2024 IEEE International Conference on Robotics and Automation (ICRA 2024)
May 13-17, 2024. Yokohama, Japan

Uncertainty-Aware Contextual Visualization for Human Supervision of OCT-Guided Autonomous Robotic Subretinal Injection

Michael Sommersperger[1], Shervin Dehghani[1], Philipp Matten[2], Hessam Roodaki[3] and Nassir Navab[1]

***Abstract*— The injection of therapeutic agents into the subretinal space might allow improved treatment of age-related macular degeneration. Various robotic systems have been developed to achieve the required precision and, in combination with intraoperative Optical Coherence Tomography (iOCT) imaging, methods for autonomous robotic guidance have been proposed. In such systems, the robot's cognition is often governed by machine learning algorithms, such as convolutional neural networks (CNNs), which provide semantic scene information from iOCT images. Although the robot performs a surgical task autonomously, human supervision is critical to monitor the robot's execution and, if necessary, stop the robot or take control to avoid trauma to the patient. In this paper, we propose a novel visualization concept for improved human supervision of autonomous robotic subretinal injection that integrates uncertainty information of the data provided to the robot. We design a focus and context visualization that renders an automatically identified instrument-aligned B-scan in the context of the 3D OCT volume. Our visualization is enriched by augmenting the uncertainty information on the instrument-aligned B-scan. To dynamically model task-specific uncertainty, we introduce a weighting scheme to assign an importance factor to each pair of classes, controlling the impact of their confusion on the overall uncertainty. We demonstrate our visualization concept on iOCT volumes acquired at different stages during subretinal injection on ex-vivo porcine eyes. We show that our processing pipeline achieves sufficient update rates for surgical display and discuss the impact of our visualization concept on the acceptance of robotic task autonomy for subretinal injection procedures.**

***Index Terms*— Medical Robots and Systems; Acceptability and Trust; Safety in HRI**

I. INTRODUCTION

Age-related macular degeneration (AMD), a leading etiological cause of visual impairment and blindness within developed nations [1], [2], predominantly manifests in the elderly population aged 60 and above. Its more aggressive manifestation, wet AMD, necessitates therapeutic intervention to mitigate vision impairment incurred through neovascularization beneath the macula, the pivotal region on the retinal surface governing central vision. A prevalent treatment modality, referred to as anti-angiogenesis therapy, involves the intravitreal administration of an anti-vascular endothelial growth factor (anti-VEGF) gene factor. Nonetheless, the therapeutic agent's delivery efficiency is constrained due to the limited permeability of the vitreous humor, a gel-like substance within the eye, rendering only a fraction of the administered drug accessible to the intended deeper target retinal layers, like the retinal pigment epithelium (RPE) or the photoreceptor layer. Consequently, recent years have witnessed the exploration of targeted retinal delivery as a potentially advantageous strategy, affording enhanced vector and drug administration efficacy. Furthermore, beyond AMD [3], diverse ocular conditions, including retinitis pigmentosa (RP) [4], and Leber's congenital amaurosis [5], have demonstrated amenability to therapeutic intervention via precise subretinal gene therapy delivery mechanisms.

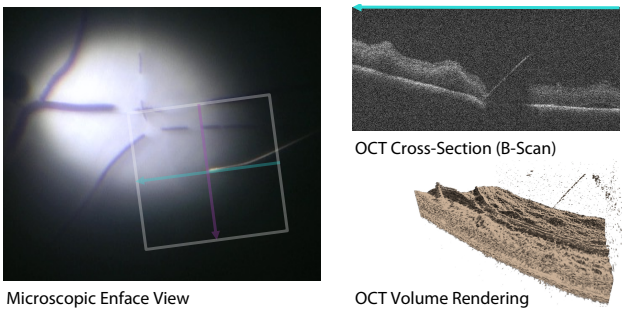

Fig. 1. View through an OCT integrated microscope during subretinal injection on an ex-vivo porcine eye (left). The white square outlines the OCT scanning area capturing a 3D volume, while the blue arrow shows the acquisition pattern generating the 2D OCT B-scan (right).

Accurate needle depth estimation and maintenance within the scant few hundred micrometers thickness of the human retina represent pivotal considerations in the realm of intraretinal injections. The intrinsic challenges posed by human hand tremor [6] and the inevitable onset of operator fatigue inherently constrain the proficiency of even the most experienced surgical practitioners to precisely and repeatably deliver the therapeutic agent to the on average 20-30 μm thin anatomical target area [7]. Consequently, these intrinsic limitations introduce challenges concerning irreversible patient trauma caused by tissue damage and an inefficient treatment due to the reflux of the therapeutic agent into the vitreous humor. To address these exigencies and enhance surgical precision, a substantive body of research has emerged, culminating in the development and efficacious deployment of robotic systems [8]–[11]. Presently, these systems function in concert with the human operator, mitigating the risk of retinal damage and drug reflux into the vitreous. Researchers have substantiated the effectiveness of such systems in ex vivo ocular models [12] and, notably, have extended their utility

Corresponding author: michael.sommersperger@tum.de
[1]M. Sommersperger, S. Dehghani, and N. Navab are with Department of Computer Science, Technische Universität München, 85748 München, Germany.
[2]P. Matten is with the Center for Medical Physics and Biomedical Engineering, Medical University of Vienna, 1090 Vienna, Austria.
[3]H. Roodaki is with Carl Zeiss Meditec AG, 73447 Oberkochen, Germany.

to clinical trials in human subjects [8], [13].

Intraoperative optical coherence tomography (iOCT), a non-invasive optical imaging modality frequently employed in the context of vitreoretinal surgery, has garnered attention as a valuable instrument for facilitating robotically guided intraocular injections, primarily owing to its inherent capacity to furnish precise depth information [16]. The acquisition of depth information through the utilization of intraoperative optical coherence tomography (iOCT) is crucial for navigating the 3D surgical scene and in particular for guiding the instrument precisely through the retinal tissue to the target location.

The majority of iOCT configurations predominantly employ OCT as a supplementary modality at specific critical junctures during surgical procedures. In these configurations, a conventional white light microscope serves as the primary imaging tool, affording a substantial field of view (FoV) to facilitate effective lateral orientation. Conversely, off-the-shelf iOCT systems typically offer limited visualization capabilities, primarily presenting only one or two cross-sectional B-scans. Notably, a specialized iteration of iOCT denoted as 4D microscope-integrated OCT has been devised to capture OCT volumetric datasets at near real-time video frame rates [17]. Rendering techniques have been proposed for the comprehensive visualization of volumetric OCT data, allowing the user to explore the data from arbitrary perspectives. However, a limitation of this approach is the absence of depth information, as some areas of the retinal tissue remain obstructed from view. Figure 1 illustrates the available multimodal data. One notable challenge encountered in the context of these diverse visualization modalities is the juxtaposition of visual outputs, often presented in parallel. This arrangement poses a cognitive challenge for surgeons, necessitating adaptability when shifting their focus between different modes of visual information.

Moreover, for a robotic system to effectively harness this wealth of information, it necessitates the concurrent achievement of real-time semantic localization of intricate anatomical structures within the retina's layers. It is common for such systems to employ an AI model to perceive the semantic localization within the input domain, which can be either OCT or conventional microscope images. Recent approaches, as proposed in [14], [15], [18], [19], leverage a CNN-based algorithm to interpret the OCT data stream in real-time and guide the robot to a target location for subretinal injection. In another study [20], using the images from a white light microscope, a more geometrical method is employed which is based on a CNN-based model. Consequently, regardless of the imaging modality, the reliability and safety of the robotic platforms are intrinsically coupled to the performance of the AI models. The AI models are prune to errors, and despite the significant progress in developing accurate models for medical image perception [15], [21], it remains crucial to identify areas where the model may fail or express uncertainty. Consequently, there emerges a compelling imperative for the development of a guided visualization framework, one characterized by intuitive interpretability, catering to both human operators and the potential integration tailored to human supervision of autonomous robotic systems.

In this work, we present such a novel visualization concept that integrates the uncertainty of a convolutional neural network (CNN)'s semantic map. This allows a human operator to monitor, verify or, if necessary, overtake control of the robotic system, potentially improving the integration and acceptance of robotic systems for autonomous subretinal injections.

The two main contributions of this work are: **(i)** A focus and context visualization concept that facilitates the in-situ visualization of an instrument-aligned volume cross-section within the OCT volume rendering. This virtual cross-section is automatically identified, centering the focus area at the tool tip. **(ii)** In addition, our proposed concept is enriched with an uncertainty map generated from the network's confidence maps. Since the semantic scene information of instrument-aligned B-scans has been employed to control a robotic system in previous works [14], [15], [19], [22], monitoring its uncertainty during autonomous procedures allows the operator to gain insights into the reliability of the robotic system's cognition. These uncertain areas are highlighted on the in-situ visualized B-scan using a perceptually linear color map to preserve the original B-scan intensities. Our method further integrates a customizable class importance weighting to control the impact of inter-class confusion on the generated uncertainty map. We demonstrate our visualization concept on OCT volumes acquired at various stages during needle insertion and subretinal injection performed on ex-vivo porcine eyes. Our results highlight the value of the generated uncertainty maps and the contextual visualization for monitoring robotic subretinal injection procedures.

II. RELATED WORK

Existing works on OCT visualization focus on direct volume rendering techniques to improve the spatial perception of the 3D OCT data, aimed at improved tool-tissue interactions and surgical decision making [23]–[27]. While these visualizations are mainly showcased for conventional manual surgery, in this work, we present efforts towards tailored visualizations for human supervision of autonomous robotic surgery, integrating uncertainty information that offers insights into the data provided to the robot.

The in-situ visualization of the instrument-aligned B-scan integrated into OCT volume rendering requires the development of contextual visualization techniques. Focus and context visualizations for medical volume rendering have been proposed in [28]. In [29], transparency modulations are leveraged to display virtual content within a 3D context, generating spatial perception of the internal structures. The in-situ augmentation of virtual objects within real objects has been demonstrated for augmented reality applications in [30]. Similar to such concepts, we augment an automatically selected instrument-aligned volume cross-section within the 3D OCT volume rendering context.

We integrate the uncertainty map generated from a selected volume cross-section within this augmentation. Previously

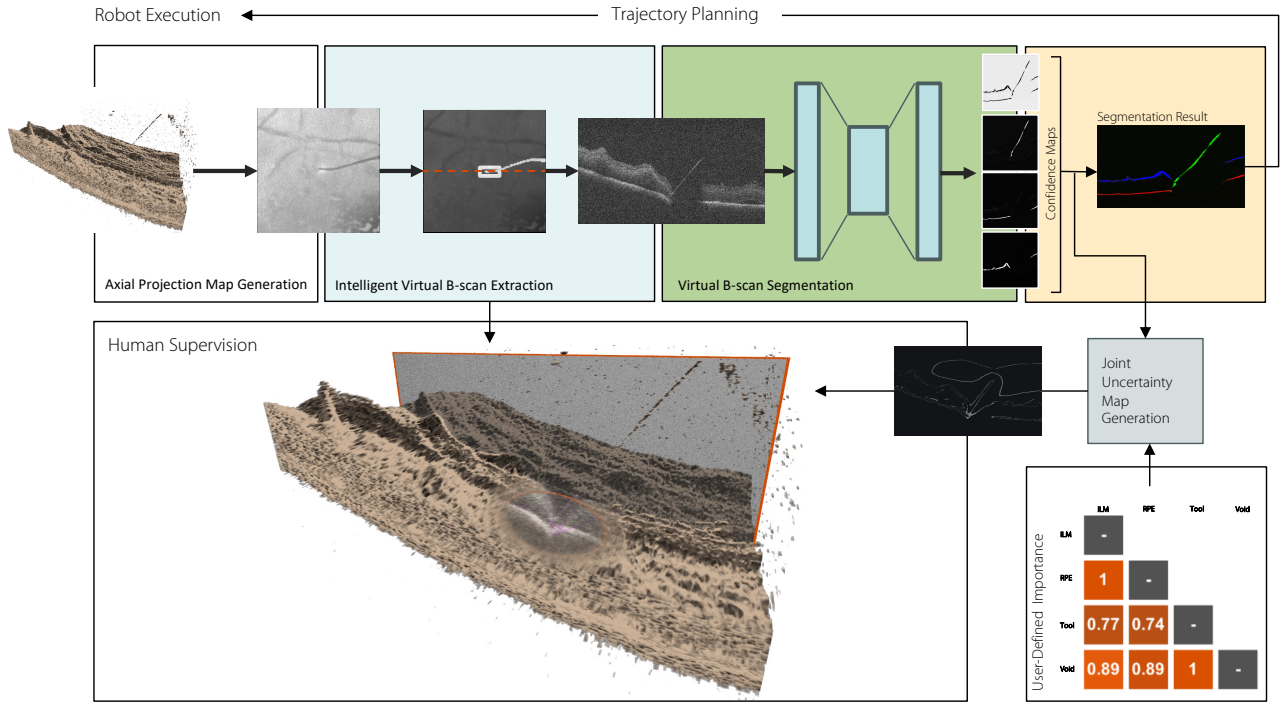

Fig. 2. Our proposed volume processing and visualization pipeline: we first determine the instrument location within the OCT volume to extract a tooltip-aligned virtual B-scan. Semantic scene information of such a B-scan typically obtained from a CNN is often employed to control the robot [14], [15]. We leverage the network's confidence maps as well as paired class importance weighting to generate an uncertainty map from the network's predictions. Finally, the virtual B-scan is augmented within a focus area to preserve the 3D context while revealing the instrument inserted into the retina.

investigated methods for uncertainty estimation have used dropout during network inference [31], [32], while others estimate the predictions' uncertainty via the entropy of the network's output [33]–[35]. Since methods leveraging the network's dropout require multiple model inferences leading to higher overall processing times, we directly integrate the confidence maps of a CNN output to calculate the entropy-based uncertainty. In addition, we introduce a paired class importance weighting that allows customization of the importance of the confusion between specific classes.

III. METHODS

The composition of the uncertainty-aware contextual visualization requires the extraction of a set of parameters from the OCT volume. These parameters include positional information, such as the plane parameters of the relevant virtual volume cross-section aligned with the instrument and the tooltip position within this cross-section, defining the position and content of the focus area. Besides this positional information, our visualization concept integrates an uncertainty map corresponding to the uncertainty of the network's semantic information at the tool-aligned virtual B-scan. To extract these components, we designed an efficient volume processing pipeline, outlined in detail in section III-A. Section III-B describes the extraction of the relevant tool-aligned virtual B-scan and the generation of the importance-based uncertainty map, while section III-C outlines the integration of the focus and context visualization into the OCT volume rendering using the generated components.

A. Processing Pipeline

To match the real-time requirements of surgical applications, our pipeline is designed for processing and visualization of OCT volumes at high update rates. Figure 2 shows an overview of the pipeline. First, the instrument is segmented in a 2D projection image to extract the instrument-aligned B-scan parameters [14]. As demonstrated in previous works [14], [15], [23], [24], [26], an efficient way of identifying the instrument is to generate a 2D projection image by averaging the volume intensities for each A-scan and employ a Unet-like network [36] with ResNet18 encoder [37] to provide a instrument segmentation map. To only consider the tooltip when extracting the instrument-aligned virtual B-scan, we extract a region of interest (ROI) around the tooltip [15]. We then fit a line l_{bscan} to the instrument mask within the ROI (orange dashed line in Fig. 2). This together with the OCT A-scan direction allows the extraction of the volume cross-section aligned with the instrument tip. As described in [14], given l_{bscan}, the normal of the plane can be defined as:

$$\vec{n} = \langle \sin\theta_z, \cos\theta_z, 0 \rangle$$

where θ_z is the rotation of the l_{bscan} on the enface plane. As described in [14], a virtual B-scan can be generated by sampling the OCT volume A-scans along this estimated line. We use nearest neighbor sampling to generate the volume cross-section along the line. Virtual B-scans extracted in a similar fashion were used in previous works [14], [15] to control or provide feedback to a robotic system for needle insertion

in subretinal injection procedures. For such algorithms, a segmentation network provides semantic scene information from the B-scans by generating a confidence map for each of the classes, each containing the probability of a specific pixel belonging to the corresponding class. As mentioned in previous works, we employ a Unet-style architecture with ResNet18 encoder to generate the confidence maps for the classes internal limiting membrane (ILM), retinal pigment epithelium (RPE), tool, and background (void). An *argmax* operation applied on the network's output channels provides the class prediction per pixel in the processed B-scan. Since the instrument is inserted into the retina from above, the position p_{tool} of the tool tip can be determined as the lowest position of the instrument in the segmentation along the A-scans. The virtual B-scan plane can be parametrized by a single point on its surface and $\vec{n}$. Therefore, p_{tool} and $\vec{n}$ uniquely define the virtual B-scan plane in the OCT volume.

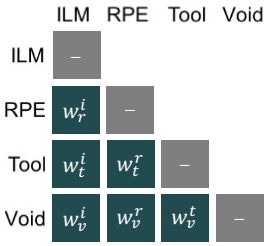

Fig. 3. Paired class importance weightings to control the impact of the classes on the overall uncertainty.

B. Uncertainty Map Generation

The output of the segmentation network consists of a confidence map for each of the classes. To determine the network uncertainty, we consider the entropy between each pair of classes. For any pair of classes α and β, the entropy at a pixel (i, j) can be calculated as:

$$ {}^{\beta}_{\alpha}H(i,j) = \sum_{k}^{\alpha,\beta} -\Pr(d(i,j)=k)\cdot\log_2(\Pr(d(i,j)=k)) \quad (1) $$

where $Pr(d(i,j)=k)$ describes the probability of the pixel at location (i,j) being classified as class k, referring to the confidence value. The paired entropy allows the possibility to assign a dynamic importance weighting for each pair of classes that controls the impact of their entropy on the overall uncertainty map. Figure 3 shows an example of the importance weightings $w_\alpha^\beta \in [0,1]$ for each pair of classes α and β. This customizable weighting is then integrated with the paired class entropy ${}^{\beta}_{\alpha}H(i,j)$ as follows:

$$ {}^{\beta}_{\alpha}H_{\mathcal{N}}(i,j) = {}^{\beta}_{\alpha}H(i,j)\cdot\Pr(d(i,j)\in\{\alpha,\beta\})\cdot w_\alpha^\beta \quad (2) $$

The factor $\Pr(d(i,j)\in\{\alpha,\beta\})$ additionally weights the entropy based on the sum of the confidence values of class α and β. Intuitively, the confusion described by the entropy between two classes is weighted less if the class probabilities are low. Once obtained ${}^{\beta}_{\alpha}H_{\mathcal{N}}(i,j)$ for each pair of classes, the overall uncertainty map is defined as the sum of the weighted entropies:

$$ U(i,j) = \sum_{\alpha,\beta} {}^{\beta}_{\alpha}H_{\mathcal{N}}(i,j) \quad (3) $$

We finally normalize the uncertainty map U and forward it to the rendering pipeline, along with the virtual B-scan plane parameters and the tooltip position.

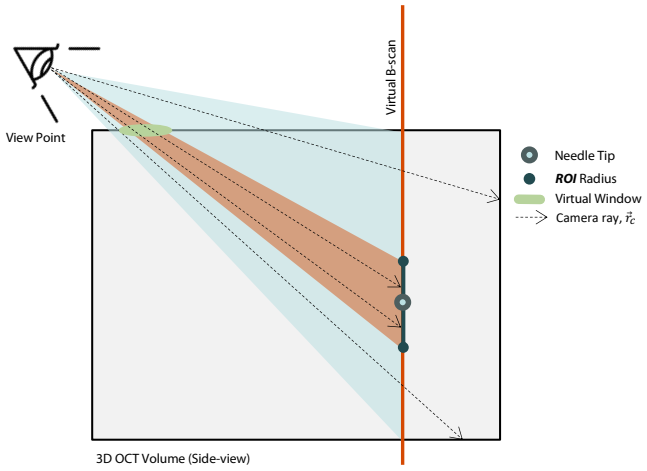

Fig. 4. Illustration of the contextual visualization. The OCT volume is outlined from a side view. Camera rays that reach the ROI around the tooltip through a virtual window (marked in orange) are contained in the focus area, while the remaining camera rays render the 3D volume context.

C. Focus and Context Visualization

A fundamental part of our concept is the focus and context visualization that facilitates the in-situ rendering of the instrument-aligned virtual B-scan in an OCT volume.

The ROI in a radius around the tooltip is considered the focus area for the in-situ visualization of the virtual B-scan. The point p_i, at which the camera ray r_c intersects with the virtual B-scan plane, is calculated as:

$$ p_i = p + \frac{(p_{tool}-p)\cdot\vec{n}}{\vec{n}\cdot\vec{r}_c}\cdot\vec{r}_c \quad (4) $$

During volume raymarching, we decide for each camera ray $\vec{r}_c$ and its intersection with the B-scan p_i, whether it passes the focus area defined by a radius r_{ROI} around the tooltip:

$$ d(p_i, p_{tool}) \leq r_{\text{ROI}}, \quad (5) $$

where d determines the Euclidean distance between two points.

If the ray does not fulfill this equation, we integrate iso-surface rendering with Phong shading to render the OCT volume and provide the 3D context around the ROI. On the other hand, if the condition is fulfilled, the ray integrates the virtual B-scan. A graphical illustration of this procedure is shown in Figure 4.

To visualize the focus area, the ray intersection with the retinal surface, determined by an iso-surface value, is shaded transparently. In particular, the opacity of the retinal surface

is defined by the distance between the projected ray point p_i and the tooltip:

$$f_t = max\left(\left(\frac{d(p_i, p_{tool})}{r_{\text{ROI}}}\right)^s, m\right) \tag{6}$$

where s regulates the smoothness of the transparency transition between the focus and contextual areas and m preserves a minimum opacity level. We choose $s = 3$ and $m = 0.05$ in our visualizations. This opacity function increases the transparency towards the tooltip while preserving the context and 3D perception of the rendering. After the camera ray has passed the virtual window at the surface, the area between the virtual window and the B-scan ROI is rendered transparent. To determine if the sampled position along the camera ray intersects with the virtual B-scan, we use the following equation:

$$\left|\frac{|n_x p_x + n_y p_y + n_z p_z + d|}{\sqrt{n_x^2 + n_y^2 + n_z^2}}\right| < d_{bscan} \tag{7}$$

If the sampled position along the camera ray intersects with the virtual B-scan, the sampled intensities are rendered in the familiar B-scan grayscale colorization, visualizing the virtual B-scan inside the volume. To augment the uncertainty $U(p)$ on the B-scan, we employ the perceptually linear L*a*b* color space due to its properties of preserving the original B-scan intensity values, which is crucial for correct interpretation of the presented data:

$$C_{L^*a^*b^*}(I(p), \delta^*) = (1.0 - \delta^*) \cdot C \tag{8}$$

where $I(p)$ is the volume intensity, $\delta^* = clamp_{[0,1]}(\frac{U(p)}{4})$ and $C = (I(p), U(p), -U(p))$. Finally, we define $\gamma = 1.0 - (U(p) - 1)^2$ and generate the final uncertainty augmentation:

$$C(p) = [RGB(\gamma \cdot (1 - \delta^*) \cdot C_{L^*a^*b^*}(I(p), \delta^*))] \tag{9}$$

Finally, we outline the virtual B-scan in the volume rendering as well as the region of interest to improve the spatial awareness of the slice positioning within the OCT volume, yielding the final uncertainty-aware contextual visualization.

IV. EXPERIMENTS

A. Experimental Setup

To train our segmentation network and demonstrate visual examples evaluating our proposed concept, we collected OCT data from experiments on ex-vivo porcine eyes. The OCT scans were collected from a Rescan 700 (Carl Zeiss Meditec, Jena) iOCT system integrated into an operating microscope. A subretinal injection cannula (INCYTO Needle-RNT) with a 23G body and 41G tip was inserted into the retina and subretinal injection was simulated using a saline solution. OCT volumes were acquired at different stages during the procedure. Each volume consists of a total of 128 B-scans with a resolution of 512×1024 pixels, covering an area of 5×5 mm at an imaging depth of 2.8 mm. To train our Unet-style CNN for the segmentation of the instrument-aligned B-scans, we acquired a dataset of 313 B-scans. The data was labeled by two biomedical engineers. We employed a combination of focal and dice loss for network training and vertical flipping for data augmentation. For instrument segmentation in the 2D projection images, we use the model and dataset of [26]. Furthermore, we acquired OCT volumes at different stages before and during needle insertion, as well as after subretinal injection to provide the visual results illustrated in Section IV-C. For the development of our system, we used Pytorch 1.13 for network training and inference and Unity 2021.3 for processing and rendering the OCT volumes.

TABLE I

AVERAGE INFERENCE TIME IN MILLISECONDS OF OUR PROCESSING AND VISUALIZATION PIPELINE.

Tool Map	B-scan	Segment	Render	Total
4.7 ± 1.3	8.0 ± 1.3	5.5 ± 0.2	17.6 ± 0.5	$\mathbf{35.7 \pm 2.3}$

B. Time profiling

To assess the potential of deploying our method for real-time visualization of autonomous robotic subretinal injections, we benchmarked the inference times of the individual pipeline components on a Windows 10 workstation with an Intel Core i7-8700K @3.7 GHz processor and Nvidia RTX 3090Ti GPU. The inference times of the individual pipeline stages are provided in Table I. In this table, *Tool Map* refers to the generation of the 2D projection image and the instrument segmentation mask, while *B-scan* refers to the estimation and generation of the instrument-aligned volume cross-section. The remaining columns refer to the segmentation of the virtual B-scan and the rendering procedure to provide the final visualization result. The uncertainty map can be generated on the GPU in less than 0.5 ms. On average, overall inference times of around 36 ms were measured, showing the potential of deploying our method in tandem with recent 4D OCT systems [17], capable of providing volumetric update rates of $10Hz$.

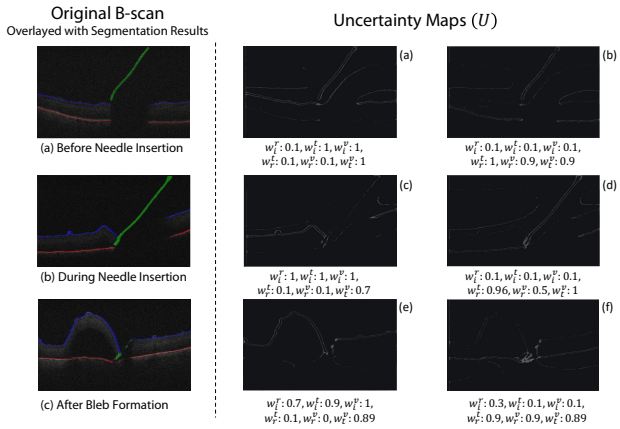

Fig. 5. Demonstrations of the network uncertainty maps of B-scans in various conditions. Customizing the importance weighting allows for highlighting the uncertainty of specific classes. The green, blue and red colors in the B-scan mark the predicted tool, ILM and RPE class labels.

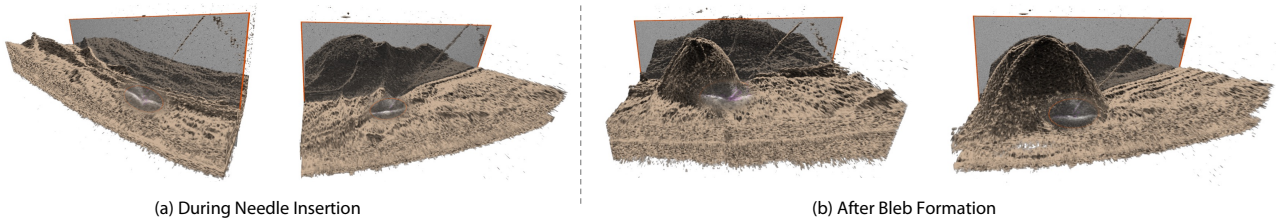

Fig. 6. Demonstrations of our rendering concept on OCT volumes acquired during needle insertion (a) and after subretinal bleb formation (b). The purple colored areas on the in-situ visualized B-scan indicate higher uncertainty, based on the generated uncertainty maps.

C. Visualization Examples

We demonstrate our proposed visualization concept on OCT volumes acquired at different stages during needle insertion and subretinal injection on ex-vivo porcine eyes. We provide a qualitative evaluation of our concept in two stages. First, we demonstrate the effectiveness of our importance-based uncertainty map to highlight the network's confusion between specific classes of interest. Figure 5 shows three representative examples of instrument-aligned B-scans at different stages during subretinal injection. The semantic information provided by the segmentation network is overlaid on the B-scans. We show that different importance values for the paired class confusions, referring to the factors in Figure 3, highlight different relevant regions in the generated uncertainty maps. While the column in the middle visualizes a higher importance weighting of the uncertainty involving the ILM and the tool classes, the column on the right emphasizes the importance of uncertainty involving the RPE and tool. The resulting maps show that in particular the uncertainties of important classes are effectively highlighted. Note that such information is not contained in the network's segmentation results.

For the final rendering of our contextual visualization, we customize the importance-weighting for the uncertainty map generation with a high emphasis on any confusion between the tool, RPE and background classes, using the parameters provided in Fig. 5(f). Figure 6 shows the results of our pipeline, demonstrated on two volumes acquired with the surgical instrument inserted into the retina and two volumes acquired after performing the injection. In Figure 6(a) the volumes are visualized at different angles demonstrating the viewpoint-agnostic visibility of the ROI around the tooltip. The examples in Figure 6(b) show the focus region within the 3D context of the subretinal bleb. In all examples, the focus area around the tooltip as well as uncertain areas are clearly visualized.

V. DISCUSSION

The visual results presented in the previous section show that our generated uncertainty maps highlight vital areas for successful subretinal injection. In particular, Figure 5 (d) and (f) show high uncertainty around the tooltip when the instrument is close to the RPE layer. In an eventual autonomous robotic procedure, this uncertainty information would provide valuable insights regarding the reliability of the data provided to the robot. If a human supervisor could monitor such uncertain information and stop or take control of the robot, potential damage to the RPE and other retinal cells caused by the needle tip could be avoided. The in-situ visualization of the tool-aligned B-scan within the 3D OCT volume shows progress towards relevance-based visualizations, compared to the side-by-side visualization in currently available systems described in Fig. 1. This allows verification of the tool positioning and monitoring of the subretinal bleb generation within one rendering and aims to minimize the mental mapping between the 2D B-scan and 3D volume visualization, reducing the cognitive load on the surgeon. In this work, presented the design and first technical implementation of the proposed visualization concept and showed its potential by demonstrating representative examples. Rather than developing a perfect segmentation model, we focus on bringing awareness to areas, in which the information provided to the robot might be uncertain. In the next step, we plan to evaluate a potential relief of the cognitive load posed on users by integrating our visualisation into a 4D OCT system with live volume acquisition and a robotic platform to autonomously perform needle insertion. Furthermore, the importance-weighting for the uncertainty map generation could, in the future, be automatically adjusted via phase detection algorithms identifying the stage of the procedure to dynamically visualize the uncertainty in crucial areas.

VI. CONCLUSIONS

We proposed a novel uncertainty-aware visualization concept that facilitates the in-situ visualization of an automatically identified instrument-aligned B-scan within the context of an OCT volume rendering. Our experiments demonstrate the real-time performance of our processing pipeline and show that our importance-based uncertainty map generation offers valuable insights into the data that is provided to robotic systems. We anticipate that advancing uncertainty visualization techniques will enhance the acceptance of robotic task autonomy, accelerate the integration of robot systems, and ultimately reduce the risks for patient trauma through improved human supervision.

VII. ACKNOWLEDGMENT

This work is partially supported, and the data is provided by Carl Zeiss Meditec.

REFERENCES

[1] R. Kawasaki, M. Yasuda, S. J. Song, S.-J. Chen, J. B. Jonas, J. J. Wang, P. Mitchell, and T. Y. Wong, "The prevalence of age-related macular degeneration in asians: A systematic review and meta-analysis," *Ophthalmology*, vol. 117, no. 5, pp. 921–927, 2010. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0161642009011786

[2] W. L. Wong, X. Su, X. Li, C. M. G. Cheung, R. Klein, C.-Y. Cheng, and T. Y. Wong, "Global prevalence of age-related macular degeneration and disease burden projection for 2020 and 2040: a systematic review and meta-analysis," *The Lancet. Global health*, vol. 2, no. 2, pp. e106–16, 2014.

[3] Y.-J. Jo, K.-H. Sonoda, Y. Oshima, A. Takeda, R.-i. Kohno, J. Yamada, J. Hamuro, Y. Yang, S. Notomi, T. Hisatomi, and T. Ishibashi, "Establishment of a New Animal Model of Focal Subretinal Fibrosis That Resembles Disciform Lesion in Advanced Age-Related Macular Degeneration," *Investigative Ophthalmology & Visual Science*, vol. 52, no. 9, pp. 6089–6095, 08 2011. [Online]. Available: https://doi.org/10.1167/iovs.10-5189

[4] N. G. Ghazi, E. B. Abboud, S. R. Nowilaty, H. Alkuraya, A. Alhommadi, H. Cai, R. Hou, W.-T. Deng, S. L. Boye, A. Almaghamsi, F. Al Saikhan, H. Al-Dhibi, D. Birch, C. Chung, D. Colak, M. M. LaVail, D. Vollrath, K. Erger, W. Wang, T. Conlon, K. Zhang, W. Hauswirth, and F. S. Alkuraya, "Treatment of retinitis pigmentosa due to mertk mutations by ocular subretinal injection of adeno-associated virus gene vector: results of a phase i trial," *Human Genetics*, vol. 135, no. 3, pp. 327–343, Mar 2016. [Online]. Available: https://doi.org/10.1007/s00439-016-1637-y

[5] F. Testa, A. M. Maguire, S. Rossi, E. A. Pierce, P. Melillo, K. Marshall, S. Banfi, E. M. Surace, J. Sun, C. Acerra, J. F. Wright, J. Wellman, K. A. High, A. Auricchio, J. Bennett, and F. Simonelli, "Three-year follow-up after unilateral subretinal delivery of adeno-associated virus in patients with leber congenital amaurosis type 2," *Ophthalmology*, vol. 120, no. 6, pp. 1283–1291, 2013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0161642012011682

[6] C. N. Riviere and P. S. Jensen, "A study of instrument motion in retinal microsurgery," in *Proceedings of the 22nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (Cat. No. 00CH37143)*, vol. 1. IEEE, 2000, pp. 59–60.

[7] M. Karampelas, D. A. Sim, P. A. Keane, V. P. Papastefanou, S. R. Sadda, A. Tufail, and J. Dowler, "Evaluation of retinal pigment epithelium–bruch's membrane complex thickness in dry age-related macular degeneration using optical coherence tomography," *British Journal of Ophthalmology*, vol. 97, no. 10, pp. 1256–1261, 2013.

[8] R. Ladha, T. Meenink, J. Smit, and M. D. de Smet, "Advantages of robotic assistance over a manual approach in simulated subretinal injections and its relevance for gene therapy," *Gene Therapy*, vol. 30, no. 3, pp. 264–270, Apr 2023. [Online]. Available: https://doi.org/10.1038/s41434-021-00262-w

[9] E. Vander Poorten, C. N. Riviere, J. J. Abbott, C. Bergeles, M. A. Nasseri, J. U. Kang, R. Sznitman, K. Faridpooya, and I. Iordachita, "Robotic retinal surgery," in *Handbook of Robotic and Image-Guided Surgery*. Elsevier, 2020, pp. 627–672.

[10] M. Zhou, Q. Yu, K. Huang, S. Mahov, A. Eslami, M. Maier, C. P. Lohmann, N. Navab, D. Zapp, A. Knoll *et al.*, "Towards robotic-assisted subretinal injection: A hybrid parallel–serial robot system design and preliminary evaluation," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 8, pp. 6617–6628, 2019.

[11] M. A. Nasseri, M. Eder, S. Nair, E. C. Dean, M. Maier, D. Zapp, C. P. Lohmann, and A. Knoll, "The introduction of a new robot for assistance in ophthalmic surgery," in *Eng. Med. Biol. Soc. (EMBC), 2013 35th Annu. Int. Conf. IEEE*. IEEE, 2013, pp. 5682–5685.

[12] K. Yang, X. Jin, Z. Wang, Y. Fang, Z. Li, Z. Yang, J. Cong, Y. Yang, Y. Huang, and L. Wang, "Robot-assisted subretinal injection system: development and preliminary verification," *BMC Ophthalmology*, vol. 22, no. 1, p. 484, Dec 2022. [Online]. Available: https://doi.org/10.1186/s12886-022-02720-4

[13] J. Cehajic-Kapetanovic, K. Xue, T. L. Edwards, T. C. Meenink, M. J. Beelen, G. J. Naus, M. D. de Smet, and R. E. MacLaren, "First-in-human robot-assisted subretinal drug delivery under local anesthesia," *American Journal of Ophthalmology*, vol. 237, pp. 104–113, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0002939421005924

[14] S. Dehghani, M. Sommersperger, P. Zhang, A. Martin-Gomez, B. Busam, P. Gehlbach, N. Navab, M. A. Nasseri, and I. Iordachita, "Robotic navigation autonomy for subretinal injection via intelligent real-time virtual ioct volume slicing," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, 2023, pp. 4724–4731.

[15] M. Sommersperger, J. Weiss, M. A. Nasseri, P. Gehlbach, I. Iordachita, and N. Navab, "Real-time tool to layer distance estimation for robotic subretinal injection using intraoperative 4d oct," *Biomedical Optics Express*, vol. 12, no. 2, pp. 1085–1104, 2021.

[16] M. J. Gerber, M. Pettenkofer, and J.-P. Hubschman, "Advanced robotic surgical systems in ophthalmology," *Eye*, vol. 34, no. 9, pp. 1554–1562, Sep 2020. [Online]. Available: https://doi.org/10.1038/s41433-020-0837-9

[17] A. Britten, P. Matten, J. Weiss, M. Niederleithner, H. Roodaki, B. Sorg, N. Hecker-Denschlag, W. Drexler, R. A. Leitgeb, and T. Schmoll, "Surgical microscope integrated mhz ss-oct with live volumetric visualization," *Biomed. Opt. Express*, vol. 14, no. 2, pp. 846–865, Feb 2023. [Online]. Available: https://opg.optica.org/boe/abstract.cfm?URI=boe-14-2-846

[18] J. W. Kim, S. Wei, P. Zhang, P. Gehlbach, J. U. Kang, I. Iordachita, and M. Kobilarov, "Deep learning guided autonomous retinal surgery using a robotic arm, microscopy, and ioct imaging," *arXiv preprint arXiv:2306.10127*, 2023.

[19] K. Mach, S. Wei, J. W. Kim, A. Martin-Gomez, P. Zhang, J. U. Kang, M. A. Nasseri, P. Gehlbach, N. Navab, and I. Iordachita, "Oct-guided robotic subretinal needle injections: A deep learning-based registration approach," in *2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2022, pp. 781–786.

[20] P. Zhang, J. W. Kim, and M. Kobilarov, "Towards safer retinal surgery through chance constraint optimization and real-time geometry estimation," in *2021 60th IEEE Conference on Decision and Control (CDC)*. IEEE, 2021, pp. 5175–5180.

[21] A. Farshad, Y. Yeganeh, P. Gehlbach, and N. Navab, "Y-net: A spatiospectral dual-encoder network for medical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2022, pp. 582–592.

[22] M. Zhou, X. Guo, M. Grimm, E. Lochner, Z. Jiang, A. Eslami, J. Ye, N. Navab, A. Knoll, and M. A. Nasseri, "Needle detection and localisation for robot-assisted subretinal injection using deep learning," *CAAI Transactions on Intelligence Technology*, 2023.

[23] M. Sommersperger, S. Dehghani, P. Matten, K. Mach, M. A. Nasseri, H. Roodaki, U. Eck, and N. Navab, "Semantic virtual shadows (svs) for improved perception in 4d oct guided surgery," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2023, pp. 408–417.

[24] M. Sommersperger, S. Dehghani, P. Matten, K. Mach, H. Roodaki, U. Eck, and N. Navab, "Intelligent virtual b-scan mirror (ivbm)," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2023, pp. 418–428.

[25] R. M. Trout, C. Viehland, J. Li, W. Raynor, A.-H. Dhalla, A. N. Kuo, L. Vajzovic, C. A. Toth, and J. Izatt, "Real-time feature-guided image fusion of posterior-segment intrasurgical optical coherence tomography and digital surgical microscopy," *Investigative Ophthalmology & Visual Science*, vol. 64, no. 8, pp. 2494–2494, 2023.

[26] J. Weiss, M. Sommersperger, A. Nasseri, A. Eslami, U. Eck, and N. Navab, "Processing-aware real-time rendering for optimized tissue visualization in intraoperative 4d oct," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part V 23*. Springer, 2020, pp. 267–276.

[27] C. Viehland, B. Keller, O. M. Carrasco-Zevallos, D. Nankivil, L. Shen, S. Mangalesh, A. N. Kuo, C. A. Toth, J. A. Izatt *et al.*, "Enhanced volumetric visualization for real time 4d intraoperative ophthalmic swept-source oct," *Biomedical optics express*, vol. 7, no. 5, pp. 1815–1829, 2016.

[28] S. Bruckner, S. Grimm, A. Kanitsar, and M. E. Groller, "Illustrative context-preserving exploration of volume data," *IEEE Transactions on Visualization and Computer Graphics*, vol. 12, no. 6, pp. 1559–1569, 2006.

[29] J. Kruger, J. Schneider, and R. Westermann, "Clearview: An interactive context preserving hotspot visualization technique," *IEEE Transactions on Visualization and Computer Graphics*, vol. 12, no. 5, pp. 941–948, 2006.

[30] C. Bichlmeier, F. Wimmer, S. M. Heining, and N. Navab, "Contextual anatomic mimesis hybrid in-situ visualization method for improving multi-sensory depth perception in medical augmented reality," in *2007*

6th IEEE and ACM International Symposium on Mixed and Augmented Reality, 2007, pp. 129–138.

[31] J. I. Orlando, P. Seeböck, H. Bogunović, S. Klimscha, C. Grechenig, S. Waldstein, B. S. Gerendas, and U. Schmidt-Erfurth, "U2-net: A bayesian u-net model with epistemic uncertainty feedback for photoreceptor layer segmentation in pathological oct scans," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. IEEE, 2019, pp. 1441–1445.

[32] Y. Gal and Z. Ghahramani, "Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning," Oct. 2016, arXiv:1506.02142 [cs, stat]. [Online]. Available: http://arxiv.org/abs/1506.02142

[33] A. Jungo and M. Reyes, "Assessing reliability and challenges of uncertainty estimations for medical image segmentation," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part II 22*. Springer, 2019, pp. 48–56.

[34] Z. Eaton-Rosen, F. Bragman, S. Bisdas, S. Ourselin, and M. J. Cardoso, "Towards safe deep learning: accurately quantifying biomarker uncertainty in neural network predictions," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part I*. Springer, 2018, pp. 691–699.

[35] T. DeVries and G. W. Taylor, "Leveraging uncertainty estimates for predicting segmentation quality," *arXiv preprint arXiv:1807.00502*, 2018.

[36] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*. Springer, 2015, pp. 234–241.

[37] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.