APPENDIX

We present detailed data processing procedures, network design, loss functions and training details of the proposed model, Meibomian gland morphological feature measuring methods, and Meibomian gland detection evaluation details.

Data Processing

Upper Eyelid Total Area

To minimize the annotation variability across different annotators, only the visible part of Meibomian glands in each image were annotated. As occluded parts of Meibomian glands hidden behind the tarsal plate should also be considered, algorithms were designed to extend Meibomian gland annotations in upper eyelids. Specifically, a principal curve[A30] was fit to each (visible) annotated Meibomian gland region to obtain its linear direction. Meibomian gland annotations were then extended along the linear principal curve direction until they intersected with the upper borders of annotated eyelid regions, which were defined at the Meibomian gland orifices. Appendix figure A1 shows some examples of upper eyelid gland completion.

Center Crop

Meibography imaging projects the three-dimensional eyelid surface to a two-dimensional image, with more distortion in the far nasal and temporal image regions. Furthermore, far temporal and nasal glands were less distinguishable and were difficult to annotate accurately due to the inter-observer differences in eyelid eversion techniques that result in Meibomian glands not being imaged at all or imaged Meibomian glands appearing blurry. Therefore, based on the eyelid region annotations, all glands located within or partially overlapping with the central 50% of the

upper eyelid were retained for analysis. The boundaries of the cropped central region are outlined in white in Appendix figure A1 for visualization. Processed images with annotations and center cropping were verified by a clinical investigator (MCL).

The gland segmentation network took a meibography image as an input and produced two intermediate outputs: the entire Meibomian gland region contour (along with pixel-wise feature embedding) and the number of glands. The predicted Meibomian gland contour, embedding, and gland number were further used by the network to predict Meibomian gland region instance segmentation.

Network Design Details

Instance Segmentation

Meibomian gland instance segmentations were generated differently in the training and evaluation phases, since ground-truth segmentation (Meibomian gland region contours) were not available in the evaluation phase. In the training phase, ground-truth segmentation (with pixel-wise feature embedding) for individual gland regions was available. Loss functions were employed to penalize the difference between network predictions and corresponding ground truths during training, so that networks could learn from the annotations to segment Meibomian glands accurately. Further details of the design and implementation of different loss terms can be found in a previous paper. [A21]

In the evaluation phase, intermediate outputs including pixel-wise embedding and gland number were extracted. K-means clustering was then performed on the embeddings to form different gland clusters. Different clusters corresponded to individual gland instance segmentations.

Ghost Gland Identification

A deep learning model was designed to identify ghost glands (Figure 3 of the paper). The neural network had three inputs: a whole meibography image, a specific gland region contour, and a corresponding cropped gland image. Inputs consisted of a global view of the meibography image and a local view of individual glands; thus, two subnetworks were used to learn to identify ghost glands. The global network learned a representation from the entire meibography image and the individual gland region contour, while the local network learned a representation from the individual gland image. The two representations were combined to predict ghost glands. A cross-entropy loss was utilized to penalize wrong predictions. Specifically, the numbers of ghost and non-ghost glands were highly imbalanced (Figure 7 of the paper). The imbalanced data distribution can result in the model having high bias and a focus on learning from the majority. The imbalance in the data distribution leads to inequality in category-wise performance as the model ignores properties of the minority (i.e., ghost glands). To avoid imbalanced category-wise performance, [A31, A32] class-aware balanced sampling[A33] was employed to ensure that each training data batch had the same number of ghost and non-ghost glands and similar classification performance across two categories, so that the model learned from different categories equally.

Loss Functions for Training the Segmentation Model

To train the Meibomian gland segmentation model, four loss terms were used: 1) A cross-entropy loss was used to generate an entire Meibomian gland region contour similar to the ground truth; 2) A regression loss was employed to minimize the difference between the predicted gland number and the actual gland number; 3) A contrastive loss was utilized to reduce intra-gland pixel embedding distances and enlarge inter-gland pixel embedding distances; 4) A regularization loss was incorporated to penalize clusters too far away from the origin to keep the activations bounded. The four loss terms cooperatively produced reasonable Meibomian gland instance segmentations. Further details of the design and implementation of different loss terms can be found in a previous paper. [A21]

Model Training Details

Meibomian gland instance segmentation

Each meibography image and corresponding Meibomian gland region segmentations were resized to 262×262 pixels. During training, 256×256 pixels were randomly cropped out of a given meibography image along with the corresponding annotations in every training epoch for data augmentation. A center crop of 256×256 pixels was made to a given meibography image and corresponding annotations during the evaluation process for both validation and evaluation datasets.

Different network architectures (ReSeg networks[A34] and stacked recurrent hourglass networks[A35]), weights of different loss terms, learning rate, and learning-rate decreasing policy were carefully assessed to obtain the best performance of the network on the validation dataset.

The best performance of the network over the validation set was the ReSeg network backbone, cross-entropy loss weight 1.0, regression loss weight 1.0, contrastive loss weight 1.0, regularization loss weight 1e-3, initial learning rate 1.0, 200 epochs in total, with learning rate decrease when the validation loss was not decreasing for 20 epochs. The algorithm performance of the model on the evaluation dataset was reported.

Ghost Gland Classification

As mentioned above, each gland in a meibography image was augmented to three images: the entire meibography image, the individual gland region contour and individual gland images. The first two images were fed to a global network while the last image was fed to a local network. Two representations from both networks were concatenated and a linear layer was employed to obtain a final vector indicating whether the gland was a ghost gland or not.

Different network architectures (SqueezeNet, [A36] resnet18, resnet34, resnet50[A37]), balanced sampling strategies, learning rate, and learning-rate decreasing policy were carefully assessed to obtain the best performance of the network on the validation dataset. The best performance of the network over the validation set employed the resnet18 backbone, class-aware balanced sampling strategy, [A33] initial learning rate 1e-3, 90 epochs in total, with learning rate decrease when the validation loss was not decreasing for 20 epochs.

Local Morphological Features

Based on individual Meibomian gland segmentations, local morphological features were quantitively measured. Morphological features included average gland local contrast, gland length, width and tortuosity.

Gland Local Contrast

Adopting an approach used in previous research, [A7] the surrounding region of each Meibomian gland in an image was obtained by extending the gland region border by 10 pixels using binary dilation. [A38] Surrounding regions are highlighted in color in Figure 4 of the paper for better visualizations. Gland pixels were also excluded if they overlapped with surrounding pixels. The average gland local contrast was defined as the difference in the average intensity of a gland region in the image and its surrounding area with negative values set to 0. Figure 4 of the paper demonstrates the process of Meibomian gland local contrast calculation with examples.

Gland Length

Gland length was calculated from the gland region using locally defined principal curves. [A39] It is important to note that these curves are different from the linear principal curves used for Meibomian gland completion where only linear direction was desired. Locally defined principal curves utilize density estimation with gaussian mixture models to represent Meibomian gland regions, and to appropriately represent the region geometry. For each gland region, corresponding principal curves are shown in Figure 5 of the paper. They usually pass through the middle of the gland and fit the geometric shape of each Meibomian gland. Gland length was defined as the product of millimeters-per-pixel and the number of pixels of each principal curve.

Millimeters-per-pixel resolution in this set of images was 0.0235, which was measured with the image of a ruler captured by the imaging instrument.

Gland Width

The average gland width was defined as the gland area (the number of pixels in a Meibomian gland region) divided by the gland length. The number of pixels was also multiplied by the millimeters-per-pixel conversion factor to obtain gland width in millimeters.

Gland Tortuosity

Convexity defect was used to represent the overall tortuosity of a gland. As in Figure 6 of the paper, a convex hull (outlined in blue) of a Meibomian gland region (outlined in red) was found for each Meibomian gland region in the image. The cavity (marked in green) of a gland is defined as the area belonging to the convex hull area but not the gland region area. The convexity defect is thus defined as the ratio of the cavity area to the convex hull area. A larger convexity defect indicates a more tortuous gland. Convexity defect is referred to as the more commonly used clinical term “tortuosity” for convenience in this study.

Gland Detection: Evaluation Metrics and Results

Evaluation Metrics

Detection performance is a key indicator of gland instance segmentation performance. Precision and recall were used as a standard metric for gland detection in this study. Specifically, an attempted match was made between each predicted gland region and a ground truth gland region. A match between the prediction and the ground truth was established for the highest mean

intersection over union (IU), provided that the mean intersection over union surpassed a threshold of 50%. Unmatched predicted glands were considered to be false positives while unmatched ground truth glands were considered to be false negatives, thereby permitting calculation of precision and recall. More details about evaluation metrics are available in the documentation for the Microsoft Common Objects in Context dataset. [A40]

REFERENCES

A1. Baudouin C, Messmer EM, Aragona P, et al. Revisiting the Vicious Circle of Dry Eye Disease: A Focus on the Pathophysiology of Meibomian Gland Dysfunction. Br J Ophthalmol 2016;100:300-6.

A2. Markoulli M, Duong TB, Lin M, Papas E. Imaging the Tear Film: A Comparison Between the Subjective Keeler Tearscope-Plus™ and the Objective Oculus® Keratograph 5M and LipiView® Interferometer. Curr Eye Res 2018;43:155-62.

A3. Pult H, Nichols JJ. A Review of Meibography. Optom Vis Sci. 2012;89:E760-9.

A4. Pult H, Riede-Pult BH. Non-Contact Meibography in Diagnosis and Treatment of Non-Obvious Meibomian Gland Dysfunction. J Optom 2012;5:2-5.

A5. Mathers WD, Shields WJ, Sachdev MS, et al. Meibomian Gland Morphology and Tear Osmolarity: Changes with Accutane Therapy. Cornea 1991;10:286-90.

A6. Hykin PG, Bron AJ. Age-Related Morphological Changes in Lid Margin and Meibomian Gland Anatomy. Cornea 1992;11:334-42.

A7. Yeh TN, Lin MC. Repeatability of Meibomian Gland Contrast, a Potential Indicator of Meibomian Gland Function. Cornea 2019;38:256-61.

A8. Daniel E, Maguire MG, Pistilli M, et al. Grading and Baseline Characteristics of Meibomian

Glands in Meibography Images and their Clinical Associations in the Dry Eye Assessment and Management (DREAM) study. Ocul Surf 2019;17:491-501.

A9. Optician online. Callinor S. How to Treat Miscros Meibomian Gland Dysfunction; June 2, 2015. Available at: https://www.opticianonline.net/features/picturing-meibomian-gland-dysfunction. Accessed April 20, 2020.

A10. Krizhevsky A, Sutskever I, Hinton GE. ImageNet Classification with Deep Convolutional NeurIPS Proceedings 2012:1097-1105. Available at: http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks. Accessed February 8, 2019.

A11. Ronneberger O, Fischer P, Brox T. 3D U-Net: Convolutional Networks for Biomedical Image Segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer; 2015:234-41.

A12. Chen LC, Papandreou G, Kokkinos I, et al. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. IEEE Trans Pattern Anal Mach Intell 2018;40:834-48.

A13. Wang J, Yeh TN, Chakraborty R, Yu SX, Lin MC. A Deep Learning Approach for Meibomian Gland Atrophy Evaluation in Meibography Images. Transl Vis Sci Technol 2019;8:37.

A14. Prabhu SM, Chakiat A, Shashank S, et al. Deep Learning Segmentation and Quantification of Meibomian Glands. Biomed Signal Process Control 2020;57:101776.

A15. Schindelin J, Arganda-Carreras I, Frise E, et al. Fiji: An Open-Source Platform for Biological-Image Analysis. Nat Methods 2012;9:676-82.

A16. Pal NR, Pal SK. A Review on Image Segmentation Techniques. Pattern Recognit

1993;26:1277-94.

A17. Malik J, Arbeláez P, Carreira J, et al. The three R's of Computer Vision: Recognition, Reconstruction and Reorganization. Pattern Recognit Lett 2016;72:4-14.

A18. Romera-Paredes B, Torr PH. Recurrent Instance Segmentation. In: European conference on computer vision. Berlin: Springer Verlag; 2016:312-29.

A19. Liu S, Qi L, Qin H, Shi J, et al. Path Aggregation Network for Instance Segmentation. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society; 2018:8759-68.

A20. Fathi A, Wojna Z, Rathod V, et al. Semantic Instance Segmentation via Deep Metric Learning. March 2017. Available at: https://arxiv.org/pdf/1703.10277.pdf. Accessed March 3, 2020.

A21. De Brabandere B, Neven D, Van Gool L. Semantic Instance Segmentation with A Discriminative Loss Function. August 2017. Available at: https://arxiv.org/pdf/1708.02551.pdf. Accessed March 3, 2020.

A22. Long J, Shelhamer E, Darrell T. Fully Convolutional Networks for Semantic Segmentation. Available at: https://www.cv-foundation.org/openaccess/content_cvpr_2015/app/2B_011.pdf. Accessed March 28, 2018.

A23. Hanley JA, McNeil BJ. The Meaning and Use of the Area Under a Receiver Operating Characteristic (ROC) curve. Radiology 1982;143:29-36.

A24. Celik T, Lee HK, Petznick A, Tong L. Bioimage Informatics Approach to Automated Meibomian Gland Analysis in Infrared Images of Meibography. J Optom 2013;6:194-204.

A25. Kim YS, Hwang K. Shape and Height of Tarsal Plates. J Craniofac Surg 2016;27:496-7.

A26. Sun W, Nasraoui O, Shafto P. Evolution and Impact of Bias in Human and Machine

Learning Algorithm Interaction. PLoS One 2020;15(8 August):e0235502.

A27. Yoo YS, Na KS, Byun YS, et al. Examination of Gland Dropout Detected on Infrared Meibography by Using Optical Coherence Tomography Meibography. Ocul Surf 2017;15:130-8.

A28. Arita R, Itoh K, Inoue K, Amano S. Noncontact Infrared Meibography to Document Age-Related Changes of the Meibomian Glands in a Normal Population. Ophthalmology 2008;115:911-5.

A29. Finis D, Ackermann P, Pischel N, et al. Evaluation of Meibomian Gland Dysfunction and Local Distribution of Meibomian Gland Atrophy by Non-contact Infrared Meibography. Curr Eye Res 2015;40:982-9.

A30. Hastie T, Stuetzle W. Principal curves. J Am Stat Assoc 1989;84:502-16.

A31. Snell J, Swersky K, Zemel TR. Prototypical Networks for Few-Shot Learning. In: Proceedings of the 31st International Conference on Neural Information Processing Systems 2017;4080-90.

A32. Liu Z, Miao Z, Zhan X, Wang J, Gong B, Yu SX. Large-Scale Long-Tailed Recognition in an Open World. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2019:2532-41.

A33. Shen L, Lin Z, Huang Q. Learning Deep Convolutional Neural Networks for Places2 Scene Recognition. ArXiv 2015. Available at: 1512.05830v1.pdf (arxiv.org). Accessed July 14, 2021.

A34. Visin F, Romero A, Cho K, et al. ReSeg: A Recurrent Neural Network-Based Model for Semantic Segmentation. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. IEEE Computer Society; 2016:426-33.

A35. Newell A, Yang K, Deng J. Stacked Hourglass Networks for Human Pose Estimation. In: European Conference on Computer Vision. Berlin: Springer Verlag ;2016:483-99.

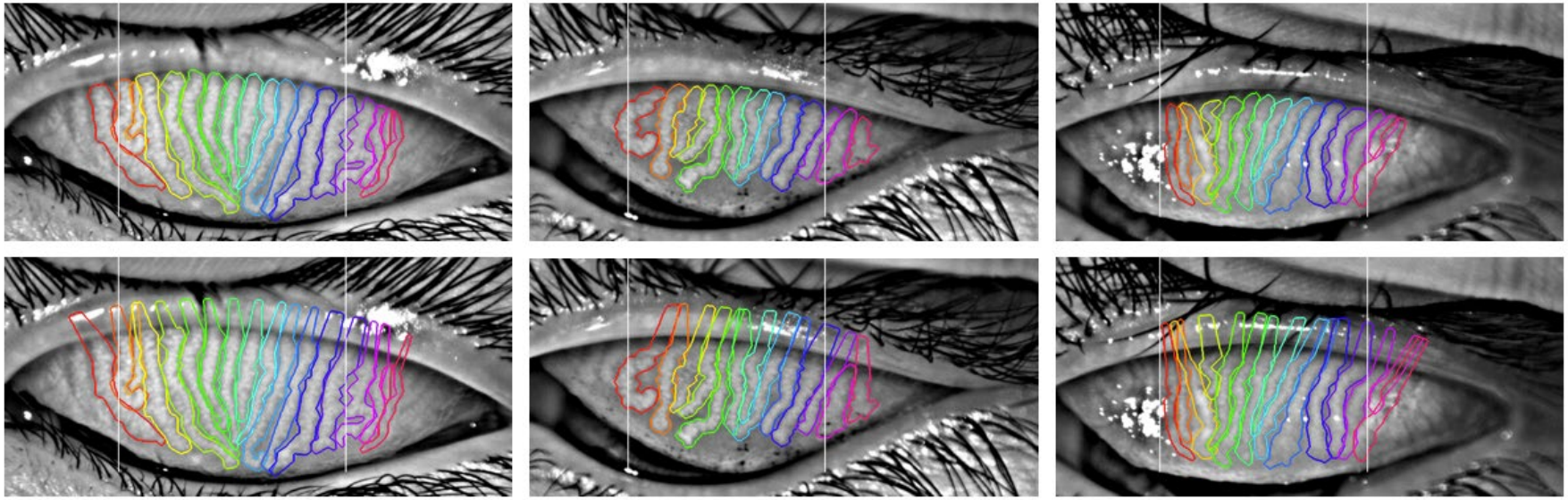
A36. Iandola FN, Han S, Moskewicz MW, et al. SqueezeNet: AlexNet-Level Accuracy with 50x Fewer Parameters and <0.5MB Model Size. Available at: https://arxiv.org/pdf/1602.07360.pdf. Accessed March 3, 2020.

A37. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2016. IEEE Computer Society; 2016:770-8.

A38. JI L, Piper J, Tang JY. Erosion and Dilation of Binary Images by Arbitrary Structuring Elements Using Interval Coding. Pattern Recognit Lett 1989;9:201-9.

A39. Ozertem U, Erdogmus D. Locally Defined Principal Curves and Surfaces. The Journal of Machine Learning Research 2011;12:1249-86.

A40. Lin TY, Maire M, Belongie S, et al. Microsoft COCO: Common Objects in Context. In: European Conference on Computer Vision 2014; 2014:740-55. Available at: LinECCV14coco.pdf (microsoft.com). Accessed July 14, 2021.



Appendix Figure A1. Processing of Meibomian gland (MG) region annotations: Upper eyelid gland annotation completion and center crop. The first and the second row correspond to original MG region annotations and MG regions after extension, respectively (see the supplementary for details). Each column corresponds to an individual example eyelid. The boundaries of the cropped regions are marked in white. Only imaged MGs falling into the cropped region were retained although all MG region annotations are shown. (Top row) To avoid annotation variability, only the visible part of each MG was originally annotated. A principal curve was fit to the (visible) annotated MG region to obtain its linear direction. Annotated MG regions were then extended along the linear principal curve direction to the upper border of the annotated eyelid region, which was defined at the MG orifice. (Bottom row) Far nasal and temporal gland annotations were not considered in this study due to their ambiguity. Based on the eyelid region annotations, imaged glands falling in the center region (outlined in white) of width equaling to 50% eyelid width were retained while rest were discarded. Glands falling on the boundaries were also retained.