Context-aware Attention U-Net for the segmentation of pores in Lamina Cribrosa using partial points annotation

Nan Ding, Hélène Urien, Florence Rossant, Jérémie Sublime, Michel Paques

Context-aware Attention U-Net for the segmentation of pores in Lamina Cribrosa using partial points annotation

Nan Ding* Hélène Urien* Florence Rossant* Jérémie Sublime* Michel Paques†
* Institut Supérieur d'Electronique de Paris (ISEP), Paris, France
† Centre d'investigation Clinique 1423, INSERM & Direction de l'Hospitalisation et des Soins, Hôpital des Quinze-Vingts, Sorbonne Université, Paris, France

***Abstract*—Glaucoma is the second leading cause of blindness in the world. Although its physiopathology remains unclear, the lamina cribrosa, a 3D mesh-like structure consisting of pores, that allow the axons passing through to join the brain, has been identified as the primary site of damage. In this work we present an extended version of U-Net for pore segmentation in 2D *en-face* OCT images with partial point annotations, i.e. having only a small portion of pore locations in each image labeled. Our method combines the attention gate and the context information to address the difficulties caused by small object segmentation in low signal to noise ratio images. Experimental results show that 71.8% of the annotated pores are successfully segmented.**

***Index Terms*—Optical coherence tomography, lamina cribrosa, pore segmentation, U-Net, attention mechanism**

I. Introduction

The lamina cribrosa (LC), a 3D mesh-like structure in the optic nerve head (ONH), has been identified as the primary site of damage in glaucoma [1], the second leading cause of blindness in the world. The LC is composed of pores (i.e. axonal pathways, Fig. 1) through which the retinal axons pass to reach the brain [2], and morphological changes in pores, like increase in surface area and elongation, have been observed in glaucoma patients [3]. In-vivo observation of LC pores is now possible thanks to recent advances in optical coherence tomography (OCT) technology and our research project aims to characterize the LC pores in glaucoma as well as the subtle changes occurring during the disease.

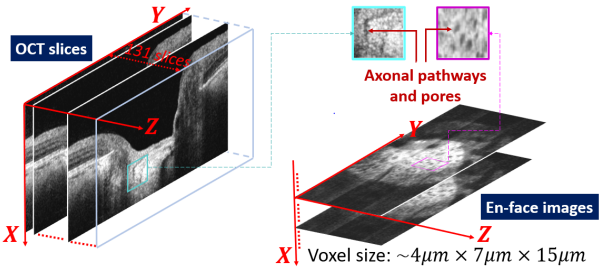


Fig. 1: 3D OCT data. *En-face* images (right) are extracted from 3D OCT slices (left). The X axis represents the depth in the ONH. Dark spots in the *en-face* images correspond to the pores of the LC.

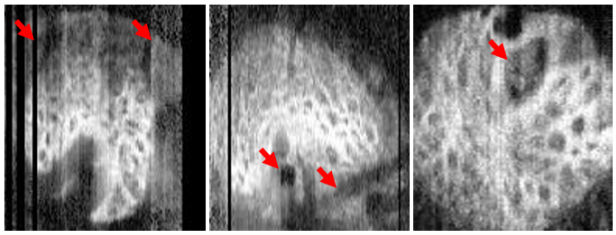
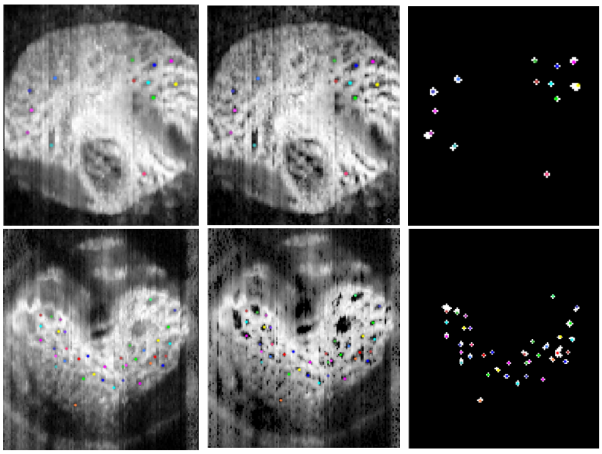


Fig. 2: Appearance variability of the *en-face* images of the LC. Pores vary in their shape, size, and location, thus they are difficult to be segmented, especially along with various artifacts shown in red arrows.

Prior in-vivo LC studies mainly focused on the total LC thickness [4], LC surface [5], and the anterior boundary of the LC [6], while only two methods have been proposed to investigate into the LC pore structure. Authors in [7] used the local thresholding technique and the 3D median filter to segment pores in 2D *en-face* images extracted from a 3D OCT volume. This method only operates on one 2D *en-face* image at a time, running the risk of not fully using the rich continuity information from the consecutive slices. In [8], pore candidates were selected as the local minima with the highest contrast in locally averaged 2D *en-face* images and axonal pathways were reconstructed from these points thanks to a tracking algorithm. However the pore detection step is not reliable, since pore features are not enough taken into account. Moreover both methods require a manual delineation of peripheral masks for the most contrasted *en-face* images in order to only process the highly identifiable regions, and such manual delineation would be time-consuming and less practical with larger datasets.

Recently, deep learning methods, especially the U-Net network [9] and its variants [10], have shown success in analyzing the LC in OCT volumes. For example, authors in [11] proposed a 3D U-Net-based network to segment the ONH tissues and thus to measure the LC thickness. Moreover, deep learning has shown its capacity to accurately segment small retinal fluid regions [12], [13] in OCT images with U-Net-based methods, despite a small dataset. Thus it is promising that U-Net variants could perform well for the pore segmentation task.

However, the above methods are developed in a fully



(a) (b) (c)

Fig. 3: (a) incomplete pore annotation. Each identified pore is marked with a point on the associated *en-face* image, by observing the continuity of local regions in consecutive images and checking the contrast. (b) pre-processed image with morphological filters to enhance pore features. (c) region growing on image (b) to produce the incomplete ground truth segmentation map.

supervised manner, which requires a large amount of data with pixel-level annotations. Obtaining such ground truth data is unrealistic for the pore segmentation as the pore contours are not very clear and there exists the ambiguity even between the experts, in addition, small size and hundreds of pores in an image make it difficult to have pixel-accurate annotation. It is thus challenging to segment the pores of the LC with the limited availability of pixel-level annotation data. As a result, the weakly supervised point annotation [14] is the most efficient for the pore segmentation task, since other methods like the the bounding boxes [15] or scribbles [16] are not suitable under dense pore intensities.

In this paper, we propose a baseline automatic LC pore segmentation model, which is, to our knowledge, the first attempt to look into this difficult problem with deep learning methods. This work is inspired from [7], [8], where the supervision of a human observer is required to reach the accuracy necessary for the medical application. Our goal is to get an accurate pore segmentation in every 2D *en-face* images, reliable enough to allow a future automatic 3D reconstruction of the axonal pathways [17].

II. Dataset

Our study population includes 17 subjects with 41 3D OCT volumes. Multiple volumes may correspond to the same eye of one patient for longitudinal studies. All OCT volumes were validated by an expert to guarantee that only patients with larger pores are recruited for the feasibility of the work, since it is difficult for some glaucoma patients to fix their sight during the examination. Volumes are acquired with the spectral-domain OCT machine (Spectralis, Heidelberg Engineering), where each acquisition centered on the ONH contains 131 2D OCT slices (Fig. 1). The theoretical transverse and axial resolutions for such a 2D slice of 496×768 pixels are about $7\mu m$ and $4\mu m$ respectively, the sampling step between two consecutive slices approaching $15\mu m$.

En-face images are extracted from the 3D OCT volume (Fig. 1) with a resolution of 131×768 pixels, and the pores were partially annotated by 2 engineers, who retained only those that show continuity in the volume and great contrast; the selected pores were marked by a single coordinate point close to the centroid (Fig. 3a). The objective of the annotation is to identify the largest pores that allow most axons to pass through since exhaustively segmenting all the pores is difficult, especially in regions with vessel shadows or scanning artifacts (Fig. 2). Finally, our dataset of 41 OCT volumes consists of 2101 *en-face* images, with an average of 8.78 ± 4.75 axonal pathways manually identified in each volume, therefore a small proportion of the totality of existing pores.

III. Segmentation with partial points annotation

Pore feature enhancement. The gray level intensity of the *en-face* images (Fig. 2a) was coded to $[0, 1]$ by linear normalization. Since pores in the *en-face* images are weakly contrasted with the surrounding tissues, we applied the following morphological filters [8] to enhance the pore features:

- Bottom-hat filter, and the complement denoted as I_{BH}, with a structuring element D_3. The objective of this filter is to enhance the contrast and leave dark areas of the image that are smaller than the structuring element.

$$I_{BH} = 1 - (I \bullet D_3 - I) \tag{1}$$

- Alternate Sequential Filter, denoted as I_{ASF} , defined by a sequence of closings and openings with increasing size of the structuring element D_i (up to $i = 3$), with morphological reconstruction by dilation or erosion at each step i. This filter aims at denoising, while retaining the main dark structures.

$$I_{open} = R^D_{I^{(i-1)}_{fas}}(I^{(i-1)}_{fas} \circ D_i), \text{ with init. } I^0_{fas} = I \tag{2}$$

$$I^{(i)}_{fas} = R^E_{I_{open}}(I_{open} \bullet D_i) \tag{3}$$

$$I_{FAS} = 1 - min(max(I^3_{fas} - I, 0), 1) \tag{4}$$

where $\circ$, and $\bullet$ denote the opening and closing operations, and both are built by using a binary structuring element with a disc of radius r, denoted as D_r. $R^D_M(I)$, $R^E_M(I)$ are respectively the morphological reconstruction by dilation and by erosion of the image I in the mask M. Such reconstruction repeats dilation or erosion operation of the image I, until the contour of I fits under the mask M. The pre-processed image $I_o(X)$ in (5) is shown in Fig. 2b, and we take $\alpha = 0.5$ to trade-off between the 2 filters to denoise and retain main pores.

$$I_o = \alpha \cdot I_{BH}(X) + (1 - \alpha) \cdot I_{FAS}(X) \tag{5}$$

Ground truth generation. The points marking the pore centroids and manually defined in the annotation phase could not be directly used for training since they represent only several pixels and there are too many false negatives in the image. To this end, a region growing algorithm was applied on I_o to generate the ground truth segmentation maps (GT, Fig. 2c): we used the partial points as seeds and we applied a region-growing algorithm to get a segmentation. The similarity criterion is the L1 distance ($DIST$) between an unallocated pixel's intensity value and the mean intensity of the current region. Among the 8-connectivity neighbors, the pixel with the smallest $DIST$ is allocated to the region if $DIST$ is less than a given threshold $DIST_{th} = 0.04$; the growing process stops when $DIST$ becomes larger than $DIST_{th}$ for all neighbors. A small threshold is used to avoid over-segmentation since some pore boundaries are barely identifiable.

Proposed Network. U-Nets [9] have been widely adopted within the medical imaging community and are known to significantly improve segmentation performances. The skip connections in U-Net helps retrieve lost spatial information at the encoder path, enabling a precise localization of the target objects. In our case, pore sizes are small, typically smaller than 5×5 pixels, given that the size of an *en-face* image is about 150×130 pixels after cropping to retain only the nonzero region. In order to further improve the detection of such small areas, pore features produced by U-Nets can be strengthened by integrating attention mechanisms [18] into the network to help capture the regions of interest (ROIs). One popular approach proposed in [19] incorporates an Attention Gate (AG) module (Fig. 5) into the U-Net. The AG allows to estimate potential areas where the pores are most likely to appear by removing feature activation in irrelevant regions, without the necessity of using explicit external ROIs as supervision. Moreover, axons pass through the ONH follow a fairly regular path: pore intensities are similar between neighboring *en-face* images, while their centroids and shapes also vary little spatially. Thus, a naive application of U-Net runs the risk of not fully using these regularity properties. To this end, we design a context-aware network by inputting 3 neighboring *en-face* images, and outputting only one segmentation map for the middle image.

The proposed context-aware attention U-Net is shown in Fig. 4. Three input images are resized to 160×160 pixels, and then they are progressively filtered by $(2\times)$ convolution blocks and down-sampled in the encoder path. The convolution block, used in both the encoder and the decoder, is composed of a convolution layer (Conv), batch normalization (BN), and rectified linear unit (ReLU). In the decoder path, each layer has an attention gate through which features from the layer l in the encoder path must pass through before concatenating to the up-sampled features in the coarser layer $(l+1)$ in the decoder path. Finally a pixel-wise softmax is applied to generate probability maps to assign each pixel the corresponding class (pore or background).

The AG module is shown in Fig 5, where the architecture is adapted from [19] for 2D pore segmentation. We define the

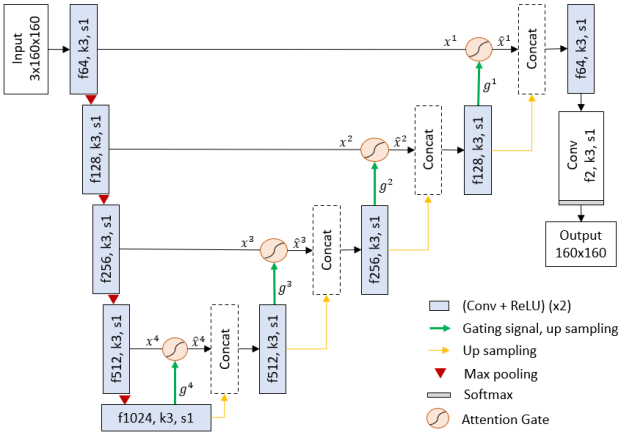


Fig. 4: Proposed context-aware attention U-Net architecture.

feature map x at pixel $i \in \{1, ..., N\}$ in layer $l \in \{1, ..., L\}$ as $x_i^l \in \mathbb{R}^{F_l}$, where F_l refers to the number of feature maps in layer l. An attention coefficient $\alpha_i^l \in [0, 1]$ is calculated by the AG to identify the ROIs. The output of AG is an element-wise multiplication $\hat{x}^l = \alpha_i^l x_i^l$.

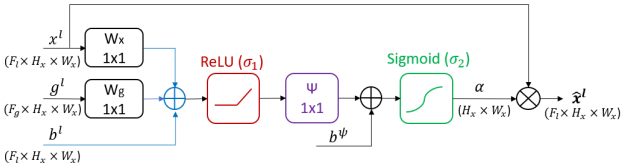


Fig. 5: Attention gate. H_x, W_x refer to the height and width of the feature map x.

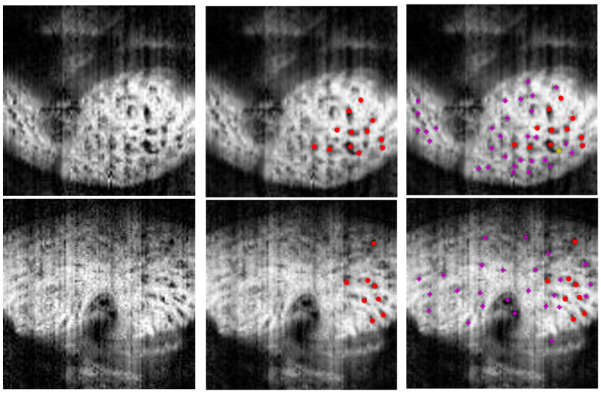
Feature maps are gradually down-sampled in the encoder to capture a large receptive field. Features on the coarse spatial grid level of layer $(l + 1)$ identify the location of target objects, and such coarse features could serve as gating signal $g^l \in \mathbb{R}^{F_g}$ to provide global information for x_i^l to disambiguate task-irrelevant feature content. Thus the additive attention coefficients α_i^l is calculated as:

$$\alpha_i^l = \sigma_2(q_{att}^l(x_i^l, g_i^l; \Theta_{att})) \tag{6}$$

$$q_{att}^l = \psi^T(\sigma_1(W_x^T x_i^l + W_g^T g_i^l + b_i^l)) + b_i^\psi \tag{7}$$

Where linear transforms $W_x \in \mathbb{R}^{F_l \times F_l}$, $W_g \in \mathbb{R}^{F_g \times F_l}$, $\psi \in \mathbb{R}^{F_l \times 1}$, and the bias $b_i^l \in \mathbb{R}^{F_l}$, $b_i^\psi \in \mathbb{R}$ together form the learnable parameter set $\Theta_{att} = \{W_x, W_g, \psi, b^l, b^\psi\}$ which characterize the AG. W_g and W_x ensure that the pixel-wise addition of x^l and g^l could be done to learn the salient regions. $\sigma_1(x)$ is the ReLU function for nonlinearity and $\sigma_2(x)$ is the sigmoid activation function for normalisation. Moreover, the linear function ψ allows to generate only one attention map α^l for all feature maps in layer l. In practice, linear transforms $\{W_x, W_g, \psi\}$ are implemented as 1×1 convolution layer.

The Generalized Dice Loss (GDL) proposed in [20] is an extension of dice loss, which is inspired from the Dice Coefficient measuring the overlap between two images. We



(a) (b) (c)

Fig. 6: Segmentation result. We extract one point for each segmented pore for the comparison of pore numbers. (a) pre-processed image. (b) partial ground truth for the evaluation. (c) our network segmentation result. The red, magenta, yellow points correspond to true positive, false positive, and the false negative pores, respectively.

TABLE I: Statistics on the pore numbers (#) of the prediction compared with the partially labeled ground truth.

	# Min	# Max	# Mean	# STD
Ground truth	1	16	7.77	3.34
Our method	8	52	33.22	10.23

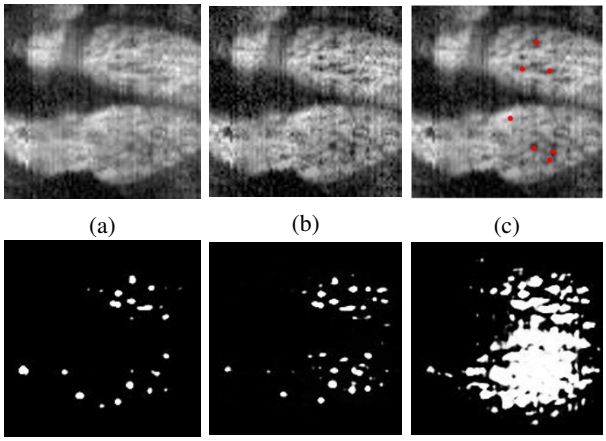
used the GDL to solve the unbalanced background/foreground problem since only few areas in our image are labeled.

$$GDL = 1 - 2\frac{\sum_{l=0}^{1} w_l \sum_{n=1}^{N} p_n g_n}{\sum_{l=0}^{1} w_l \sum_{n=1}^{N} p_n + g_n}, w_l = \frac{1}{\sum_{n=1}^{N} g_n}$$

Where $GT = \{g_1, ..., g_N | g_n \in \{0, 1\}\}$ is the GT segmentation map for an image of N pixels, and $P = \{p_1, ..., p_N | p_n \in [0, 1]\}$ is the output segmentation map. The weight w_l is used to provide in-variance to different label set properties.

IV. Experiments

Implementation details. Experiments were carried out on the dataset described in Sec II. Data augmentation is performed by randomly combining horizontal/vertical flipping, rotating, brightness and contrast changing, Gaussian noise, and elastic deformation [21]. We used a nested 4-fold cross-validation with 13 subjects for training, 2 subjects for validation and 2 subjects for testing. Since for a test fold we had an ensemble of 4 trained models, the prediction on the test fold was obtained by averaging the 2 models that achieve the highest recall and precision to increase the robustness of the model. The learning rate is initialized as 3×10^{-4}, and would be reduced by a factor of 5 if the exponential moving average of training loss did not improve by at least 5×10^{-3} within the last 30 epochs. Adam optimizer is used with a weighted decay of 3×10^{-5}, and a batch size of 32 for 500 epochs on a TITAN RTX GPU.



(a) (b) (c) (d) (e) (f)

Fig. 7: Segmentation results using different ratios of points annotation. (a) source image. (b) pre-processed image. (c) partial ground truth with the pre-processed image. (d)-(f) results using different ratios of annotated points (full, 80%, and 60% respectively.

TABLE II: Pore segmentation results using different ratio of annotation. Note that our full annotation is partial.

	Pixel-level		Object-level	
	Dice	Jaccard	Precision	Recall
Full	0.280	0.171	0.261	0.718
80%	0.192	0.104	0.166	0.714
60%	0.081	0.056	0.074	0.503

Evaluation. We used the pixel-level Dice similarity Coefficient ($Dice$) and Jaccard index ($Jaccard$) to evaluate the segmentation performance, however, as our ground truth segmentation map is incomplete and is generated by an automatic method, it is not surprising that the $Dice$ and $Jaccard$ are low for the experiments. Thus, for a more reliable evaluation, we complete the evaluation by adding the object-level recall ($Recall$) and precision ($Precision$) metrics to measure the the pore detection performance. The $Recall$ and $Precision$ are calculated by extracting local maximum points of the output probability map, while only one point is retained when multiple neighbouring points have been extracted.

As mentioned above, our ground truth is incomplete both for the training process and for the evaluation, in order to get a more comprehensive validation of the results, we randomly select 10 images from the test dataset with 194 images, and try to point out all the pores without ambiguity. For these *a posteriori* annotated images, we averaged the object-level true positives (TP), false positives (FP), false negatives (FN), and finally the $Recall$ for the predictions.

Prediction with partial points annotation. We firstly compared the segmentation results with the ground truth, as shown in Fig. 7 and Table I. We can see that our method is able to segment pores marked in the ground truth (high recall performance), even though the ground truth used for the training phase is incomplete. Moreover, pores that are not

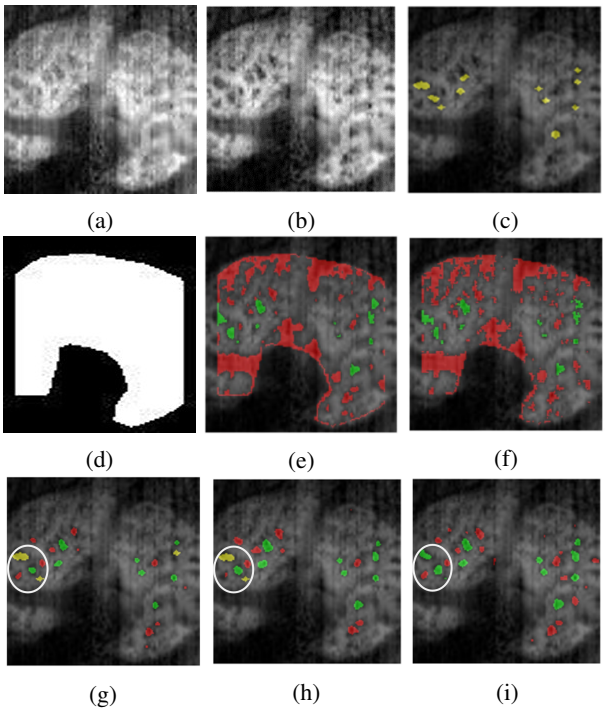


Fig. 8: An example of segmentation results. (a) source *i*mage. (b) pre-processed image. (c) incomplete ground truth mask. (d) peripheral mask used for Chan_Vese and W-Net. (e) Chan_Vese. (f) W-Net. (g) U-Net. (h) Attention U-Net. (i) our model. Green, red and yellow areas represent correctly segmented pores (TP), incorrectly classified pores (FP), and missed pores (FN) in the segmentation map respectively.

TABLE III: Segmentation results on the test dataset (194 images). We favor the $Recall$ metric since the ground truth is incomplete.

	Pixel-level		Object-level	
Model	Dice	Jaccard	Precision	Recall
Chan_Vese [22]	0.074	0.054	0.119	0.645
W-Net [23]	0.081	0.052	0.100	0.689
U-Net [9]	0.241	0.142	0.254	0.633
Attention U-Net [19]	0.250	0.145	**0.277**	0.622
Our method	**0.280**	0.171	0.261	**0.718**

identified in the ground truth are segmented as well, which means, pore features are well learned during the training even with the partial ground truth. However, it is inevitable that the model predicts some false positives in artifact areas (see Fig. 7c) since they appear like dark spots that are similar to pores.

To explore the performance of our pore segmentation method with partial ground truth when the ratio of the number of annotated axonal pathways changes, we randomly select 60% and 80% of the pathways in the ground truth and trained the network with such sub-partial points annotation dataset. The segmentation results are shown in Table II. The $Recall$ metric is comparable between the "Full" and 80% of the ground truth, but the former annotation performs much better on other metrics, which could be explained by that that with a

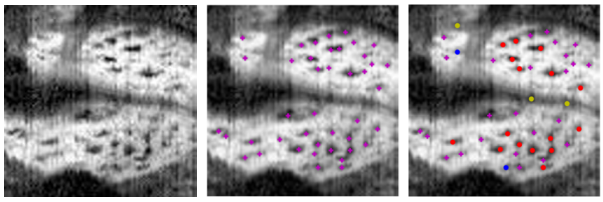


Fig. 9: Example of *a postriori* validation result by the expert for a more reliable evaluation. (a) pre-processed image. (b) our network prediction. (c) manual validation by the expert. Blue and yellow markers are the object-level FN and FP pores respectively. Red points in (c) are the partial ground truth for the evaluation in Table III, while the magenta points are TP pores that are not in the partial ground truth.

TABLE IV: *A posteriori* evaluation by the expert (10 images).

	Object-level			
Model	TP	FP	FN	Recall
Chan_Vese [22]	23.8	8.4	12.9	0.64
W-Net [23]	27.2	15.5	8.7	0.76
U-Net [9]	29.8	2.0	5.1	0.85
Attention U-Net [19]	27.4	**1.8**	5.6	0.83
Our method	**32.7**	2.2	**3.2**	**0.91**

smaller ratio of annotated points, the network tends to predict more pore candidates, even though some of them are simply small noisy areas, and we observe that pore features are not well learned. However, with 60% of the pathway annotation, sometimes the network is almost incapable of predicting pore candidates with very few pathways for training phase (about 5 for an OCT volume) (Fig. **??**f). The deleted pores, as well as the missing pores in the ground truth, are too much penalized to get a fair segmentation result with the small dataset.

Comparison with other state-of-the-art methods. We compare the proposed method with traditional active contour approach proposed by Chan_Vese [22], the unsupervised method W-Net [23], as well as the supervised methods of original U-Net [9] and Attention U-Net [19]. We manually delineated a peripheral mask to only retain the most contrasted area of the LC for Chan_Vese [22] and W-Net [23] approaches, since they are not capable of detecting the LC area automatically. Fig. 8 shows the visual comparison of segmentation performances, we can observe that active contour and unsupervised W-Net approaches are sensitive to acquisition noises and artifacts, while U-Net based methods are more efficient in predicting the pore candidates. The proposed method is able to predict more pore candidates in low contrast regions and in border regions, thanks to the context-aware design that is able to refer local consecutive potential pore areas.

The proposed method is robust in identifying true positive pores with a high $Recall$ value, as shown in Table III. The $Dice$, $Jaccard$ and $Precision$ are low for all models, which could be explained by two factors: on the one hand, only partial pores are annotated in the ground, so that false positive regions in the output segmentation map may correspond to

pores that are not labeled in the ground truth; on the other hand, the thresholding value $DIST_{th}$ is low, resulting in pore sizes in the ground truth tend to be smaller.

To this end, we inversely asked the expert to validate or not the pores suggested by different methods and also to point out the missing pores. Using this method, we bypass the partial ground truth problem. Obviously this method is time-consuming, hence why we did it only on 10 images of the test dataset. This is what we show in Fig. 9 and Table IV where the results produced by different methods are evaluated *a posteriori* instead of relying on the partial ground truth for score evaluation. We can observe that our model predicts more TP pores thanks to the context-aware design, and meanwhile the attention gate helps to eliminate FP pores by gradually attenuating the activation of their surrounding background as the network goes to shallower layers in the decoder path, resulting in a more accurate prediction of pore location and pore sizes. Finally, The missing pores (FN) are mainly located in the beginning or the end of the pathway, where the image is of low signal to noise ratio, as well as the artifact areas, where in both cases it is hard even for experts to identify without referring the continuity.

V. Conclusion

We proposed a simple yet effective method for LC pore segmentation. Our algorithm is based on a context-aware U-Net with added attention gates and has proved to be competitive with other state-of-the-art methods. This is a difficult task not only because the images are of low resolution and noisy, but because of varying advises of expert physicians to build the ground truth: disagreements exist from one expert to another on which spots are really pores or not. This is a clear limitation of our work that we will have to work on in the future, perhaps by proposing a dynamic system in which user can interact in real time with the neural network to confirm or not the presence of a perceived LC pore. Our future work also includes 3D modeling of pores to quantify deformations in glaucomatous eyes to better understand the disease.

References

[1] HA Quigley, E Addicks, W Green, and AE Maumenee, "Optic nerve damage in human glaucoma. ii. the site of injury and susceptibility to damage," *Archives of ophthalmology*, vol. 99, pp. 635–49, 05 1981.

[2] Albert Dichtl, Jost Jonas, and Gottfried Naumann, "Course of the optic nerve fibers through the lamina cibrosa in human eyes," *Graefe's archive for clinical and experimental ophthalmology = Albrecht von Graefes Archiv für klinische und experimentelle Ophthalmologie*, vol. 234, pp. 581–5, 10 1996.

[3] Abhiram Vilupuru, Nalini Rangaswamy, Laura Frishman, Earl Smith, Ronald Harwerth, and Austin Roorda, "Adaptive optics scanning laser ophthalmoscopy for in vivo imaging of lamina cribrosa," *Journal of the Optical Society of America. A, Optics, image science, and vision*, vol. 24, pp. 1417–25, 06 2007.

[4] Kazuko Omodaka, Takaaki Horii, Seri Takahashi, Tsutomu Kikawa, Akiko Matsumoto, Yukihiro Shiga, Kazuichi Maruyama, Tetsuya Yuasa, Masahiro Akiba, and Toru Nakazawa, "3d evaluation of the lamina cribrosa with swept-source optical coherence tomography in normal tension glaucoma," *PLOS ONE*, vol. 10, pp. e0122347, 04 2015.

[5] Nripun Sredar, Kevin Ivers, Hope Queener, George Zouridakis, and Jason Porter, "3d modeling to characterize lamina cribrosa surface and pore geometries using in vivo images from normal and glaucomatous eyes," *Biomedical optics express*, vol. 4, 2013.

[6] Mei Tan, Sim Ong, Sri Thakku, Ching-yu Cheng, Tin Aung, and Michael Girard, "Automatic feature extraction of optical coherence tomography for lamina cribrosa detection," *Journal of Image and Graphics*, vol. 3, 01 2015.

[7] Zach Nadler, Bo Wang, Gadi Wollstein, Jessica Nevins, Hiroshi Ishikawa, Larry Kagemann, Ian Sigal, Daniel Ferguson, Daniel Hammer, Ireneusz Grulkowski, Jonathan Liu, Martin Kraus, Chen Lu, Joachim Hornegger, James Fujimoto, and Joel Schuman, "Automated lamina cribrosa microstructural segmentation in optical coherence tomography scans of healthy and glaucomatous eyes," *Biomedical optics express*, vol. 4, pp. 2596–2608, 11 2013.

[8] Florence Rossant, Kate Grieve, Stéphanie Zwillinger, and Michel Paques, "Detection and tracking of the pores of the lamina cribrosa in three dimensional sd-oct data," 11 2017, pp. 651–663.

[9] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," 2015.

[10] Nahian Siddique, Sidike Paheding, Colin P. Elkin, and Vijay Devabhaktuni, "U-net and its variants for medical image segmentation: A review of theory and applications," *IEEE Access*, vol. 9, 2021.

[11] Sripad Krishna Devalla, Prajwal K R, Bharathwaj Sreedhar, Shamira Perera, Jean Martial Mari, Khai Chin, Tin Tun, Nicholas Strouthidis, Tin Aung, Alexandre Thiery, and Michael Girard, "Drunet: A dilated-residual u-net deep learning network to digitally stain optic nerve head tissues in optical coherence tomography images," *Biomedical Optics Express*, vol. 9, 03 2018.

[12] Zailiang Chen, Dabao Li, Hailan Shen, Hailan Mo, Ziyang Zeng, and Hao Wei, "Automated segmentation of fluid regions in optical coherence tomography b-scan images of age-related macular degeneration," *Optics & Laser Technology*, vol. 122, pp. 105830, 02 2020.

[13] Ruwan Tennakoon, Amirali K Gostar, Reza Hoseinnezhad, and Alireza Bab-Hadiashar, "Retinal fluid segmentation in oct images using adversarial loss based convolutional neural networks," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. IEEE, 2018.

[14] Hui Qu, Pengxiang Wu, Qiaoying Huang, Jingru Yi, Zhennan Yan, Kang Li, Gregory M Riedlinger, Subhajyoti De, Shaoting Zhang, and Dimitris N Metaxas, "Weakly supervised deep nuclei segmentation using partial points annotation in histopathology images," *IEEE transactions on medical imaging*, 2020.

[15] Hoel Kervadec, Jose Dolz, Shanshan Wang, Eric Granger, and Ismail Ben Ayed, "Bounding boxes for weakly supervised segmentation: Global constraints get close to full supervision," in *Medical Imaging with Deep Learning*. PMLR, 2020.

[16] Yigit B Can, Krishna Chaitanya, Basil Mustafa, Lisa M Koch, Ender Konukoglu, and Christian F Baumgartner, "Learning to segment medical images with scribble-supervision alone," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Springer, 2018.

[17] Bo Wang, Katie Lucy, Joel Schuman, Ian Sigal, Richard Bilonick, Chen Lu, Jonathan Liu, Ireneusz Grulkowski, Zach Nadler, Hiroshi Ishikawa, Larry Kagemann, James Fujimoto, and Gadi Wollstein, "Tortuous pore path through the glaucomatous lamina cribrosa," *Scientific Reports*, vol. 8, 05 2018.

[18] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan Gomez, Lukasz Kaiser, and Illia Polosukhin, "Attention is all you need," 06 2017.

[19] Ozan Oktay, Jo Schlemper, Loic Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Hammerla, Bernhard Kainz, Ben Glocker, and Daniel Rueckert, "Attention u-net: Learning where to look for the pancreas," 04 2018.

[20] Carole Sudre, Wenqi Li, Tom Vercauteren, Sébastien Ourselin, and Manuel Jorge Cardoso, "Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations," 07 2017.

[21] Patrice Simard, David Steinkraus, and John Platt, "Best practices for convolutional neural networks applied to visual document analysis.," pp. 958–962, 01 2003.

[22] Tony Chan and L.A. Vese, "Active contour without edges," *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, vol. 10, pp. 266–77, 02 2001.

[23] Clément Royer, Jeremie Sublime, Florence Rossant, and Michel Paques, "Imaging unsupervised approaches for the segmentation of dry armd lesions in eye fundus cslo images," *Journal of Imaging*, vol. 7, pp. 143, 08 2021.