Peer Review File

Article information: https://dx.doi.org/10.21037/atm-21-6736

First External Peer Review

Reviewer A

Congratulations for developing this time-effective algorithm.

I recommend following changes in words, terminology, language etc. for better understanding and clarification of the subject discussed in the manuscript:

Comment 1: Page 4 Line 59 sensitivity (agreement)

Reply 1: We appreciate your advice on language expression. We still prefer the word "agreement" because it better expresses the meaning of “agreement between the two measurements” and we use it consistently throughout the manuscript.

Changes in the text: N/A

Comment 2: Page 4 Line 64 deep learning algorithm

Reply 2: Thank you very much. We are willing to modify our text according to your suggestion.

Changes in the text: We have modified our text as advised (see Page 4, line 73).

Comment 3: Page 5 Line 79 etc. (and so on)

Reply 3: Thank you very much. We are willing to modify our text according to your suggestion.

Changes in the text: We have modified our text as advised (see Page 5, line 88).

Comment 4: Page 6 Lines 107-8 basic fact (ground truth)

Reply 4: Thank you for your suggestion. We still prefer "ground truth" as it conveys the meaning better.

Changes in the text: N/A

Comment 5: Page 7 Line 124 compliance (line)

Reply 5: Thank you very much. We are willing to modify our text according to your suggestion.

Changes in the text: We have modified our text as advised (see Page 7, line 135).

Comment 6: Page 8 Line 134 power/2).

Reply 6: Thank you very much. We are willing to modify our text according to your suggestion.

Changes in the text: We have modified our text as advised (see Page 8, line 145).

Comment 7: Page 10 Line 196 convoluted (convolved)

Reply 7: Thank you very much. We are willing to modify our text according to your suggestion.

Changes in the text: We have modified our text as advised (see Page 11, line 208).

Comment 8: Page 10 Line 197 basic fact (ground truth) edge map

Reply 8: Thank you for your suggestion. We still prefer "ground truth" as it conveys the meaning better.

Changes in the text: N/A

Comment 9: Page 12 Line 222 similarity/ overlapping (agreement)

Reply 9: Thank you very much. We are willing to modify our text according to your suggestion.

Changes in the text: We have modified our text as advised (see Page 12, line 239).

Comment 10: Page 12 Line 240 shows (showed)

Reply 10: We appreciate your attention to details. We have corrected the language mistake.

Changes in the text: We have modified our text as advised (see Page 13, line 257).

Comment 11: Page 13 Line 248 basic fact (ground truth)

Reply 11: Thank you for your suggestion. We still prefer "ground truth" as it conveys the meaning better.

Changes in the text: N/A

Comment 12: Page 14 Line 272 there are (there're) some studies that

Reply 12: Thank you very much. We are willing to modify our text according to your suggestion.

Changes in the text: We have modified our text as advised (see Page 14, line 289).

Comment 13: Page 15 Line 291 inaccuracy (error)

Reply 13: We appreciate your attention to details. We are willing to modify our text according to your suggestion.

Changes in the text: We have modified our text as advised (see Page 15, line 309).

Comment 14: Page 15 Line 293 a (an)

Reply 14: We appreciate your attention to details. We have corrected the language mistake.

Changes in the text: We have modified our text as advised (see Page 15, line 310).

Comment 15: Page 15 Line 295 corresponding (coordinating)

Reply 15: Thank you very much. We are willing to modify our text according to your suggestion.

Changes in the text: We have modified our text as advised (see Page 15, line 312).

Comment 16: Page 15 Line 306 deep learning algorithm

Reply 16: We appreciate your attention to details. We are willing to modify our text according to your suggestion.

Changes in the text: We have modified our text as advised (see Page 16, line 324).

Comment 17: Page 16 Line 318 learnt (trained)

Reply 17: Thank you very much. We are willing to modify our text according to your suggestion.

Changes in the text: We have modified our text as advised (see Page 16, line 336).

Comment 18: If you feel 'ground truth' conveys the meaning better, then continue using it in the text. I feel you wish to say 'basic fact' instead.

Reply 18: Thank you for your suggestion. We still prefer "ground truth" because it can more accurately convey the meaning in this article, and this term is widely used in articles related to deep learning algorithm. Therefore, comment 4,8,11 were not revised.

Changes in the text: N/A

Reviewer B

Comment 1: Line 87 – Choroidal thickness has never been used as a screening tool in any disease. It varies with multiple ocular and systemic diseases and conditions, and even with day time. Its thinning or thickening has been associated with different pathologies, as being part of its pathogenesis. However, it has never been used as a screening tool because no normative database has ever been published and a thinned choroid may be attributed to multiple causes. Using choroidal thickness as a diagnostic tool for high myopia does not make much sense because high myopia is not defined according to choroid, but to axial length or refractive error.

Reply 1: As the Reviewer suggested, choroidal thickness could not be used as a diagnostic tool for high myopia and we agree with that. This paper is not intended to use choroidal thickness as a diagnostic tool for high myopia, but develop a better algorithm to segment the choroid in highly myopic eyes. The importance of monitoring choroidal thickness in highly myopic eyes has been discussed in many previous studies and growing evidence indicated that the choroid plays a critical role in the pathophysiology of myopia. Therefore, when high myopia was already diagnosed by axial length or refractive error, choroidal thickness could be an indicator for predicting myopic progression at an early stage. And the cut-off value of choroidal thickness could be used for differentiating pathological myopia from high myopia and classify myopic maculopathy.

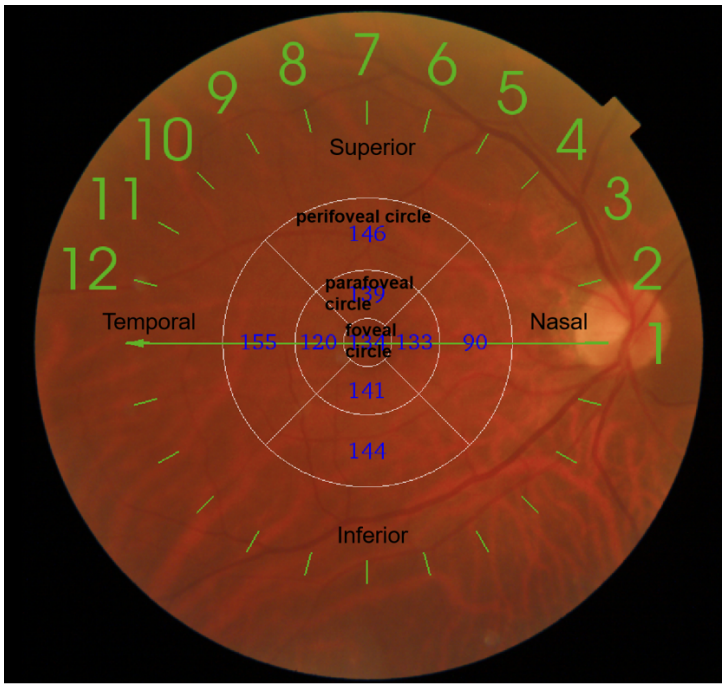
Thanks for your advice, which makes us find that the expression of the article is ambiguous. We have modified our text to “limiting its use as a potential indicator for monitoring disease progression in HM patients and for mass population screening” as advised. After this modification, the “tool” in our paper all represents the “GCS-Net”. In order to prevent your confusion caused by the imprecise expression of the conclusion in the abstract, we have also modified the conclusion to “The GCS-Net proposed in our study provides a reliable and fast tool to quantify ChT in HM patients and could potentially be used as a tool for monitoring ChT in ocular diseases related to the choroid.”.

Changes in the text: We have modified our text as advised (see Page 4, line 70-72; Page 5, line 95-96).

Comment 2: Line 141 – Choroidal thickness was measured using the ETDRS grid. Sectors in this grid are wide enough so that multiple points can be measured inside. What was the concrete point that was measured in each ETDRS sector?

Reply 2: As stated in our article, the ETDRS grid has nine distinct regions, each of which corresponds to multiple 12-line radial B-scans. We averaged the thicknesses of the corresponding choroidal regions for all images of the same region. For example,

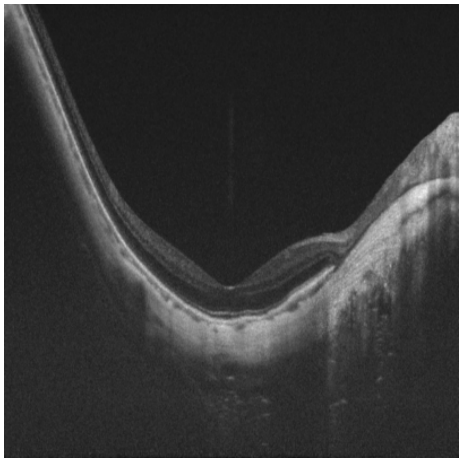
to measure the nasal region in the parafoveal circle, images 1-4 and 10-12 are selected and the thickness of the nasal area in the parafoveal circle is the average thickness of the choroid corresponding to this region in these 7 images. This choroidal calculation method is consistent with the calculation principle of the built-in Topcon software.



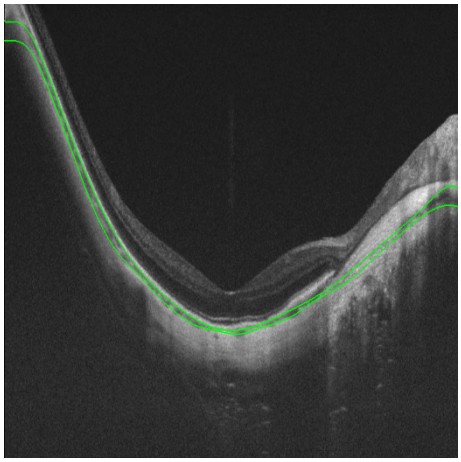
Changes in the text: We have modified our text to "the average ChT" to convey the meaning better (see Page 8, line 158).

Comment 3: Line 146 – The Topcon Atlantis DR-1 SS-OCT includes automatic segmentation of retina and choroid, and it gives automatic measurements of choroidal thickness. Automatic measurements have been proved to be more accurate than manual ones. Were these automatic measurements considered anytime?

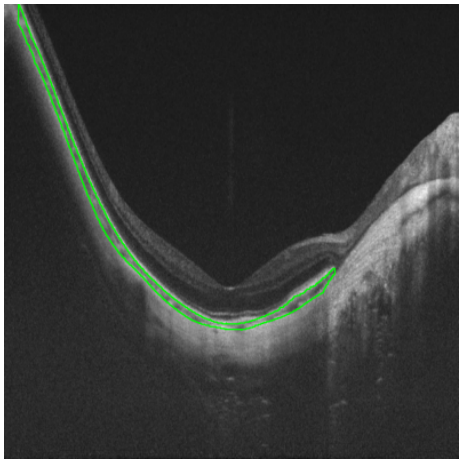
Reply 3: Although the Topcon Atlantis DR-1 SS-OCT has included automatic segmentation of the choroid, its accuracy in high myopia is limited. For example, the figure below shows that the automatic segmentation by the built-in Topcon software is not satisfied enough in some highly myopic eyes. All the automatic segmentation obtained with the built-in Topcon software were checked before manual segmentation. Actually, our proposed GCS-Net could obtain better results on choroidal segmentation than the built-in Topcon software.



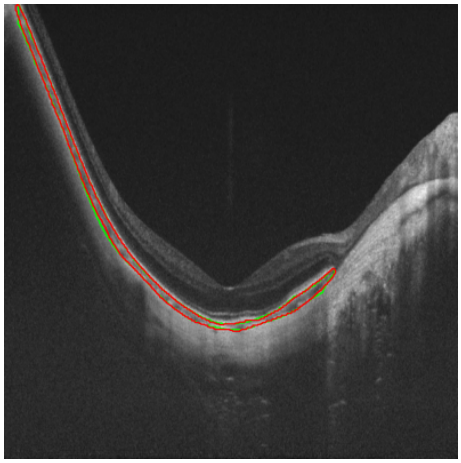
Original B-scan



Automatic segmentation by the built-in Topcon software (green outline)



Ground truth (green outline)



Automatic segmentation by the proposed GCS-Net (red outline)

Changes in the text: N/A

Comment 4: Line 150 – Were manual measurements performed on the same OCT with its caliper or were they performed with an external program?

Reply 4: On the same OCT with its caliper.

Changes in the text: N/A

Comment 5: Line 154 – What was labeled, choroidal thickness or diagnosis?

Reply 5: The boundaries of choroid, which are Bruch's membrane and the choroidal-scleral interface, were labeled.

Changes in the text: We added this sentence in Methods (see Page 8, line 152-153).

Comment 6: Line 155 – "by a retina specialist (YF) who checked the final segmentation at least once" –Were all images manually segmented or automatic

segmentation was used? In case that manual segmentation was performed, could you explain how this was done?

Reply 6: All manually segmented images, but not the automatically segmented images, were checked by a retina specialist (YF). The automatic segmentation was evaluated in the validation dataset by IoU, DSC, sensitivity and specificity. And the calculation of choroidal thickness after automatic segmentation was compared with manual segmentation in the test dataset. Actually, all the images used in this paper have been labeled manually. However, the automatic segmentation by GCS-Net did not referee the results of human. The manual segmentation plays two roles: one is as the material of the training dataset, and the other is as ground truth for comparison with automatic segmentation. Thanks for your kind advice, which makes us find that the expression of the article is not accurate enough.

Changes in the text: We have modified our text to "by a retina specialist (YF) who checked the manual segmentation at least once" (see Page 9, line 166).

Comment 7: Line 202 – Positive or negative means that the patient has or has not a disease or condition, that to say, we are considering a qualitative variable. Nevertheless, choroidal thickness outcomes are always a quantitative variable. What was considered as true positive or false positive?

Reply 7: Thank you very much for your review and comments. For your question, the true positive and false positive here are not for the choroidal thickness, but the evaluation metrics of the automatic segmentation of the AI system. Our evaluation metrics rely on the computation of these values. For image segmentation, true positive represents the number of pixels predicted as foreground by automatic segmentation and labeled as foreground in the ground truth, false positive represents the number of pixels predicted as foreground but labeled as background in the ground truth, and false negative represents the number of pixels predicted as background but labeled as foreground in the ground truth. In choroid segmentation, foreground means the choroid regions while background means non-choroid regions. We also added another figure to illustrate (new Figure 2).

Changes in the text: We have revised the text in the Method section (see Page 11, line 220-222) and added another figure (see Page 26, line 525-535 and new Figure 2).

Comment 8: Line 220 – Correlations are worthless in this type of studies. It is expected that correlations are found because both measurements are made on the same images. This analysis is not really appropriate here.

Reply 8: Thanks for your suggestion. Correlation analysis was intended to study the correlation between two measurements. We accept that correlations are worthless in

this paper and delete all the relevant content. Original Figure 4 was also deleted and the ICC data was added in Table 3.

Changes in the text: We deleted all the contents of correlation analysis and added some data in Table 3 (see Page 3, line 61-62; Page 4, line 67; Page 12, line 235-242; Page 13, line 273; Page 27, line 547-549; Table 3).

Comment 9: Results and discussion is based on the basis that choroidal thickness can be used as the only tool for diagnosing high myopia. Therefore, the outcomes about sensitivity, specificity and so on. If we accept this, we could only state that this new method in artificial intelligence is good for differentiating patients below or above 26mm of axial length. Choroidal thickness should not be considered a diagnosing tool for high myopia. Hence, all manuscript should be rewritten.

Reply 9: Results and discussion are based on the basis that choroidal thickness is a valuable indicator for myopia and also our proposed GCS-Net can segment the choroid of highly myopic eyes with great accuracy and efficiency. Sensitivity and specificity are used to evaluate the performance of the AI model in segmenting the choroid. Sensitivity reflects the proportion of correctly segmented foreground parts in the ground truth. Specificity reflects the proportion of correctly segmented background parts in the ground truth. The formulas used to calculate sensitivity and specificity are as follows:

$$\text{sensitivity} = \frac{\text{true positive}}{\text{true positive+false negative}} \quad \text{specificity} = \frac{\text{true negative}}{\text{true negative+false positive}}$$

Hopefully, the replies above and changes in the manuscript have made this paper easier to understand. Please contact with us if you have any questions. I am happy to make further modifications and explanations.

Changes in the text: We have revised the text in the Method section (see Page 11, line 223-225).

Second External Peer Review

Reviewer A

I congratulate the authors for their efforts revising the manuscript. Now it is far more understandable and I am sure that it will improve current knowledge about choroid in a near future. However, there are still some issues to be addressed before publication:

Comment 1: Introduction, line 94 – Screening is a term used for diagnosing, and as previously said choroidal thickness cannot be used as a diagnostic or screening tool for high myopia. It may be used for monitoring its progression but not for diagnosing. Therefore this should be deleted from the manuscript in line 94: 'and for mass population screening'.

Reply 1: Thank you very much. We are willing to modify our text according to your suggestion.

Changes in the text: These words have been deleted from the manuscript as advised (see Page 5, Line 95).

Comment 2: Methods – If automatic measurements of choroidal thickness in the ETDRS grid were not used, how many manual measurements were performed? This should be better clarified: number of measurements, locations of these measurements, using caliper, which OCT slabs, references used for distances (both vertical and horizontal).

Reply 2: The manual measurement was calculated by the built-in Topcon software with its caliper in the test dataset with 266 eyes (number of measurements), and the ETDRS grid was also used to investigate the average choroidal thickness in different regions (locations of these measurements). This calculation method has been used in many published studies. The average thickness of the corresponding choroid regions for 12 radial B-scans was calculated by converting the pixel counts into μm (locations of these measurements & OCT slabs). The image size of each B-scan is 1024 (the horizontal direction) × 992 (the vertical direction) pixels, which corresponds to a total area of 9 × 2.6 mm^2 (references used for distances). All the information has been described in the Methods section. Thanks for your kind advice, which makes us find that the expression is not clear enough. We have revised our text to "the results of automatic segmentation were compared with manual segmentation calculated by the built-in Topcon software with its caliper in the ETDRS grid."

Changes in the text: We have modified our text as advised (see Page 12, Line 237-239).

Comment 3: Methods – Please, add within the text what is considered as a true positive, a true negative, a false positive, and a false negative. Not only the explanation in figure 2.

Reply 3: Thank you very much. We are willing to modify our text according to your suggestion.

Changes in the text: We added the explanation in the manuscript as advised (see Page 11, Line 222-227).

Comment 4: I recommend the authors to give an explanation in the discussion section why automatic segmentation with internal algorithm in DR-1 Topcon SS-OCT is not correct enough. Those two pictures included in the response to reviewers are very convincible (Original B-scan / Automatic segmentation by the built-in Topcon software (green outline)).

Reply 4: Thank you for your suggestion. We considered possible reasons for the error due to over-smoothing and artifact interference and added an explanation in the Discussion section.

Changes in the text: We added an explanation in the manuscript as advised (see Page 14-15, Line 292-295).

Comment 5: Discussion – line 346. 'and also as a mass population screening tool for ocular diseases related to the choroid'. This sentence should be removed because it may be controversial.

Reply 5: Thank you very much. We are willing to modify our text according to your suggestion.

Changes in the text: This sentence has been removed as advised (see Page 18, Line 359).