AI at the forefront of the eye: Triaging tool for confocal microscopy images of human cornea

Vlada Rozova[1,2] VLADA.ROZOVA@RMIT.EDU.AU
Kh Tohidul Islam[2] KHTOHIDUL@UNIMELB.EDU.AU
Laura E Downie[3] LDOWNIE@UNIMELB.EDU.AU
Holly Chinnery[3] HOLLY.CHINNERY@UNIMELB.EDU.AU
Karin Verspoor[1,2] KARIN.VERSPOOR@RMIT.EDU.AU
[1] *School of Computing Technologies, RMIT University, Melbourne, Australia*
[2] *School of Computing and Information Systems, University of Melbourne, Melbourne, Australia*
[3] *Department of Optometry and Vision Sciences, Melbourne School of Health Sciences, Faculty of Medicine, Dentistry and Health Sciences, University of Melbourne, Melbourne, Australia*

Editors: Under Review for MIDL 2022

Abstract

Corneal confocal microscopy is used in both ophthalmology and neurology to identify and monitor the immunological and neural effects of ocular and systemic diseases. However, its use in research and clinical settings is limited by the lack of reliable, time-efficient methods to process acquired data. A typical imaging session yields a stack of images varying in quality and field of view that require careful filtering prior to further analysis. Here, we present a framework for automated quality assessment and selection of distinct human corneal confocal microscopy images suitable for downstream analysis.

Keywords: image quality assessment, medical imaging, convolutional neural networks, Fourier transform.

1. Introduction

The cornea is the only tissue where sensory nerves and immune cells can be non-invasively imaged in humans, using in vivo confocal microscopy. A typical confocal imaging session will yield a few hundred images per patient, varying in quality (focusing consistency, degree of artefact) and overlapping in the field-of-view. Currently, these images are manually filtered and analysed to characterise populations of immune cells, including their morphology and density. Reliable quantification of such cellular features requires selection of a high-quality non-overlapping image set of manageable size (Vagenas et al., 2012). Yet such a selection process is time-consuming, poorly reproducible, and subject to bias (De Silva et al., 2017).

We report on the development of a framework that takes as input a stack of raw images acquired during a confocal microscopy session and outputs a predefined number of high-quality images capturing non-overlapping regions to be used for downstream analysis.

2. Methods

Workflow Our proposed workflow consists of several steps. In step 1, a stack of raw images is loaded from a folder of images corresponding to a single patient at a given visit.

In step 2, the image quality score for each image is predicted, and converted to a class label. Images predicted as "low-quality" are discarded. In step 3, the remaining high-quality images are clustered, by computing a matrix of pairwise distances; similar images are assigned to the same cluster. For each cluster, the image with the highest quality score is retained. Finally, in Step 4, we randomly select and return the number of images requested by the user or the full constructed subset, whichever is smaller.

Image set The in vivo confocal microscopy images (400 x 400 um) of human corneas were acquired as part of clinical research studies approved by the University of Melbourne Human Research Ethics Committee and with informed consent obtained from all participants. When the image quality was deemed acceptable for downstream analysis, an image was labelled as high-quality. Our image set contained 4517 confocal images of central cornea obtained from 15 individuals with 66% high-quality images.

Image quality prediction and selection (step 2) To train a model for image quality prediction, we randomly allocated image sets from 11, 3, and 1 patient(s) into training, validation and test sets, respectively, ensuring an equal proportion of positive samples in all sets. To improve image contrast, we performed simple contrast stretching, re-scaling the image to saturate pixel values falling within the 1^{st} and 99^{th} percentiles. We used the VGG-16 deep learning architecture (Simonyan and Zisserman, 2015), replicating each grayscale image to create three channels for model input. We trained the network from scratch using Stochastic Gradient Descent with Momentum optimiser with cross-entropy loss and a learning rate of $1e^{-3}$. To convert predicted probabilities to class labels, we selected a threshold value by optimising for F_1 score on the validation set. For model comparison and evaluation, we computed ROC AUC, Precision, and Recall.

Image clustering (step 3) To develop the clustering method, we used subsets of high-quality images from two randomly selected patients. All images were reviewed and organised into separate folders based on their visual similarity. The training data consisted of 255 images split into 11 groups, while the test set included 272 images organised into 13 groups. To each image $I(x, y)$, we applied contrast stretching and sample-wise centring followed by a 2D Fourier transform $F(u, v)$. We defined the distance between two images as the Frobenius norm of the differences between spectral amplitudes: $d(I_1, I_2) = \|abs(F_1) - abs(F_2)\|_F$. The resulting matrix of pairwise distances D was normalised to contain values between 0 and 1. To perform image clustering, the pre-computed matrix D is input to the DBSCAN clustering algorithm (Ester et al., 1996). The maximum distance between two neighbours, ϵ, was optimised using the training data by maximising the Recall at Precision = 1.0. This clustering was evaluated by calculating macro-Precision and Recall.

3. Results

Figure 1 depicts examples of confocal microscopy images of human cornea used in the development of the proposed triaging tool.Images A-C show high-quality vs. low-quality images. The optimal threshold value was determined to be equal to 0.42. The final model achieved ROC AUC of 0.90 with Precision = 0.84 and Recall = 0.89. Images D-F present examples of images allocated into different groups based on their visual similarity. Clustering in step 3 with $\epsilon = 0.56$ achieved a macro-Precision = 1.0 (+/- 0.0) and Recall = 0.97 (+/- 0.04).

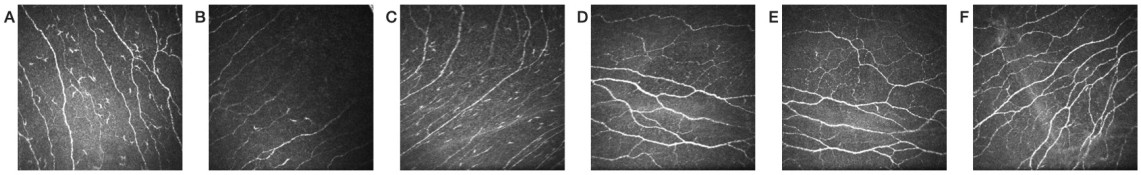

Figure 1: Example of confocal microscopy images: (A) a high-quality image; (B-C) low-quality images due to poor focus and image "smearing"; (D-E) a pair of highly similar images; (F) a dissimilar image from the same patient.

4. Discussion

For image quality prediction, we compared several commonly used deep learning models with VGG16 giving the best result despite having a shallower architecture. We favoured a solution yielding higher Precision since discarding some high-quality images is preferred over retaining low-quality images for downstream analysis.

Measuring image similarity is especially challenging; human visual inspection simultaneously analyses a range of criteria to determine whether two frames show the same region or not, including differences in brightness, contrast, or focus, as well as observing that one image is a shifted, rotated or non-linearly transformed version of the other. Our preliminary results suggest that comparing spectral amplitudes to generate a matrix of pairwise distances produce similar results to training a Siamese neural network (Koch et al., 2015) yet does not require extensive training.

References

Manikkuwadura Eranda Harshan De Silva, Alexis Ceecee Zhang, Amalia Karahalios, Holly Rose Chinnery, and Laura Elizabeth Downie. Laser scanning in vivo confocal microscopy (IVCM) for evaluating human corneal sub-basal nerve plexus parameters: protocol for a systematic review. *BMJ Open*, 7(11):e018646, 2017. ISSN 2044-6055.

Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. KDD'96, page 226–231. AAAI Press, 1996.

Gregory Koch, Richard Zemel, and Ruslan Salakhutdinov. Siamese neural networks for one-shot image recognition. In *32nd International Conference on Machine Learning*, 2015.

Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In *3rd International Conference on Learning Representations*, 2015.

Dimitrios Vagenas, Nicola Pritchard, Katie Edwards, Ayda M. Shahidi, Geoff P. Sampson, Anthony W. Russell, Rayaz A. Malik, and Nathan Efron. Optimal image sample size for corneal nerve morphometry. *Optometry and Vision Science*, 89(5):812–817, 2012.