Uncertainty estimation of a geographic atrophy OCT segmentation algorithm: How do we identify cases where the algorithm may be mistaken?

Luis de Sisternes, PhD[1]; Lars Omlor, PhD[2]; Warren Lewis, MS[3]; Sophie Kubach, MS[1]; Varsha Pramil, MS[4]; Harris A. Sheikh, BS[4]; Ruikang K. Wang, PhD[5]; Nadia K. Waheed, MD, MPH[4]; Niranchana Manivannan, PhD[1]

[1]Carl Zeiss Meditec, Inc., Dublin, CA, USA; [2]Corporate Research and Technology, Carl Zeiss Inc., Dublin, CA, USA; [3]Bayside Photonics, Inc., OH, USA; [4]New England Eye Center, Boston, MA, USA; [5]University of Washington, Seattle, WA, USA

Poster #3017 – F0287

PURPOSE

- We developed a geographic atrophy (GA) segmentation model for OCT data based on deep learning.
- One of the main obstacles for clinical application of this automated algorithm is understanding for which specific cases we can expect reliable results and when it is likely to fail.
- This work proposes an approach to estimate the regional uncertainty of a deep learning segmentation model.

METHODS

- We generated uncertainty maps for GA segmentation using a model ensemble trained through cross-validation (Figure 1).
- As each model in the ensemble was trained with different data, observing the variance in results across them can be used to estimate epistemic uncertainty (unseen manifestations).
- Data augmentation in the test phase simulating different signal level conditions can estimate the aleatoric uncertainty (signal levels the model is unfamiliar with).
- Analyzing both sources of uncertainty, we generated maps that relate to model confidence.
- These maps were evaluated by analyzing the segmentation performance at different levels of uncertainty compared to two expert graders (R1 and R2), excluding from the analysis those regions with uncertainty higher than a chosen level (Figure 2).

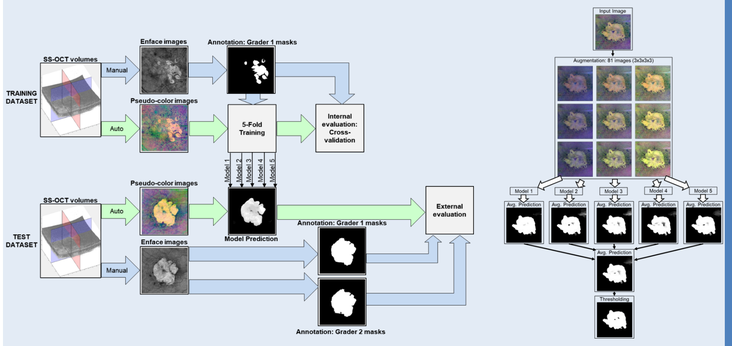
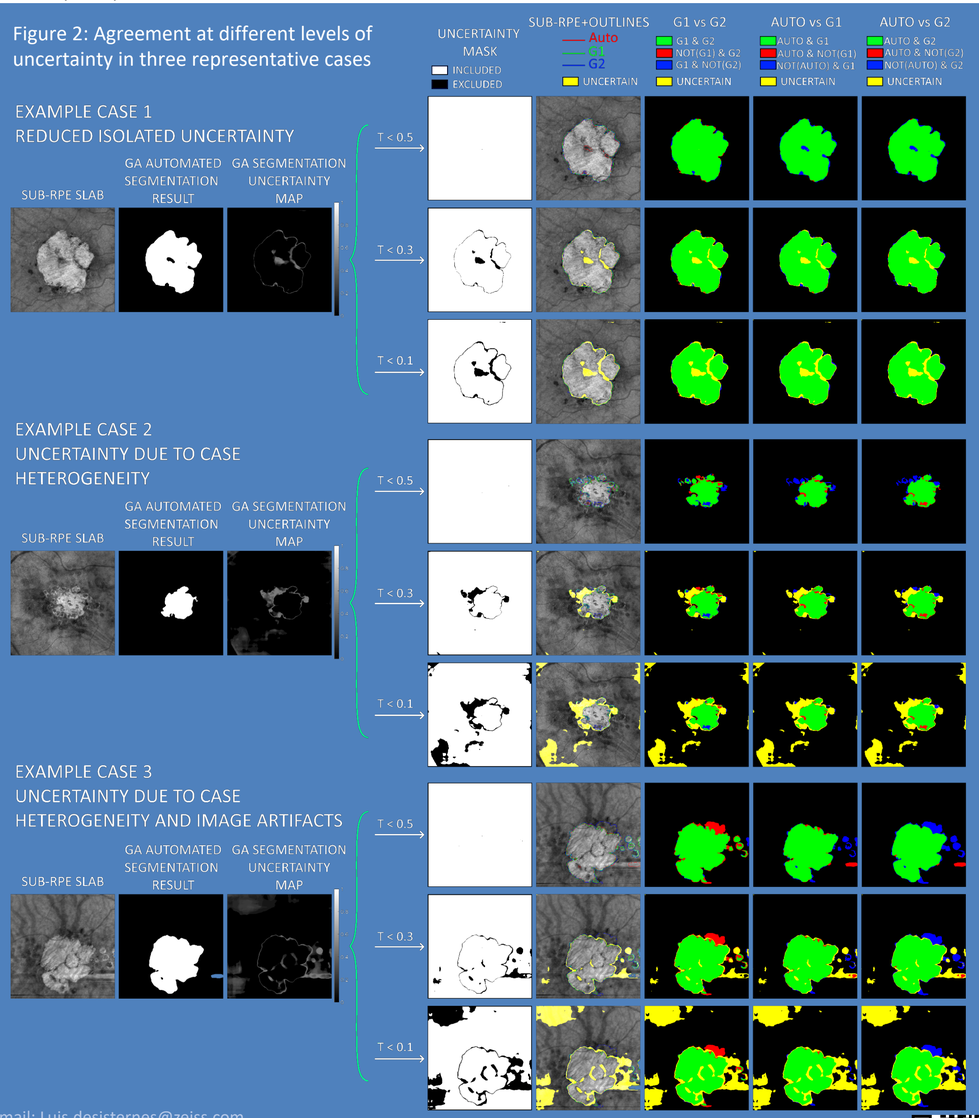


Figure 1: Model training / test workflow (left) and test augmentation (right)

Figure 2: Agreement at different levels of uncertainty in three representative cases



Email: Luis.desisternes@zeiss.com

Disclosures: LdS (E), LO (E), WL (C), SK (E), RW (C, F), NW (F), NM (E): Carl Zeiss Meditec, Inc. ; NW (C): Nidek Medical Products; NW (C): Boehringer Ingelheim; NW (C): Topcon; NW (S): Gyroscope Therapeutics; NW (F): Heidelberg; NW (F): Nidek Medical Products; NW (F): Topcon; NW (I): Ocudyne

RESULTS

- We processed 180 OCT scans (PLEX® Elite 9000, ZEISS, Dublin, CA) from GA eyes and 45 from non-GA AMD eyes.
- Regions of segmentation disagreement between the automated method and each grader had significantly higher values of uncertainty (Figure 3A).
- Automated segmentation accuracy (Dice Index) increased with lower uncertainty thresholds (excluding larger image regions from the analysis) while the comparison between two graders did not improve (Figure 3B), showing the ability of the uncertainty maps to indicate regions where the algorithm was not confident.
- Choosing a threshold producing a segmentation performance similar to the intergrader agreement (Dice=0.91) deemed as uncertain approximately 6% of the image on average while maintaining detection sensitivity.
- Uncertain regions corresponded primarily to locations without GA where the algorithm might make a false positive decision (Figure 3C).

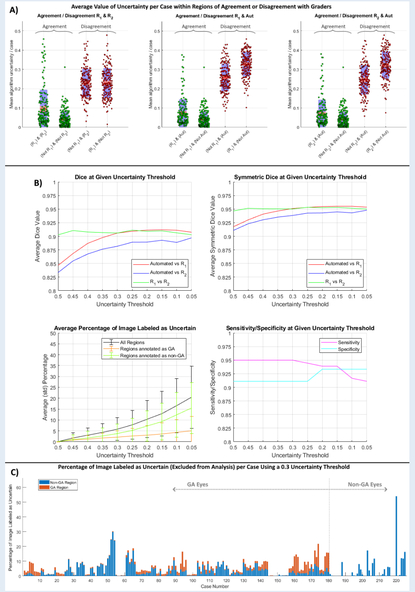


Figure 3

CONCLUSIONS

- We introduce a method to generate uncertainty maps for an automated GA segmentation model.
- These maps inform about segmentation confidence and can be used as feedback for manual review.