First, all original images were cropped with a valid area of 549×260 pixels to eliminate interfering factors and improve the accuracy of DL algorithm. Since original images are of the same format and have the same valid area, automated cropping is performed by setting the coordinates, the length and width of the valid area. We cropped the original 950 images into 1,900 images of the same size and format (Figure 2).

Then, images that were unfocused, reflected by illumination, covered by eyelashes, or had other conditions that could interfere with recognition of conjunctiva and MGs were excluded. After a preliminary quality control, a total of 1878 resized images were included. Due to the higher requirement of MGs annotation, a secondary quality control was carried out by the same two specialists and another 311 images without clear MGs structures were excluded.

Image annotation

After scanning, the conjunctiva and MGs area was manually annotated by the same two eyelid specialists independently as ground truth using VGG Image Annotator (version 1.0.5, USA) (Figure 3 and 4). The disagreement of image annotation among the two specialists was settled through the consultant with a third higher level eyelid specialist.

Training algorithm

Mask R-CNN [13], a flexible and efficient framework pretrained with Microsoft Common Objects in Context (MS COCO) dataset[14], was applied to train our model. Through transfer learning[15], we retrained this R-CNN model with our comparatively small sample image datasets, and fine-tuned the parameters of each layer to recognize the conjunctiva and MGs area. TensorFlow, an end-to-end open source platform developed by the Google Brain team, was used to build and deploy our model (Figure 5).

Model training

Images with annotation were randomly allocated to the training dataset and the validation dataset in a ratio of 8:2. The training dataset was used to train DL algorithm and the validation dataset was used to verify the performance of DL algorithm.

In the training stage, model weights based on the MS COCO dataset were used as a pre-training model to load the training dataset of conjunctiva and MGs, respectively. The model weights of conjunctiva and

MGs training datasets were then trained separately as a preparation for validation. We trained the model for 50 epochs with a learning rate of 0.001.

Model verification

In the validation stage, a total of 376 images of validation dataset for conjunctiva and 314 images for MGs were loaded separately. Then, the trained model weights of conjunctiva and MGs were loaded separately. Finally, the masks of conjunctiva and MGs, the mask area ratio of conjunctiva and MGs, and the processing time were output and recorded. The masks output by the model was compared with the previous manual annotation to evaluate the model's performance (Figure 6).

Performance evaluation

The model's performance was evaluated by mAP (mean average precision) and validation loss. The mAP is the mean value of average precisions for each class, reflecting the accuracy of area detection/segmentation on the validation dataset. Validation loss is the loss value of the verification dataset. The smaller the value, the better the training result. If the essential features of the training dataset are obtained by a DL model to a certain extent, it can be used well for the verification dataset. However, excessive learning will lead to over-fitting and DL model will keep obtaining the abnormal features of the training set, and then the validation loss may rise.

The Mask R-CNN loss function is defined as:

$$L = L_{class} + L_{box} + L_{mask} \quad [13]$$

The mask area ratio of conjunctiva and MGs was used to calculate the proportion of normal MGs area. Model processing time was recorded for subsequent comparison with manual processing time.

Comparison between Mask R-CNN model and doctors

To evaluate the Mask R-CNN model's diagnostic ability for MGs loss, 58 meibography images (29 upper eyelids and 29 lower eyelids) that independent from the training and validation dataset was randomly collected as a test set. The performance of the DL model was compared with 2 expert specialists, 4 seniors, and 4 novices from Renmin Hospital of Wuhan University Eye Center. Before evaluation, 10 specialists accepted the same training about how to evaluate MGs.

The evaluation of MGs loss was graded as "less than 1/3" or "larger than 1/3 and less than 2/3" or "larger than 2/3" [16]. The evaluation process was recorded and timed by the same staff.

In order to determine the accuracy of the DL model, another two specialists used VGG Image Annotator to annotate the test set and consulted with a third higher level specialist when there was disagreement. After annotation, we calculated the ratio of MGs loss in each image. Taking manual annotation as the gold standard, the results of Mask R-CNN model and the 10 specialists were analyzed and compared.

Statistical analysis

The statistical analysis was performed using SPSS 20 (IBM, Chicago, Illinois, USA). Data were expressed as the mean ± standard deviation for metric values and as a frequency (percentage) for categorical variables. A two-tailed Student's t test was used to compare differences in accuracy and processing time of the Mask R-CNN model and specialists, and a P value of <0.05 was considered significant for the measured variables. The correlation coefficient was also used to evaluate the correlation between the Mask R-CNN model and specialists.

Results

The performance of Mask R-CNN on identification of conjunctiva

The Mask R-CNN model marks 376 images in the conjunctival test set with an average accuracy of mAP more than 97.6%. Validation loss is under 0.35. The average processing time per image is 0.163s (61.216 s for 376 images) (Fig. 7).

The performance of Mask R-CNN on identification of meibomian glands

The Mask R-CNN model marks 314 images in the meibomian glands test set with an average accuracy of mAP more than 92.0%. Validation loss is under 1.0. The total time for the model to process all 314 test set images is 114.399s, and the average processing time per image is 0.356s (Fig. 7).

Comparison between Mask R- CNN and Doctors

According to the results of the evaluation of the test set by Mask R-CNN and specialists, processing time and accuracy were compared. The results of time comparison showed that the total time taken by specialists to evaluate 58 images averaged 592.59s, with the fastest being 453.56s and the slowest being 718.04s. On average, 10.22(± 1.37)s is required for each picture, the fastest one takes 7.82s per picture, and the slowest one averages 12.38s. On the other hand, the computer takes a total of 27.96s to evaluate these 58 images, and each picture takes an average of 0.48s, almost 21 times faster than specialists (Table 1).

Table 1
Comparison of time spent by machine and doctors

	Total time (second)	Time per image (second)
Expert specialist 1	580.58	10.01
Expert specialist 2	475.60	8.20
Senior 1	567.24	9.78
Senior 2	655.40	11.30
Senior 3	609.00	10.50
Senior 4	453.56	7.82
Novice 1	687.30	11.85
Novice 2	601.46	10.37
Novice 3	577.68	9.96
Novice 4	718.04	12.38
Average (mean ± SD)	592.59	10.22 ± 1.37
Mask R-CNN	27.96	0.48

From the result of accuracy comparison, specialists can only roughly evaluate the meibomian gland structure in the image as "less than 1/3 loss", " larger than 1/3 and less than 2/3 loss" or "larger than 2/3 loss". The evaluation results of specialists were integrated (Fig. 8). We can easily tell that evaluation of the same image are different among the ten specialists.

The results of the evaluation of 58 images by Mask R-CNN was compared with results of manual annotation (taken as gold standard), and correlation coefficient was calculated, r = 0.976 (Fig. 9). Calculate the absolute value of the difference between the two sets of data, the average is 0.0355, SD = 0.0212, which means that the difference between the gold standard and Mask R-CNN is very small (3.55% ±2.12%).

Discussion

The prevalence of dry eye and MGD has increased dramatically in recent years, affecting billions of people worldwide. With growing understanding of such diseases, clinical diagnosis cares more about the objective and quantitative observation of meibomian glands. There are problems such as huge workloads for clinicians, inaccuracy of clinical results, weak association between severity and cure, especially in some developing countries. All deficiencies mentioned above restrict the diagnosis, treatment and long-term managements. Accurate assessment of meibomian glands in patients, can help

clinicians determine treatment plan, grade the severity of the disease and assess the effectiveness of treatment that patients have received. It is of great importance to explore an effective and accurate meibomian gland function assessment method, and to achieve the automation and intelligence assessment is to satisfy the necessity of personalized treatment and chronic disease management in large population.

AI is developing rapidly and increasing numbers of intelligent products are entering into daily life. It has the potential to revolutionize disease diagnosis and management by rapidly reviewing immense amounts of images time-consuming for human clinicians[17–19]. Deep learning(DL) is the most advanced branch of machine learning(ML) and at the fundamental of AI. It excels at the problem of learning from big data and make predictions afterwards, which is well-suited for our study. We chose a pretrained Mask R-CNN algorithm to perform our task. Mask R-CNN was developed based on Faster R-CNN and outperformed other single models for object instance segmentation and detection[13]. Without competition among classes, Mask R-CNN removes the harsh quantization of RoIPool, aligning extracted features properly with the input and preserving the explicit per-pixel spatial correspondence. MS COCO is a large-scale object detection, segmentation, and caption dataset[14]. There are 1.5 million object instances, 80 object categories, and 91 stuff categories. Each type of object has many images, and each image contains accurate segmentation information.

In our study, the accuracy of the DL model is much higher than the specialists[20]. The DL-based model has an objective and accurate advantage that doctors cannot match. It can stably and accurately identify the meibomian gland features and give the test results immediately. The speed of the model is almost 21 times that of doctors, revealing the practical applications and unique advantages of deep learning algorithm in improving the accuracy of the results and saving time. With the help of the deep-learning-based model, doctors can evaluate meibomian gland more accurately, objectively and quickly, so as to better personalize the diagnosis and treatment of patients.

There are some limitations in our study. The dataset is comparatively small and the model only distinguishes between the “presence” or “loss” of meibomian glands. In clinics, the meibomian gland abnormalities also include different types of structural anomalies and structural defects, so the accuracy of image recognition remains to be further refined when used in real clinical setting[21, 22]. In the later stage, our team will also establish multi-center databases with other hospitals to make data sources more universal and more abundant, to further improve the accuracy of the model and promote the development of artificial intelligence in ocular-surface-related diseases.

Conclusions

The deep learning based meibomian gland intelligent assessment model can achieve extremely high accuracy, helping specialists to evaluate and assess meibography images better and faster, which primarily provides a reliable basis in support of individualized treatment and chronic disease management.

Declarations

Authors' contributions

Yanning Yang conceived and supervised the overall study. Yi Yu and Yiwen Zhou contributed to writing of the report and conduct the study. Yi Yu, Yiwen Zhou and Lianlian Wu collected images. Yabiao Zhou, Yuejiao Tan trained the model. Yi Yu and Yiwen Zhou contributed to the statistical analysis. Yanning Yang guaranteed the article.

All authors reviewed and approved the final version of the manuscript.

Conflict of interest statements

All authors declared no conflict of interest.

Acknowledgements

This work was partly supported by the grant from the National Natural Science Foundation of China [Grant Nos. 81770899 to Yang Yanning]

References

1. Craig JP, Nichols KK, Akpek EK, Caffery B, Dua HS, Joo CK, Liu Z, Nelson JD, Nichols JJ, Tsubota K, Stapleton F (2017) "TFOS DEWS II Definition and Classification Report". Ocul Surf 15:276–283
2. Dai Y, Zhang J, Xiang J, Li Y, Wu D, Xu J (2019) "Calcitriol inhibits ROS-NLRP3-IL-1beta signaling axis via activation of Nrf2-antioxidant signaling in hyperosmotic stress stimulated human corneal epithelial cells". Redox Biol 21:101093
3. Clayton JA (2018) "Dry Eye" N Engl J Med 379:e19
4. Yu J, Asche CV, Fairchild CJ (2011) "The economic burden of dry eye disease in the United States: a decision tree analysis". Cornea 30:379–387
5. Chhadva P, Goldhardt R, Galor A, "Meibomian Gland Disease: The Role of Gland Dysfunction in Dry Eye Disease," Ophthalmology 124, S20-S26 (2017)
6. Teo CHY, Ong HS, Liu YC, Tong L (2020) Meibomian gland dysfunction is the primary determinant of dry eye symptoms: Analysis of 2346 patients. Ocul Surf 18:604–612
7. Lemp MA, Crews LA, Bron AJ, Foulks GN, Sullivan BD (2012) Distribution of aqueous-deficient and evaporative dry eye in a clinic-based patient cohort: a retrospective study. Cornea 31:472–478
8. Eom Y, Lee JS, Kang SY, Kim HM, Song JS, "Correlation between quantitative measurements of tear film lipid layer thickness and meibomian gland loss in patients with obstructive meibomian gland dysfunction and normal controls," Am J Ophthalmol 155, 1104–1110 e1102 (2013)
9. Arita R, Fukuoka S, Morishige N (2017) Functional Morphology of the Lipid Layer of the Tear Film. Cornea 36(Suppl 1):S60–S66

10. Nelson JD, Shimazaki J, Benitez-del-Castillo JM, Craig JP, McCulley JP, Den S, Foulks GN, "The international workshop on meibomian gland dysfunction: report of the definition and classification subcommittee," Invest Ophthalmol Vis Sci 52, 1930–1937 (2011)
11. Finis D, Ackermann P, Pischel N, Konig C, Hayajneh J, Borrelli M, Schrader S, Geerling G (2015) Evaluation of Meibomian Gland Dysfunction and Local Distribution of Meibomian Gland Atrophy by Non-contact Infrared Meibography. Curr Eye Res 40:982–989
12. Ting DSW, Pasquale LR, Peng L, Campbell JP, Lee AY, Raman R, Tan GSW, Schmetterer L, Keane PA, Wong TY (2019) Artificial intelligence and deep learning in ophthalmology. Br J Ophthalmol 103:167–175
13. He K, Gkioxari G, Dollar P, Girshick R (2020) Mask R-CNN. IEEE Trans Pattern Anal Mach Intell 42:386–397
14. Tan AJ, Zhou GX, He MF (2020) Rapid Fine-Grained Classification of Butterflies Based on FCM-KM and Mask R-CNN Fusion. Ieee Access 8:124722–124733
15. Torrey L, Shavlik J, "Transfer learning," in Handbook of research on machine learning applications and trends: algorithms, methods, and techniques (IGI global, 2010), pp. 242–264
16. Koh YW, Celik T, Lee HK, Petznick A, Tong L (2012) Detection of meibomian glands and classification of meibography images. J Biomed Opt 17:086008
17. Kermany DS, Goldbaum M, Cai W, Valentim CCS, Liang H, Baxter SL, McKeown A, Yang G, Wu X, Yan F, Dong J, Prasadha MK, Pei J, Ting MYL, Zhu J, Li C, Hewett S, Dong J, Ziyar I, Shi A, Zhang R, Zheng L, Hou R, Shi W, Fu X, Duan Y, Huu VAN, Wen C, Zhang ED, Zhang CL, Li O, Wang X, Singer MA, Sun X, Xu J, Tafreshi A, Lewis MA, Xia H, Zhang K, "Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning," Cell 172, 1122–1131 e1129 (2018)
18. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, Venugopalan S, Widner K, Madams T, Cuadros J, Kim R, Raman R, Nelson PC, Mega JL, Webster DR (2016) Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. JAMA 316:2402–2410
19. Long E, Lin H, Liu Z, Wu X, Wang L, Jiang J, An Y, Lin Z, Li X, Chen J (2017) An artificial intelligence platform for the multihospital collaborative management of congenital cataracts. Nature biomedical engineering 1:1–8
20. Feng Y, Gao Z, Feng K, Qu H, Hong J (2014) Meibomian gland dropout in patients with dry eye disease in China. Curr Eye Res 39:965–972
21. Arita R, Fukuoka S, Morishige N (2017) New insights into the morphology and function of meibomian glands. Exp Eye Res 163:64–71
22. Eom Y, Choi K-E, Kang S-Y, Lee HK, Kim HM, Song JS, "Comparison of meibomian gland loss and expressed meibum grade between the upper and lower eyelids in patients with obstructive meibomian gland dysfunction," Cornea 33, 448–452 (2014)

Figures

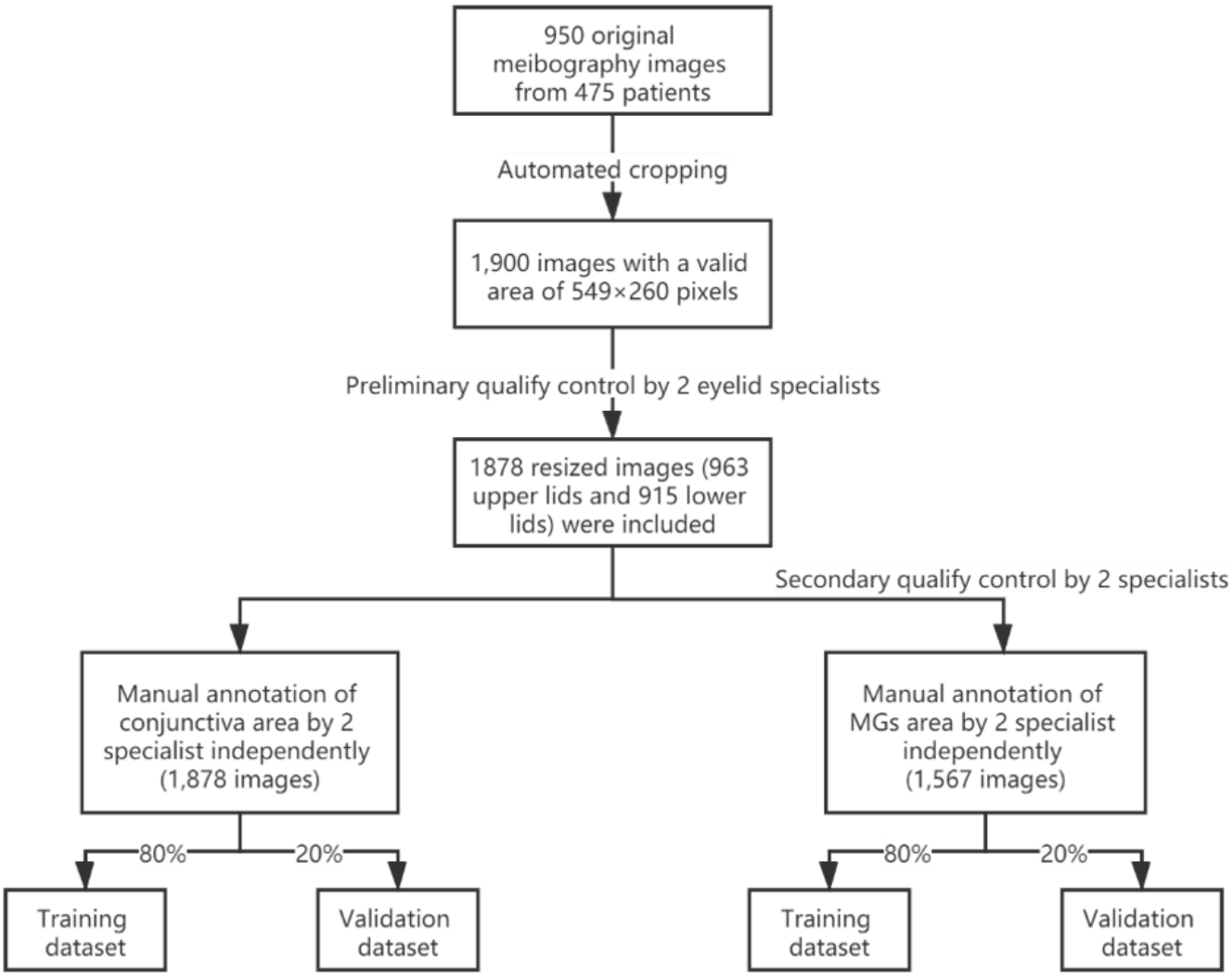

Figure 1

Workflow of image obtaining, pre-processing, annotating and dataset division.

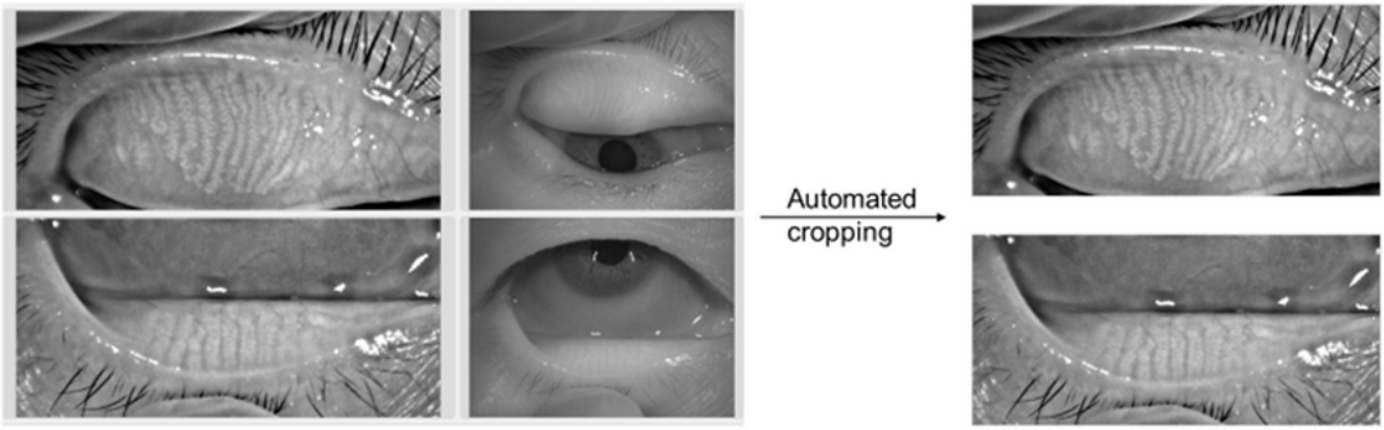

Figure 2

Original images were cropping automatedly by setting the width and location of images.

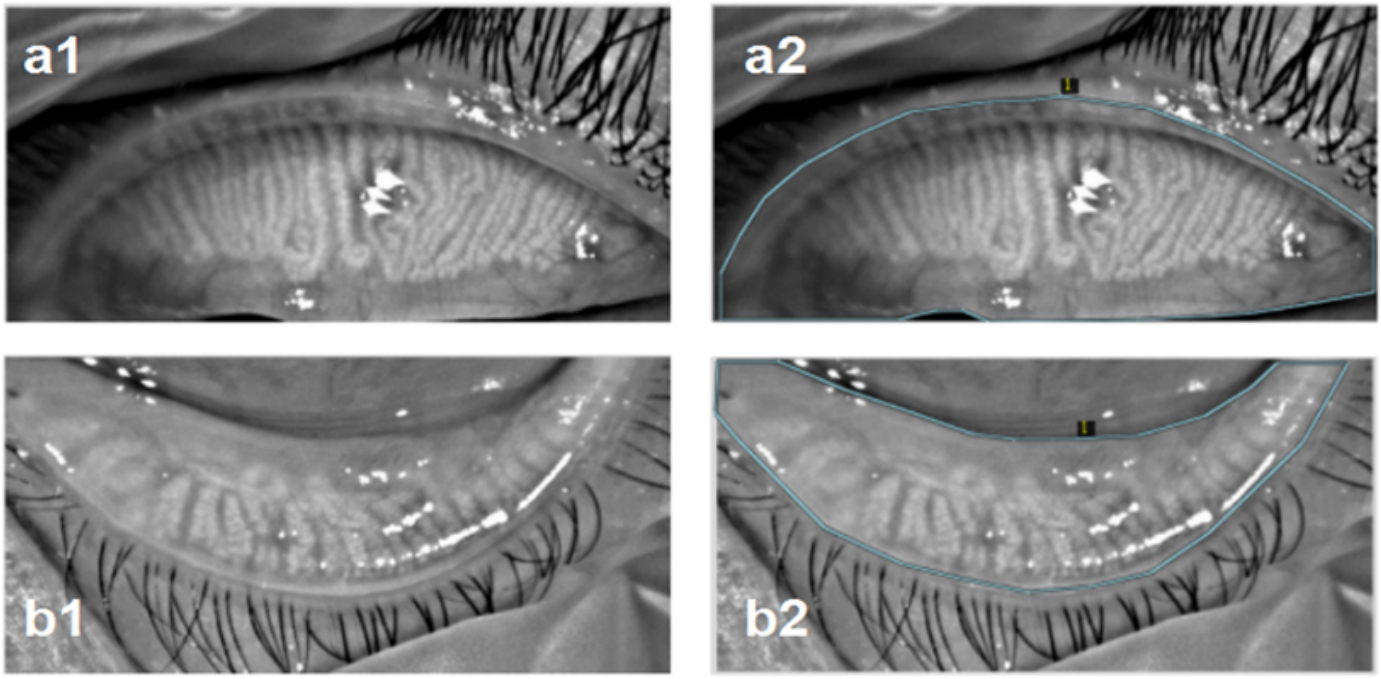

Figure 3

Representative images of conjunctiva area manual annotations. a1. A upper eyelid. a2. Annotation of a upper eyelid. b1. A lower eyelid. b2. Annotation of a lower eyelid.

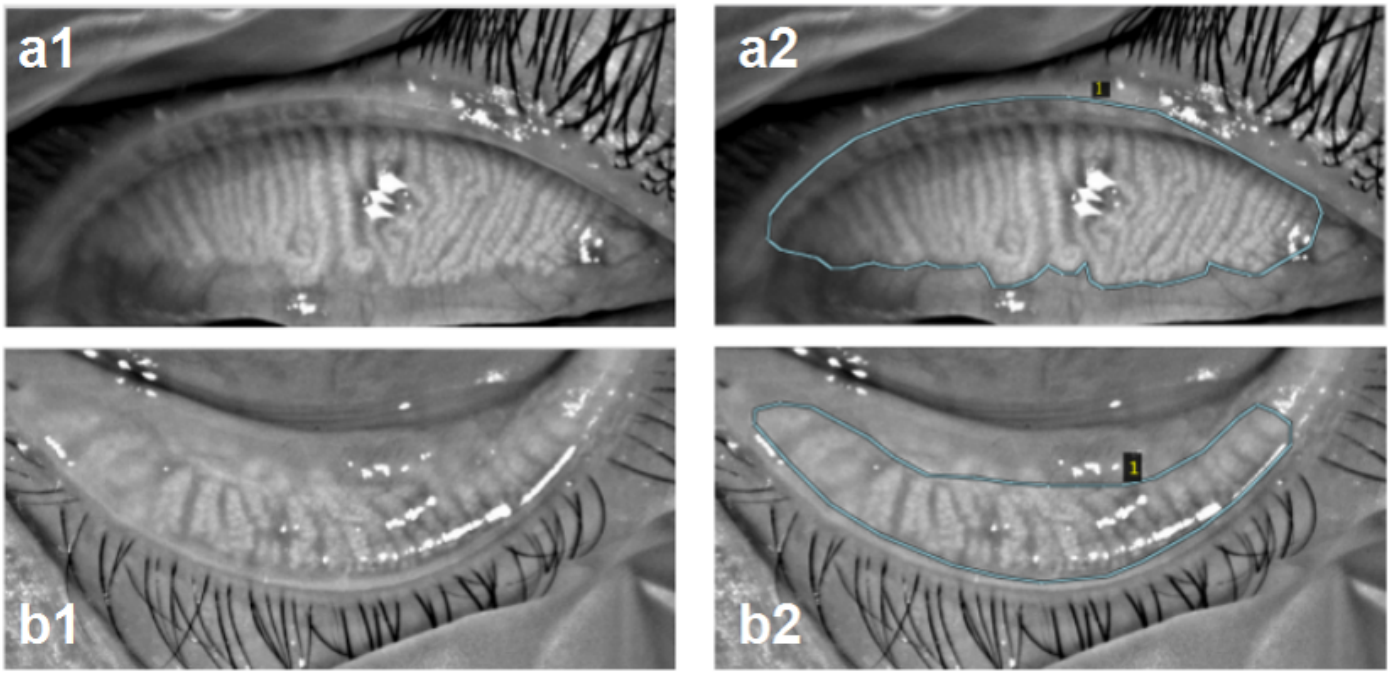

Figure 4

Representative images of meibomian glands manual annotations. a1. A upper eyelid. a2. Annotation of a upper eyelid. b1. A lower eyelid. b2. Annotation of a lower eyelid.

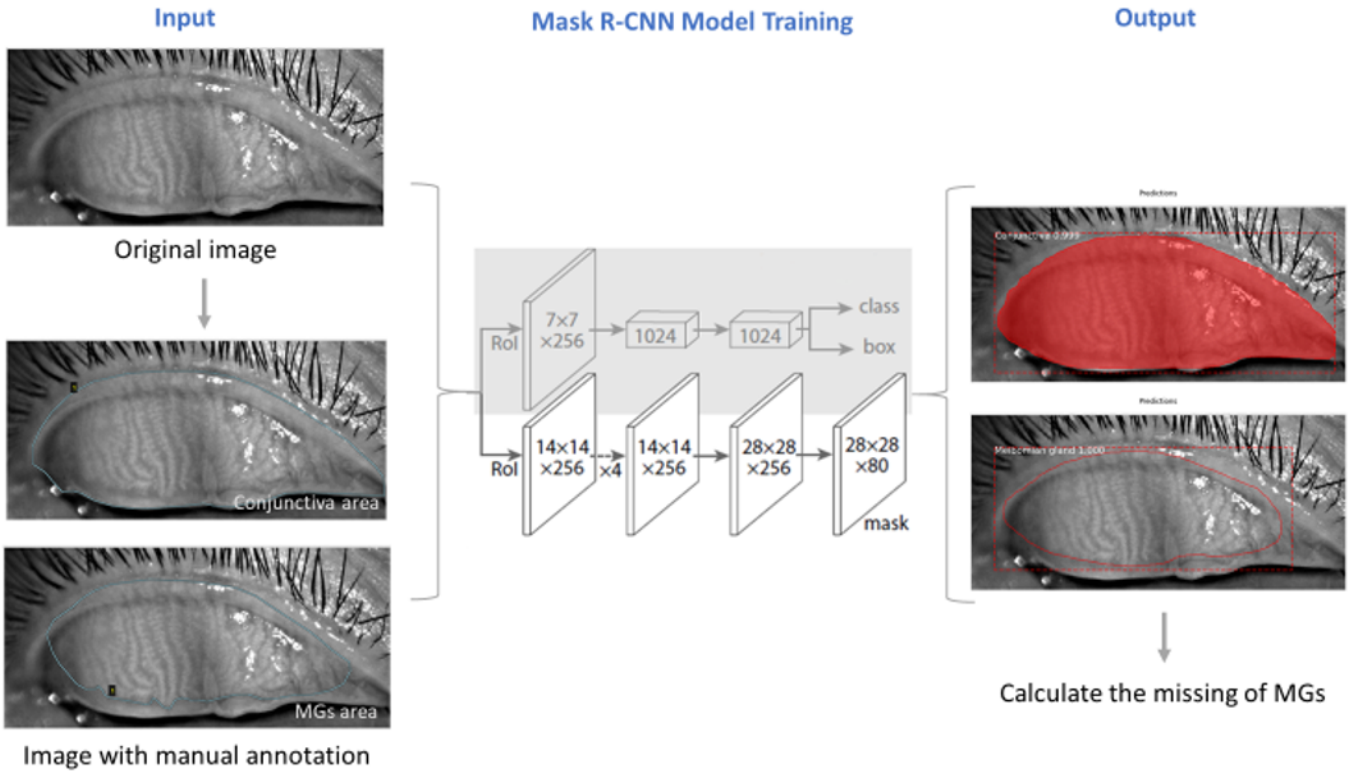

Figure 5

Process for detection of conjunctiva and MGs structure in meibography images. Data flow is from left to right: meibography images were subsequently processed using Mask R-CNN based model and finally output with masks. The mask area ratio of conjunctiva and MGs was used to calculate the proportion of normal MGs area.

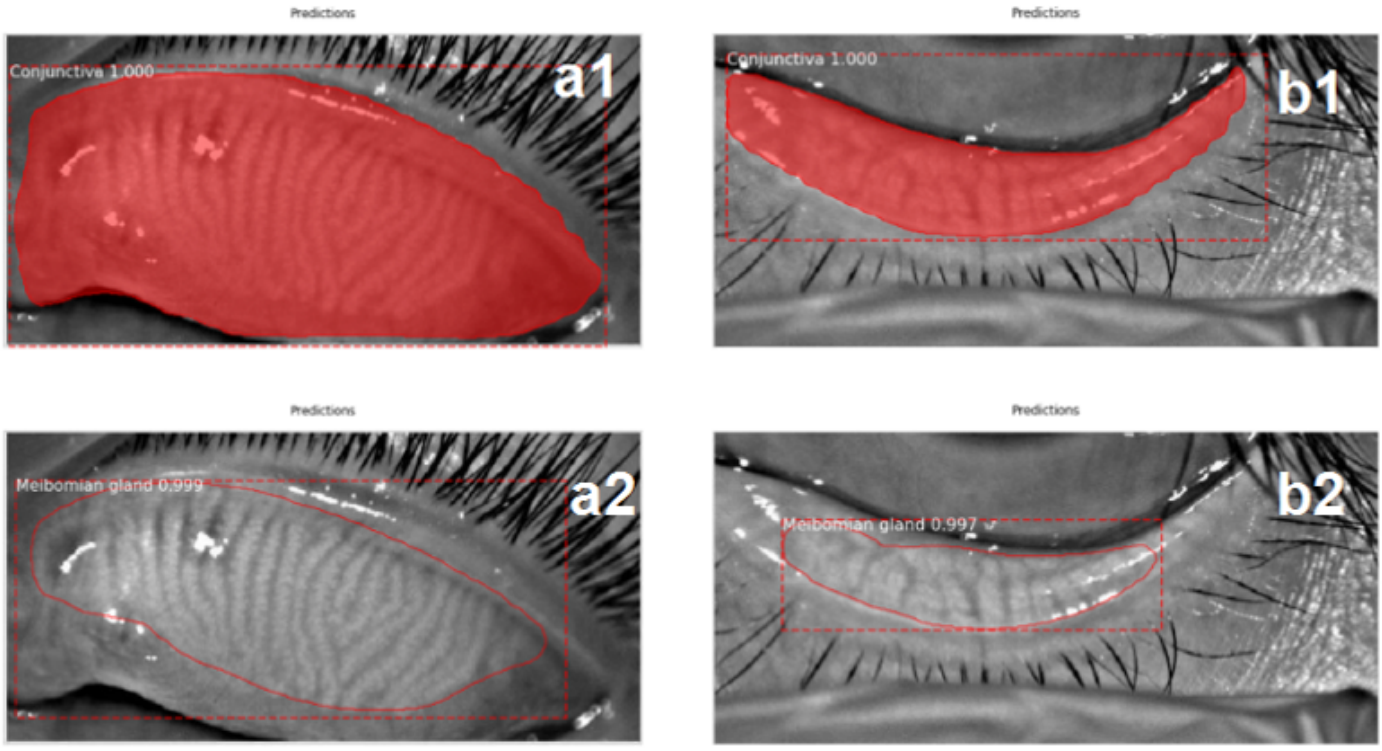

Figure 6

Mask R-CNN identifies the range of conjunctiva and meibomian gland. a1, Mask R-CNN identifies the range of conjunctiva of upper eyelid. a2, Mask R-CNN identifies the range of conjunctiva of upper eyelid. b1, Mask R-CNN identifies the range of conjunctiva of lower eyelid. b2, Mask R-CNN identifies the range of meibomian gland of lower eyelid.

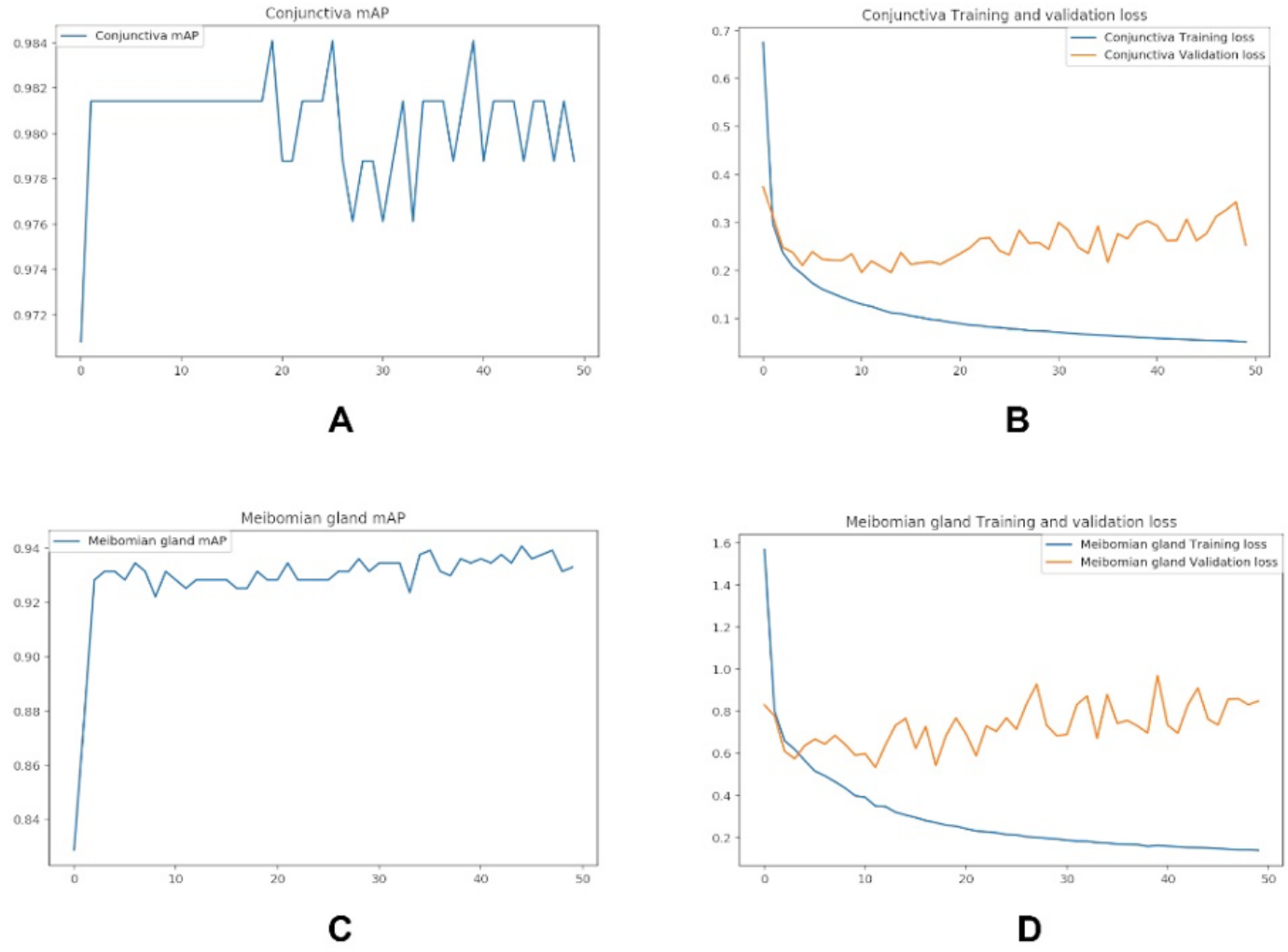

Figure 7

Training curve of model recognition of conjunctiva. A. Validation loss < 0.35. B. mAP > 0.976. Training curve of model recognition of meibomian glands. C. Validation loss < 1.0. D. mAP > 0.92.

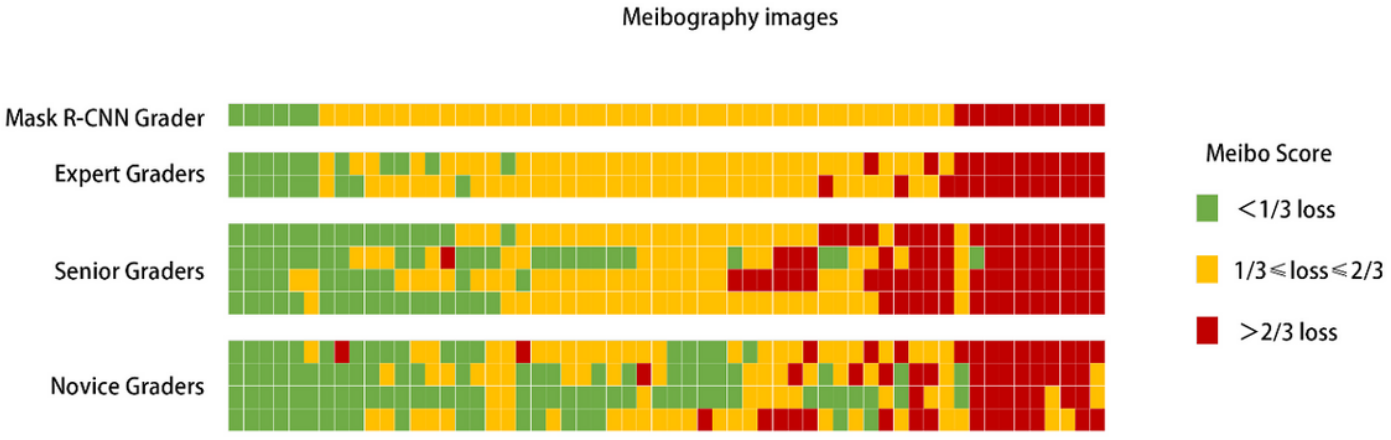

Figure 8

Ophthalmologist's grading results of 58 images. We collected 58 images independent from training and validation dataset. In this picture, green, yellow and red correspond to meibomian glands loss less than 1/3, between 1/3 and 2/3, and more than 2/3, respectively. Each small grid represents an image. The consistency of less than 1/3 loss and more than 2/3 loss were better than between 1/3 and 2/3.

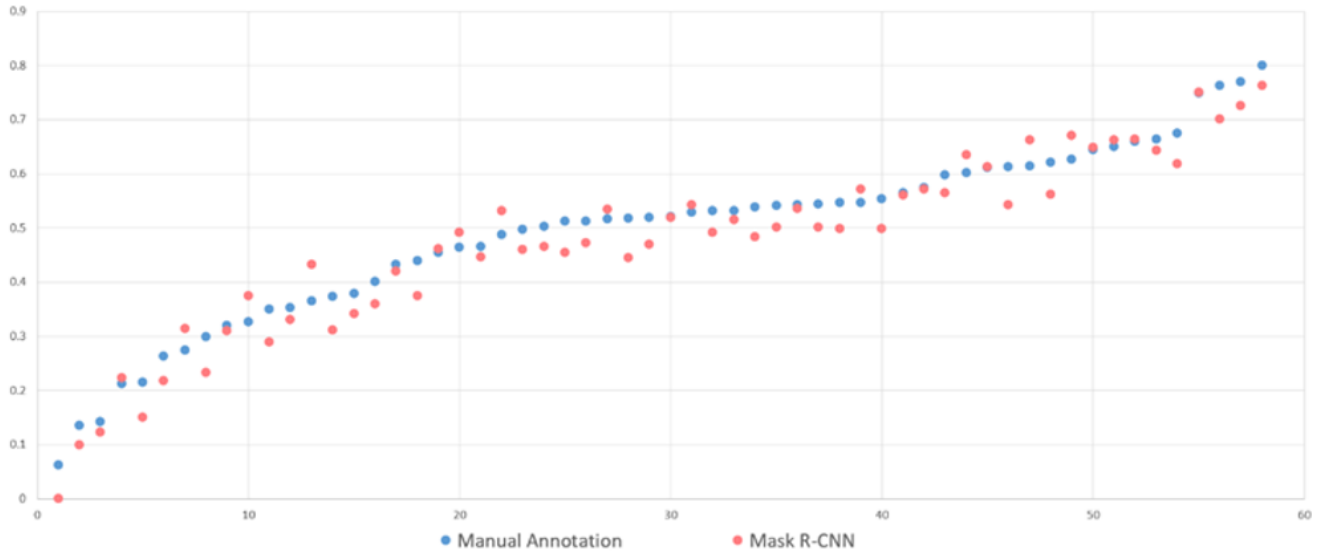

Figure 9

Comparison of Mask R-CNN marking results and manual annotation results. The results of the evaluation of 58 images by Mask R-CNN was compared with results of manual annotation (taken as gold standard), and correlation coefficient was calculated.