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

Yi Yu · Yiwen Zhou · Miao Tian · Yabiao Zhou · Yuejiao Tan · Lianlian Wu · Hongmei Zheng · Yanning Yang

Received: 8 February 2021 / Accepted: 12 March 2022 / Published online: 19 September 2022
© The Author(s), under exclusive licence to Springer Nature B.V. 2022

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). Noninvasive 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 that 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 480 ms 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.

Keywords Deep learning · Artificial intelligence · Dry eye · Meibomian gland dysfunction · Meibography

Background

Meibomian glands (MGs) are ocular adnexa and the sebaceous glands distributed along the rims of the eyelids. Meibomian lipids secreted by meibomian glands known as meibum constitute the lipid layer...
and help preventing tears from excessive evaporation [1]. Thus, MGs play an essential role in the stability of tear film. Meibomian gland dysfunction (MGD), which is characterized by anatomic changes or functional abnormalities of the meibomian glands [2], is considered to be the main cause of dry eye (DE). Over 85% of patients who clinically diagnosed with DE have been reported to have co-morbid signs of MGD [3]. Therefore, it is important to evaluate both the morphology and function of the meibomian gland in order to understand the pathophysiology of MGD, then make clinical diagnosis and facilitate targeted treatments [4].

Morphological features of MGs are important to evaluate the health of MGs [5, 6]. Traditionally, MGD can be diagnosed with slit-lamp microscope 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 [7]. Other important features include MG thickness, density, length, distortion and inter-glandular space. Noninvasive infrared meibography, recently known as an effective tool, allows for real-time, objective observation of the morphology of MGs [8]. However, the assessment of meibography images is 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 [9]. However, the automated recognition of MGs, grading of MGD severity and 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 subjects (18–65 years old) were collected. The exclusion criteria include: (1) subjects with history of ocular injury or surgery, (2) using ocular or systemic medications known to affect the ocular surface or tear film, (3) ocular or systemic diseases that affect the anatomy of the anterior segment or tear film. All the meibography images (including the upper and lower eyelids) were recorded as JPG format using the Oculus® Keratograph 5 M (OCULUS, Germany) from January 2017 to January 2019 in 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 (Fig. 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 1900 images of the same size and format (Fig. 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 MG structures were excluded.
After scanning, the conjunctiva and MGs area were manually annotated by the same two eyelid specialists independently as ground truth using VGG Image Annotator (version 1.0.5, USA) (Figs. 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 [10], a flexible and efficient framework pre-trained with Microsoft Common Objects in Context (MS COCO) dataset [11], was applied to train our model. Through transfer learning [12],...
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 (Fig. 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...
The masks’ output by the model was compared with the previous manual annotation to evaluate the model’s performance (Fig. 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 overfitting 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 [10]:

\[ L = L_{\text{class}} + L_{\text{box}} + L_{\text{mask}} \]

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 are independent from the training and validation dataset were 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” [13]. 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 0.163 s (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 114.399 s, and the average processing time per image is 0.356 s (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 592.59 s, with the fastest being 453.56 s and the slowest being 718.04 s. On average, 10.22(± 1.37)s is required for each picture, the fastest one takes 7.82 s per picture, and the slowest one averages 12.38 s. On the other hand, the computer takes a total of 27.96 s to evaluate these 58 images, and each picture takes an average of 0.48 s, almost 21 times faster than specialists (Table 1).

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 the evaluation of the same image is different among the ten specialists.
The results of the evaluation of 58 images by Mask R-CNN were compared with results of manual annotation (taken as gold standard), and correlation coefficient was calculated, \( r = 0.976 \) (Fig. 9). Calculating 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 [14, 15]. 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 makes predictions afterwards, which is well-suited for our study. We chose a pre-trained 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 [10]. 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

|                      | 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                    |

**Table 1** Comparison of time spent by machine and doctors
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 [16]. 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 can only be applied to distinguish the loss of meibomian glands for now. In clinics, the meibomian gland abnormalities also include morphological changes such as shortening, entanglement, distortion and segmentation [17], so the accuracy of image recognition remains to be further refined when used in real clinical setting [18, 19]. In the later stage, our team will also establish multicenter 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.

Acknowledgements

The authors thank all of trainees and clinical collaborators for their contributions.

This work was partly supported by the Key Research and Development program of Hubei Province [Grant Nos. 2020BCB055 to Yang Yanning].

Authors’ contributions YY conceived and supervised the overall study. YY and YZ contributed to writing of the report and conducted the study. YY, MT and LW collected the images. YZ and YT trained the model. YY and YZ contributed to the statistical analysis. HZ and YY guaranteed the article.

Funding The authors have not disclosed any funding.
Availability of data and material  All the data are available.

Declarations

Conflict of interest  All authors declared no conflict of interest and agreed to publish.

Ethical approval  This study was approved by the Ethics Committee of Renmin Hospital of Wuhan University, and under trail registration number 2019 K-K010. This study does not involve animal experiments adherence to the Declaration of Helsinki.

References

1. Eom Y, Lee JS, Kang SY, Kim HM, Song JS (2013) 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(6):1104–1110. https://doi.org/10.1016/j.ajo.2013.01.008

2. Chhadva P, Goldhardt R, Galor A (2017) Meibomian gland disease: the role of gland dysfunction in dry eye disease. Ophthalmology 124(11):S20–S26. https://doi.org/10.1016/j.jophtha.2017.05.031

3. Lemp MA, Crews LA, Bron AJ, Foukls GN, Sullivan BD (2012) Distribution of aqueous-deficient and evaporative dry eye in a clinic-based patient cohort: a retrospective study. Cornea 31(5):472–478. https://doi.org/10.1097/ICO.0b013e318225415a

4. Arita R, Fukuoka S, Morishige N (2017) Functional morphology of the lipid layer of the tear film. Cornea 36(Suppl 1):S60–S66. https://doi.org/10.1097/ICO.0000000000001367

5. Lin X, Fu Y, Li L, Chen C, Chen X, Mao Y, Lian H, Yang W, Dai Q (2020) A novel quantitative index of meibomian gland dysfunction, the meibomian gland tortuosity. Transl Vis Sci Technol 9(9):34. https://doi.org/10.1167/tvst.9.9.34

6. Chan T, Wan KH, Shih KC, Jhanji V (2018) Advances in dry eye imaging: the present and beyond. Br J Ophthalmol 102(3):295–301. https://doi.org/10.1136/bjophthalmol-2017-310759

7. Nelson JD, Shimazaki J, Benitez-del-Castillo JM, Craig JP, McCulley JP, Den S, Foukls GN (2011) The international workshop on meibomian gland dysfunction: report of the definition and classification subcommittee. Invest Ophthalmol Vis Sci 52(4):1930–1937. https://doi.org/10.1167/iovs.10-6997b

8. 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 meiography. Curr Eye Res 40(10):982–989. https://doi.org/10.1007/s11942-014-9719-29

9. Li JO, Liu H, Ting D, Jeon S, Chan R, Kim JE, Sim DA, Thomas P, Lin H, Chen Y, Sakamoto T, Loewenstein A, Lam D, Pasquale LR, Wong TY, Lam LA, Ting D (2021) Digital technology, tele-medicine and artificial intelligence in ophthalmology: a global perspective. Prog Retin Eye Res 82:100900. https://doi.org/10.1016/j.preteyeres.2020.100900

10. He K, Gkioxari G, Dollar P, Girshick R (2020) Mask R-CNN. IEEE Trans Pattern Anal Mach Intell 42(2):386–397. https://doi.org/10.1109/TPAMI.2018.2844175

11. Ouyang N, Wang W, Ma L, Wang Y, Chen Q, Yang S, Xie J, Su S, Cheng Y, Cheng Q, Zheng L, Yuan Y (2021) Diagnosing acute promyelocytic leukemia by using convolutional neural network. Clin Chim Acta 512:1–6. https://doi.org/10.1016/j.cca.2020.10.039

12. Cai C, Wang S, Xu Y, Zhang W, Tang K, Ouyang Q, Lai L, Pei J (2020) Transfer learning for drug discovery. J Med Chem 63(16):8683–8694. https://doi.org/10.1021/jacs.9b02147

13. Koh YW, Celik T, Lee HK, Petznick A, Tong L (2012) Detection of meibomian glands and classification of meiography images. J Biomed Opt 17(8):86008. https://doi.org/10.1117/1.JBO.17.8.086008

14. Kermany DS, Goldbaum M, Cai W, Valentim C, Liang H, Baxter SL, McKeown A, Yang G, Wu X, Yan F, Dong J, Prasadaha MK, Pei J, Ting M, Zhu J, Li C, Hewett S, Dong J, Ziyar I, Shi A, Zhang R, Zheng L, Hou R, Shi W, Fu X, Duan Y, Huan V, Wen C, Zhang ED, Zhang CL, Li O, Wang X, Singer MA, Sun X, Xu J, Tafreshi A, Lewis MA, Xia H, Zhang K (2018) Identifying medical diagnoses and treatable diseases by Image-Based deep learning. Cell 172(5):1122–1131. https://doi.org/10.1016/j.cell.2018.02.010

15. Li F, Wang Y, Xu T, Dong L, Yan L, Jiang M, Zhang X, Jiang H, Wu Z, Zou H (2021) Deep learning-based automated detection for diabetic retinopathy and diabetic macular oedema in retinal fundus photographs. Eye (Lond). https://doi.org/10.1038/s41433-021-01552-8

16. 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(10):965–972. https://doi.org/10.3109/02713683.2014.891748

17. Villani E, Marelli L, Dellavalle A, Serafino M, Nucci P (2020) Latest evidences on meibomian gland dysfunction diagnosis and management. Ocul Surf 18(4):871–892. https://doi.org/10.1016/j.jtos.2020.09.001

18. Arita R, Fukuoka S, Morishige N (2017) New insights into the morphology and function of meibomian glands. Exp Eye Res 163:64–71. https://doi.org/10.1016/j.exer.2017.06.010

19. Eom Y, Choi KE, Kang SY, Lee HK, Kim HM, Song JS (2014) Comparison of meibomian gland loss and expressed meibum grade between the upper and lower eyelids in patients with obstructive meibomian gland dysfunction. Cornea 33(5):448–452. https://doi.org/10.1097/ICO.0000000000000992

Publisher’s Note  Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other
