Real-time tool to layer distance estimation for robotic subretinal injection using intraoperative 4D OCT

MICHAEL SOMMERSPERGER,1,2,* JAKOB WEISS,2 M. ALI NASSERI,2,3 PETER GEHLBACH,1 IULIAN IORDACHITA,1 AND NASSIR NAVAB1,2

1Johns Hopkins University, Baltimore, MD 21218, USA
2Technical University of Munich, Germany
3Klinikum Rechts der Isar, Augenklinik, Munich, Germany
*michael.sommersperger@tum.de

Abstract: The emergence of robotics could enable ophthalmic microsurgical procedures that were previously not feasible due to the precision limits of manual delivery, for example, targeted subretinal injection. Determining the distance between the needle tip, the internal limiting membrane (ILM), and the retinal pigment epithelium (RPE) both precisely and reproducibly is required for safe and successful robotic retinal interventions. Recent advances in intraoperative optical coherence tomography (iOCT) have opened the path for 4D image-guided surgery by providing near video-rate imaging with micron-level resolution to visualize retinal structures, surgical instruments, and tool-tissue interactions. In this work, we present a novel pipeline to precisely estimate the distance between the injection needle and the surface boundaries of two retinal layers, the ILM and the RPE, from iOCT volumes. To achieve high computational efficiency, we reduce the analysis to the relevant area around the needle tip. We employ a convolutional neural network (CNN) to segment the tool surface, as well as the retinal layer boundaries from selected iOCT B-scans within this tip area. This results in the generation and processing of 3D surface point clouds for the tool, ILM and RPE from the B-scan segmentation maps, which in turn allows the estimation of the minimum distance between the resulting tool and layer point clouds. The proposed method is evaluated on iOCT volumes from ex-vivo porcine eyes and achieves an average error of 9.24 µm and 8.61 µm measuring the distance from the needle tip to the ILM and the RPE, respectively. The results demonstrate that this approach is robust to the high levels of noise present in iOCT B-scans and is suitable for the interventional use case by providing distance feedback at an average update rate of 15.66 Hz.

© 2021 Optical Society of America under the terms of the OSA Open Access Publishing Agreement

1. Introduction

Age-related macular degeneration (AMD) is the leading cause of blindness in patients over the age of 65 in developed countries [1]. Due to demographic changes and aging, the number of cases is increasing worldwide and is predicted to reach 288 million by 2040 [2]. The anatomic macula is an approximately 5 mm area of the retina containing the 1.5 mm highly specialized area known as the clinical macula or fovea, which is responsible for sharp vision. In advanced "wet" AMD, blood vessels grow through the barrier Bruch membrane and leak fluid or blood into and under the macula, which if left untreated, leads to irreversible damage to the photoreceptors and vision loss. Currently, AMD is not curable, but its progression in the advanced wet form can be slowed by intravitreous injection of anti-vascular endothelial growth factor drugs, which are globally accepted as the present standard of care [3,4]. Evolving therapeutic interventions include but are not limited to, subretinal stem cell therapy [5], gene therapy [6,7], photoreceptor [8] and RPE [9,10] cell transplants, and most recently gene editing [11] technology. Many of these emerging
approaches require or would benefit from access to the subretinal space. Enhanced safety, robust repeatability, higher precision, with fewer demands on the surgeon are all desirable elements of next treatment modalities. In order to achieve targeted delivery of therapeutic agent into the subretinal space, typically a microsurgical injection needle is directed through the internal limiting membrane (ILM), traverses the retina and delivers its payload into the potential subretinal space between the photoreceptors and the retinal pigment epithelium (RPE). High precision and control minimizes injurious contact with delicate photoreceptors and retinal pigment epithelial cells, assures proper localization of drug, and improves consistency and repeatability.

Although ophthalmic surgeons are trained to perform very delicate procedures with sub-millimeter precision, their hand tremor is estimated to be as high as 182 $\mu$m RMS [12] in amplitude. As the average thickness of the retina is 250 $\mu$m [13], and the injection target is an anatomical area around 20-30 $\mu$m [14], constraining the acceptable error, assistance with injection precision has to be pursued. To enable the required precision and open the possibility for targeted injection, a number of robotic concepts [15–20] have been introduced in the last decades. In 2016, surgeons at Oxford’s John Radcliffe Hospital performed the first such robotic eye surgery worldwide [21] and showed the feasibility of robot assisted ophthalmic interventions. However, crucial tasks for subretinal injection, such as the verifiable positioning of the needle tip at the correct insertion depth, are still challenging due to a number of factors including the very thin and flexible needle body [22].

Intraoperative optical coherence tomography (iOCT) has been used in various studies to provide visual guidance during ophthalmic interventions [6,23–25]. To date iOCT is the only imaging technology capable of detecting small retinal structures at micrometer resolution while providing live-feedback during surgery. The cross-sectional B-scan images can be used to determine the position of the tool relative to the retina and also to estimate the needle insertion status during subretinal injection. While in other vitreoretinal procedures the microscopic field of view provides the surgeon with all essential information and the OCT is used to integrate additional high-resolution information, in subretinal injection the cross-sectional view offers information that is not apparent from the microscopic view, such as imaging of the anatomy and the insertion target located below the retina surface, as well as the current insertion status, which emphasizes the importance of OCT for targeted and reproducible injections. Developments in swept source [26] and spiral scanning [27] OCT have enabled volumetric imaging at near video rate and therefore have opened new possibilities for 4D OCT guided ophthalmic surgeries. While prior work has demonstrated the feasibility of real-time visualization of 4D imaging data [26–29], real-time processing for advanced surgical guidance poses a significant challenge due to the high data rates of up to several GB/s. Additional challenges of iOCT include shadowing artefacts of the surgical instruments, which occur due to high reflectance at the tool surface that obscure relevant parts of the underlying retina, as well as high noise levels compared to diagnostic OCT, which add difficulties to tasks such as retinal layer and instrument segmentation. In the past, only limited work has been conducted that achieved automatic real-time feedback calculating the distance between the surgical tool and the retina, for example by using single real-time generated 2D B-scans [30]. A major challenge of determining the distance from only a single cross-sectional image is to precisely position the scan area to capture the tooltip. Manual adjustment after each robot movement is laborious, very time consuming and thus, adding, rather than diminishing, complexity to the procedure. Although automatic positioning of a single B-scan could be achieved via instrument segmentation of the microscopic image [31–33], consistently capturing the extremely small tip of a 41 gauge subretinal injection needle with a single B-scan image is still challenging and its failure leads to very high errors in the distance calculation. On the other hand, by acquisition of 3D volumes the scan area only has to be roughly aligned to the tool and does not need to be adjusted after each instrument movement. Furthermore, it can
easily be ensured that the OCT volume contains the instrument tip and shadowing artefacts can be restored from the surrounding retina anatomy.

Overall, rapid and continuous distance estimation between the needle tip and the retinal surface is important for the control and targeting of the robot. Once the tool is touching the retina, the robot is mainly advanced along one degree of freedom in the insertion direction to minimize tissue damage and patient trauma. The distance between the needle tip and the inner RPE surface determines the maximum advancement of the robot, in order to reach the subretinal space without damaging important retinal and retinal support cells.

In this work, we introduce an efficient method to robustly estimate the distance between the tip of the injection needle and the ILM and RPE surface boundaries from iOCT volumes. We propose a novel efficient pipeline for processing iOCT volume rates at speeds that are suitable for image-guided surgeries by reaching update rates compatible with OCT systems that have shown to enable 4D OCT-guidance retinal surgery visualizing surgical instruments and manipulation of tissue [27]. This is supported by a learning-based segmentation of tool, ILM and RPE surface boundaries from iOCT B-scans. Our pipeline achieves efficiency by narrowing down the region of interest (ROI) in a 2D projection. The B-scans within the ROI are segmented using a convolutional neural network (CNN) and subsequent extraction and processing of surface point clouds yield an estimation of the distance between instrument tip and the two retinal layers of interest. Herein, we evaluate the iOCT B-scan segmentation performance for our use case and compare it to a baseline model for retinal layer segmentation of diagnostic OCT B-scans. Additionally, we analyse various segmentation networks, addressing the time requirement imposed on our method and the interplay between network size and performance. Finally, the estimated distances are validated on 17 iOCT volumes acquired from ex-vivo porcine eyes.

To our knowledge, there is no direct related work on end-to-end needle tip to layer distance estimation using iOCT volumes. However, several existing works are closely related to the sub-components of the proposed pipeline. Thus, we discuss the state of art in regards to these three relevant applications: the localization and tracking of the needle tip in iOCT volumes, the distance estimation between a surgical tool and the retina surface, and the segmentation of retinal layers in diagnostic OCT B-scans.

To locate the position of the tooltip, Zhou et al. [34] track the geometry of a specifically shaped needle body above the retina in OCT volumes and fit the CAD model of the instrument to estimate the position of the tooltip, assuming that the instrument does not bend when inserted into the retina. Recently, deep learning approaches [35,36] have been introduced to detect the needle in B-scans. In [36], the authors propose a two-step approach, in which they first identify the tool above the retina and then estimate potential tool locations under the retina surface in the remaining B-scans. The one-stage detector RetinaNet is used to localize the needle in the candidate areas, reaching a detection accuracy of 99.2% and an error of 23 µm on their test data.

Addressing the distance estimation between a surgical tool and the retina surface, Roodaki et al. [30] propose to use only a single 2D OCT B-scan to provide the surgeon with distance information during tasks where the tool is exclusively located above the retina. Instead of OCT volumes, a real-time pattern of two perpendicular B-scans is acquired during surgery. The segmentation of the retina and tool surface is then generated by traditional filtering and thresholding methods. The tool and retina surfaces are distinguished by detecting the instrument shadow in the B-scans. Other works [37–39] integrate an OCT probe into surgical instruments to obtain distance feedback to the retinal surface from single A-scans. In particular, Cheon et al. [40] use an instrument-integrated OCT probe to calculate the insertion depth of the needle into the retina. However, the tool has to be aligned perpendicular to the target surface, and integration of an OCT probe makes the instruments more complicated.

To determine the 3D position of the needle tip and the retinal layers, we segment the surface boundaries of the ILM and the RPE layer as well as the tool surface in the iOCT B-scans.
Previous works on retinal layer segmentation are reported exclusively in the context of diagnostic OCT imaging. Traditional approaches can be categorized in A-scan [41–43] and B-scan [44–46] methods based on noise reduction and pre-processing techniques. Subsequent SVM and graph approaches reach errors around 6 µm, but come with computation times of several seconds or minutes [47]. Due to the intraoperative application, such segmentation approaches are computationally too expensive and cannot be used as part of our pipeline. In recent years, deep learning methods for retinal layer segmentation gained popularity. In 2017, Roy et al. introduced ReLayNet [48], a convolutional network for end-to-end segmentation of 10 retinal layers and fluid masses. They achieved fast computation times of 10 ms with a comparably small input image resolution of 512x64 pixels per B-scan. In following works, U-Net-like architectures were further adapted to also obtain feedback regarding the model uncertainty [49], improving performance and generalization at costs of higher computation times. A cascade network [50] consisting of a U-Net architecture as well as a fully convolutional network was shown to improve the layer segmentation by combining the U-Net output with an additional relative position map. Borkovkina et al. [51] reported that by reducing the parameters of a conventional U-Net and optimization on the GPU, retinal layer segmentation can be achieved at very high frame rates. Instead of generating layer segmentation maps, Shah et al. [52] directly predicted a position vector of the retinal layer boundaries along the A-scans. However, for this method, a significant amount of training data as well as continuous retinal layer structures across the A-scans are required. Recently, Tran et al. [53] reformulated retinal layer segmentation as a language modeling problem. They split the OCT B-scans in small column-wise bands, such that the known sequence of retinal layers along the incoming scan direction of the A-scans can be modeled as a sequence of words, predicting the current layer by a history of previous and future pixels along the A-scans.

None of these approaches of diagnostic OCT image segmentation includes surgical tools inside the B-scans. In interventional microscope-integrated OCT B-scans, continuous retinal layer structures can not be assumed due to surgical instruments occluding the retinal tissue below. Moreover, diagnostic OCT devices provide significantly better image quality of the cross-sectional B-scans, and many of the segmented retinal layers are not visible in intraoperative B-scans. Therefore, these methods can not easily be transferred to interventional OCT imaging.

To the best of our knowledge, there is no published research on the simultaneous segmentation of multiple retinal layers and a surgical tool in iOCT B-scans. Due to the time constraints, previous intraoperative works using iOCT volumes rely on simple image processing techniques. One main challenge is to maximize the segmentation performance, while keeping the computational costs of the pipeline at a minimum. In the field of cornea surgery, Keller et al. [54] introduce an intraoperative method to segment the surgical tool and the cornea boundaries from iOCT volumes. They process the volume in smaller sub-groups and segment only every other B-scan to save computational time at the cost of under-sampling. Still, the authors report average computation times of 427 ms for processing one volume. We introduce an efficient pipeline, automatically selecting only the relevant B-scans of the volume within a ROI around the needle tip. This enables us to lower the total computation time for the learning-based segmentation of needle surface and retinal layer boundaries required for the tooltip to layer distance estimation.

2. Methods

Our approach for intraoperative distance estimation consists of a sequence of steps. Figure 1 shows an overview of this pipeline. First, a set of 2D projection maps is generated by computing various features along the A-scans of the volume. In the next step, a combination of the generated images serves as a multi-channel input for an instrument segmentation network. From the resulting tool mask, a small area containing the needle tip is identified. Further processing is performed exclusively in this region of interest.
Fig. 1. An overview of our pipeline. (a) A set of 2D feature maps is generated from the newly acquired volume. (b) Instrument segmentation based on the projection images yields a binary tool map. (c) Estimation of a small ROI around the needle tip (indicated by the blue rectangle). (d) The tool and retinal layer boundaries of the selected B-scans within the ROI are segmented. (e) The three point clouds corresponding to instrument, ILM and RPE surface boundaries are generated from the segmentation maps. The shadowed retinal layers are reconstructed considering the surrounding anatomy. (f) Noise is removed from the point clouds by applying Euclidean clustering. The distance between instrument and retinal layers is finally obtained from the resulting point clouds.

The instrument, ILM, and RPE surface segmentation of the relevant B-scans contained in the tip ROI represents the central component of our pipeline. From the cross-sectional segmentation maps, we generate the 3D surface point clouds of the three classes. To cope with tool shadowing artifacts occluding the retina and undetected surfaces, we inpaint the holes in the retinal layer point clouds considering the neighboring surface areas. Afterwards, Euclidean clustering is applied to filter out noise. The cluster with the most points of each class determines the final needle, ILM, and RPE point clouds. Finally, the minimal distance between the tool, the ILM and the RPE point clouds is calculated. In the following sections, the pipeline steps are explained in more detail.

2.1. Instrument tip area localization

The use of 2D projection images from iOCT volumes by computing a set of features for each A-scan has been shown to be an effective way to reduce computational complexity in other applications of real-time 4D iOCT processing [29]. We follow this work and compute the same four enface projection maps encoding average and maximum intensity, argmax and centroid maps. The combined projections serve as a four-channel input image for the instrument segmentation network described in [29] which outputs a binary mask for the instrument. For training the segmentation we use their data set and the techniques outlined in their work. The resulting segmentation mask is afterwards dilated using a small kernel size of $3 \times 3$ pixels, such that the binary map includes the needle as well as the surrounding retinal anatomy.

The purpose of the instrument map is primarily to determine a region around the tooltip, which is exclusively used for all further analysis. As the number of segmented B-scans should be minimized to lower the computation times, the positioning of the instrument is of utmost importance. Figure 2(a) shows the relative positioning of the injection needle to the B-scan...
acquisition direction. By aligning the tip part of the instrument parallel to the cross-sectional images, the number of segmented B-scans can be dramatically reduced. Taking into account the contour, insertion direction and size of the tool, the relevant ROI around the needle tip can be identified from the binary segmentation map. Figure 2(b) shows a minimal rectangle fitted around the instrument contours, which is emphasised in green. The four vertices of the rectangle are separated into two body and two tip vertices by evaluating the distance of the points to the image center: as the needle tip is included in the OCT volume, two vertices are positioned at the border of the image or outside the image, while the remaining two are located inside the image. The tip points $t_1$ and $t_2$ can be identified as the closest points to the image center, shown in Fig. 2(b). From $t_1$ and $t_2$ each two new points are generated. From $t_1$ the points $b_1$ and $b_4$ are generated with $b_1 = t_1 + s_b \cdot v_i$ and $b_4 = t_1 - s_i \cdot v_i$, where $s_b$ and $s_i$ are fixed scalars (0.15 and 0.2) determining the size of the resulting ROI. These parameters were chosen empirically according to the resolution of the OCT scans. Once specified, $s_b$ and $s_i$ do not need to be adjusted, and higher values do not change the final pipeline output, but potentially lead to higher computation times. Further, $v_i$ is the needle insertion vector, defined by the edges of the minimal rectangle, which connect the tip and body vertices. Accordingly, $b_2$ and $b_3$ are generated from $t_2$ with $b_2 = t_2 + s_b \cdot v_i$ and $b_3 = t_2 - s_i \cdot v_i$. Finally, the upright bounding box is obtained from the four newly generated points $b_1$, $b_2$, $b_3$ and $b_4$, illustrated as the blue bounding box in Fig. 2(c). The B-scans containing the instrument and the neighboring anatomy, identified from the masked pixels within the ROI, are afterwards segmented and processed.

2.2. B-scan segmentation

To be applicable to image-guided robotic interventions, the distance calculation approach depends on a rapid segmentation of tool and retinal layers. The high and varying noise levels in iOCT B-scans are challenging for the previously used threshold-based segmentation methods. Therefore, we introduce a more complex segmentation method using a deep-learning approach. The selection of a network that can deliver good segmentation results while keeping the computational costs at a minimum is particularly important for the interventional pipeline. Instead of segmenting the full
ILM and RPE layers, we are only interested in the layer boundaries. Shah et al. [52] report that in order to directly predict the layer boundary position within the A-scans, continuous layer structures and a significant amount of training data are required. They found that UNet architectures are more effective for training small data sets or B-scans without preserved continuous layer structures. As the availability of iOCT data is extremely limited, and the instruments and resulting shadowing artifacts introduce discontinuities in the layer structures, we consequently adopt a U-Net-like architecture to obtain the tool and layer boundary segmentation mask.

We use a UResNet18 as our segmentation network, where the encoder consists of a ResNet18 [55] architecture and an up-convolutional part with similar structure is appended and used as decoder. This architecture is generated using the FastAI [56] dynamic U-Net API. Each encoder-block is connected with the corresponding decoder block via skip connections to preserve high-level features. A combination of focal and Dice loss is used to regulate the high imbalance of the surface classes and the resulting pixel, and to maximize the overlap of the predicted surface boundaries and ground truth labels by maximizing the Dice score. Our data set for training and validation is described in section 3.1. Figure 3 shows an example of an input B-scan and the corresponding labeling of the tool and the retinal layer boundaries. We use 80% of the data set for training and the remaining 20% for validation. Our test set consists of 17 iOCT volumes. The original B-scan images have a resolution of 512x1024 pixels, where each column corresponds to one A-scan. However, during analysis of the previous step, we discovered that the tip ROI does not contain more than 100 A-scans. Therefore, we split each B-scan of the training set in smaller bands of 256 A-scans and scale the axial resolution by half, which reduced the network computation times by a factor of 2.7 compared to training at full B-scan resolution. As a side-effect, we could increase the batch size for training to 12 samples leading to a smoother gradient and, thus, better training. Furthermore, horizontal flipping is used to add more variability to our data set.

![Figure 3](image)

**Fig. 3.** (a) An example of an unprocessed iOCT B-scan containing an injection needle. The incoming light is fully reflected by at the instrument surface leading to shadowing artefacts, such that the retinal structures underneath are not imaged. (b) The ILM surface (red), the anterior surface of the RPE (green) and the needle surface (blue) are overlaid to show an example of the labeling used for training. Both images have an original resolution of 512x1024 pixel and are resized for clear visibility in this figure.

The ResNet18 encoder of the network is pre-trained on ImageNet. We use the AdamW [57] optimizer to update the model weights, first, freeze the encoder, and train five epochs to tune the network’s decoder. Then, the whole network is trained for 15 epochs with a sliced learning rate between $10^{-5}$ and $(3 \cdot 10^{-3})/5$ distributed over the network layers. Finally, the last three decoder
layers are fine-tuned for another five epochs. We use FastAI [56] with Pytorch for training the network. Since there is no related work on the segmentation of retinal layers and instruments in iOCT B-scans, we compare our network to a baseline model for retinal layer segmentation in diagnostic OCT, which is able to generate segmentation maps at high inference times. We evaluate the influence of different loss functions on the surface segmentation and test the final segmentation performance by evaluating the distance output of the complete pipeline showing its robustness to high noise levels in iOCT images.

2.3. **Point cloud processing**

The 3D point clouds of the ILM surface, the anterior RPE boundary, and the needle surface can be generated from the segmented maps. Along each A-scan, the first occurrence of each class is found, its position within the volume is converted to the corresponding location in 3D space, and the new point is added to the respective point cloud.

**Retina reconstruction** The retina point clouds generated in the previous step have some issues in areas where the tissue is not detected correctly or is not imaged due to the shadow of the metallic needle. To reconstruct these regions, a surface depth map of each retinal layer and the corresponding mask indicating the not detected surfaces is obtained from the segmentation maps. An efficient image inpainting method [58] can be applied to fill the missing parts in the depth map using the mask and considering the neighboring depth values. We propose to fill the holes of the retinal layer surface point clouds with the values in the reconstructed depth map. Figure 4 shows the subsequent reconstruction steps. This method is applied to the ILM and the RPE surface point clouds, respectively.

![Fig. 4.](image) Reconstruction of the missing retinal layer surfaces. The following steps are analogously applied to reconstruct the ILM and RPE surface: (a) Shadowing artefacts underneath the tool lead to missing retina surfaces in the initial point clouds. (b) A surface depth map of the retinal layer is generated. (c) Based on this depth map, a mask indicating the missing areas in the point cloud is obtained. (d) The holes in the depth map are filled using the neighboring depth values. (e) Finally, the holes in the retinal layer point clouds are reconstructed from the inpainted depth map.

**Filtering** Applying Euclidean clustering to each of the points clouds and identifying noise as geometric outliers allows for the removal of potential noise in all point clouds before calculating the minimum distance. The voxel size of the volume determines the distance tolerance separating two clusters. For each surface class, the cluster containing the most points is selected as the final point cloud.
Distance calculation  After this fast post-processing step, the minimum distance between the needle tip and the retinal layers can be directly computed from the surface point clouds. By iterating through the points of the tool point cloud, the tool tip point is identified. Since the iOCT scanner is integrated into an operating microscope, the imaging pathway of the retina is restricted via the pupil. In vitreoretinal procedures, the surgeon inserts the instrument through a trocar, directed towards the retina. Therefore, we consider the tip in the tool point cloud as the point with the highest depth value in A-scan direction. To calculate the minimum distance between the tool and the ILM, all tool points to all ILM points are compared and the shortest Euclidean distance as well as the tool point closest to the ILM are obtained. Because a minimal area around the tip is extracted during pre-processing, this final step is computationally very fast.

In case the closest tool point is the needle tip and is located above the ILM, the proposed pipeline recognizes that the tool is located above the retina and returns the estimated distance. Otherwise, instrument contact with the retina is assumed if the minimum distance between the tool and the ILM point cloud is smaller than a threshold defined relative to the voxel size. Further, to detect whether the ILM has been penetrated, the point of the tool point cloud closest to the ILM is compared to the needle tip point. If the closest tool point is not the tip point, we can assume the needle has penetrated the layer, since, consequently, the needle tip has to be located below the ILM surface point cloud. We apply the same analysis to estimate the distance between the tool and the RPE surface point cloud and to detect the contact between the needle and the RPE.

3. Experimental setup and evaluation methods

In the following sections we describe the iOCT data sets we used in our experiments for training, validation and testing. We further introduce the metrics for the evaluation of the B-scan segmentation performance, as well as for the validation of the final pipeline outputs. Finally, we describe the three loss functions that were considered for comparison in our experiments.

3.1. Materials

The training and validation set, as well as the test set for our experiments consist of iOCT B-scans from ex-vivo porcine eyes acquired with a Rescan 700 (Carl Zeiss Meditec, Jena) iOCT system integrated into an operating microscope. Each volume consist of a total of 128 B-scans with a resolution of 512×1024 pixels. To generate the ground truth B-scan segmentation maps, the surface boundaries of the ILM and the anterior RPE as well as the needle surface are manually labeled under supervision of a retinal expert. The data set used for training and validation consists of B-scans including microsurgical needles as well as the retinal anatomy. From 75 iOCT volumes acquired from 22 ex-vivo porcine eyes, 595 B-scans where selected showing the instrument and its immediate vicinity. Needles with 41G and 27G tip were used as instruments. The volumes are acquired at scan sizes of 3×3 and 5×5 mm in width and height at a scan depth of 2mm.

We obtained two different data sets for testing. In the first, we used an INCYTO Needle-RNT for subretinal injection with a 23G body and 41G tip. We acquired four volumes with the tip above the ILM, and three volumes with the needle positioned between the ILM and the RPE. This data set uses a realistic needle diameter as well as an intact anterior segment of the porcine eye. However, the anterior segment structures deteriorate rapidly post mortem in porcine eyes leading to relatively poor image quality in this data set. We call this data set our Low Quality data set, as it may represent cases of challenging intraoperative scenarios. The second data set consists of 10 volumes acquired with a 27G needle located above the retina. In this data set we removed the anterior segment ("open sky") to improve the OCT image quality. By this effort we have created B-scans that are more representative of the usual iOCT scan quality during in-vivo
surgery. We refer to this as our High Quality data set in the following discussions. Figure 5 shows a representative B-scan for each of the two data sets.

![B-scans](image)

**Fig. 5.** The two data sets differ in noise levels and thus in the quality of the iOCT B-scans. (a) An example B-scan of the Low Quality data set. The scans suffer from high levels of speckle noise and diminished signal, leading to reduced instrument visibility and a less distinctive transition between vitreous and retinal layers. (b) An example B-scan of the High Quality data set, which, compared to the first data set, is less corrupted by the typical speckle noise of OCT and exhibits clearer instrument visibility and higher contrast between vitreous and retina.

### 3.2. Evaluation metrics

To evaluate the model performance, we introduce three metrics measuring the detection and positional error of the predicted segmentation masks. Since each A-scan corresponds to a column in the B-scans, we calculate the average detection accuracy for each class in the B-scan columns, referred to as $ACC_{Tool}$, $ACC_{ILM}$, and $ACC_{RPE}$. If a class was correctly detected within the A-scan, we calculate the L1 error between the output and the ground truth location of the surface, referring to the row index of the first occurrence along the A-scan scanning direction. Accordingly, we refer to the surface errors of the classes as $L_1_{Tool}$, $L_1_{ILM}$, and $L_1_{RPE}$, respectively. The standard deviations of the L1 errors for the three classes are consequently referred to as $SD_{Tool}$, $SD_{ILM}$ and $SD_{RPE}$.

The most important aspect of the proposed system is the accuracy of the distance calculation between tooltip and the retinal layers. Therefore, we evaluate the end-to-end performance by determining the Euclidean distance error between the pipeline outputs and the ground truth distances. To obtain the ground truth distance between the tool and the two retinal layers in the two test sets described in section 3.1, we generate the 3D point clouds of the manually generated ground truth B-scan segmentation maps. The final ground truth distance is then determined as the minimum Euclidean distance between tool and retinal layer point clouds without applying additional post-processing.

### 3.3. Loss functions

As the tool surface and the retinal layer boundaries represent only small parts within the B-scans, the classes are highly imbalanced. Addressing these issues, in our experiments we investigate the behaviour of three loss functions for imbalanced data sets. One possibility to address class imbalance is to use a weighted cross-entropy (WCE) loss, where classes with low occurrence in the
data set are weighted higher than dominant background classes. We apply an inverse weighting of the class probabilities $p_{\text{tool}}$, $p_{\text{ilm}}$, and $p_{\text{rpe}}$ of the tool, ILM and RPE class as well as the probability $p_{\text{res}}$ of class containing all residual pixels, leading to a weighting with $(1 - p_{\text{tool}})$, $(1 - p_{\text{ilm}})$, $(1 - p_{\text{rpe}})$ and $(1 - p_{\text{res}})$ for the respective classes. An effective alternative for segmentation with imbalanced classes is the focal loss function [59] introduced in 2017, which is defined as:

$$FL(p_t) = -\alpha(1 - p_t)^\gamma \log(p_t)$$  \hspace{1cm} (1)$$

The parameters $\alpha$ and $\beta$ are hyper-parameters and can be fine-tuned. The function assigns smaller weights to easy examples and focuses on learning harder examples. As the third loss function we deploy a combination of the focal and Dice loss [60]. The combination of distribution- and region-based loss functions has been shown to improve the model performance in previous works [48,50]. As the Dice loss optimizes the Dice score and therefore the overlap between the model output and the ground truth segmentation, it was shown to work well for the layer-boundary segmentation problem. We weigh the focal and Dice loss equally and use their sum as the final combined loss function.

4. Results

To evaluate our proposed system, we separately investigate the B-scan segmentation performance and the final distance outputs of the pipeline. In the next sections we first evaluate the B-scan segmentation of the UResNet18 and compared it to a baseline model for diagnostic OCT layer segmentation, as well as other segmentation networks. Subsequently, we evaluate the end-to-end distance estimation between the needle and the two retinal layers on 17 iOCT volumes. Furthermore, we analyse the influence of different loss functions on the B-scan segmentation, as well as on the final distance estimation and investigate the robustness to the varying noise levels of the OCT scans. Finally, we show the feasibility of our method for the interventional use case by evaluating the computation times of the pipeline and its individual components.

4.1. B-scan layer surface segmentation

To evaluate segmentation of the iOCT B-scans, we first compare the three different loss functions defined in section 3.3 regarding their performance on the instrument and retinal layer segmentation. We then compare the UResNet18 to a baseline model for retinal layer segmentation in diagnostic OCT B-scans, as well as a standard U-Net and a network for real-time semantic segmentation.

4.1.1. Loss function evaluation

Since the tool surface and the retinal layer boundary classes are highly imbalanced, the choice of the loss function is important to achieve a good segmentation performance. Analyzing our training set, the occurrences of tool, ILM and RPE surfaces have shown to be very low, with corresponding probabilities of $p_{\text{tool}} = 0.203\%$, $p_{\text{ilm}} = 0.752\%$ and $p_{\text{rpe}} = 0.526\%$. Consequently, the residual pixels, which do not belong to any of these surface classes, have a class probability of $p_{\text{res}} = 98.519\%$. For performance comparison of the three loss functions specified in section 3.3, we assign the weights of the WCE loss according to these probabilities. During hyper-parameter tuning, we determined the best values for the parameters of the focal loss function, $\alpha$ and $\gamma$, to be 0.95 and 1.0, respectively. The same values are applied to the parameters of the focal loss within the combined focal and Dice loss function. To compare the suitability of the three loss functions, we compute the average surface detection accuracy for the three classes, as well as the average positional error and the standard deviation of the segmentations within the A-scans, as described in section 3.2. Table 1 shows the results of the comparison. The loss functions have similar class detection accuracy and only differ slightly in the positional error of the segmented surface boundaries. From our results, we conclude that all three discussed loss functions represent viable
options for training this problem. However, the most important aspect is their impact on the final distance estimates between needle tip and the retinal layers, which we evaluate in section 4.2.

Table 1. Performance comparison of the U-Net-style architecture with ResNet18 encoder trained on different loss functions. The average detection accuracy per A-scan and the average pixel error of the detected and the ground truth surface locations along the A-scans are evaluated for the three surface classes.

| Loss Function | ACC Tool | ACC ILM | ACC RPE | L1 Tool | L1 ILM | L1 RPE | SD ILM | SD RPE | SD Tool |
|---------------|----------|---------|---------|---------|--------|--------|--------|--------|---------|
| WCE Loss      | 96.0%    | 90.8%   | 99.8%   | 5.0px   | 5.3px  | 1.0px  | 7.8px  | 0.3px  | 2.9px   |
| Focal Loss    | 95.1%    | 90.0%   | 99.7%   | 4.1px   | 4.4px  | 1.2px  | 7.1px  | 0.2px  | 2.7px   |
| Focal+Dice Loss | 95.2%   | 90.6%   | 99.6%   | 4.0px   | 5.0px  | 1.7px  | 7.7px  | 0.2px  | 2.3px   |

4.1.2. Model architecture comparison

As there is no published research on retinal layer and tool surface segmentation in iOCT B-scans, we compare our segmentation network with three baseline semantic segmentation networks: ReLayNet [48], a baseline model for retinal layer segmentation in diagnostic OCT, standard U-Net [61], as well as ERFNet [62], which is specifically designed for real-time semantic segmentation. As speed plays an important role in our application, we assess the inference speed of the models. In addition to the above accuracy and positional metrics, we obtain the average inference times in a python environment without optimization, emphasizing the interplay between number of parameters and network performance. These metrics are reported in Table 2.

Table 2. Comparison of the UResNet18 architecture to ReLayNet, a baseline model for retinal layer segmentation in diagnostic OCT B-scans, a standard U-Net and ERFNet, a light-weight network for real-time semantic segmentation. The number of network parameters is specified in million (M) and the segmentation performance as well as the average network inference times are evaluated.

|         | Params | ACC Tool | ACC ILM | ACC RPE | L1 Tool | L1 ILM | L1 RPE | SD ILM | SD RPE | SD Tool | Time   |
|---------|--------|----------|---------|---------|---------|--------|--------|--------|--------|---------|--------|
| U-Net   | 13.3 M | 94.8%    | 90.1%   | 99.5%   | 9.6px   | 5.0px  | 1.7px  | 7.2px  | 0.6px  | 3.6px   | 28.3ms |
| ERFNet  | 2.0 M  | 95.6%    | 90.3%   | 99.7%   | 8.0px   | 5.2px  | 1.2px  | 7.3px  | 0.6px  | 3.8px   | 22.8ms |
| ReLayNet| 8.6 M  | 95.0%    | 90.1%   | 99.4%   | 5.8px   | 5.5px  | 2.0px  | 7.8px  | 0.7px  | 3.4px   | 25.2ms |
| UResNet18 | 31.1 M | 95.2%    | 90.6%   | 99.6%   | 4.0px   | 5.0px  | 1.7px  | 7.7px  | 0.2px  | 2.3px   | 32.5ms |

Overall, the ILM, RPE and tool surface classes are detected with similar accuracy across all networks. The L1 errors of ILM and RPE are comparable and differ only slightly. The UResNet18 and UNet share the smallest error regarding the ILM segmentation, while the ERFNet reaches the smallest RPE segmentation error. The networks especially show a difference in the segmentation accuracy of the tool surface. The UResNet18 achieves the lowest L1 tool error and clearly outperforms the other networks, while the standard U-Net shows the highest error. Similarly, the lowest standard deviations of the tool and RPE errors are achieved by the UResNet18. In contrast, comparing the average network speed, the ERFNet reaches the lowest inference time, while the UResNet18, containing the most parameters, also is the most computationally intensive network. Figure 6 shows examples comparing the UResNet18 and the ERFNet outputs with the manual ground truth segmentations. In both examples, the networks can segment the two retinal layers similarly well. However, the ERFNet is not able to generate a good segmentation of the tool. The second row of Fig. 6 shows a challenging B-scan example with the tool inserted into the retina. While the ERFNet fails to detect the tooltip, the UResNet18 is able to determine the pixels at the tip. The false positives of the tool class are filtered out during the subsequent point cloud processing step.
Fig. 6. Examples of output and ground truth segmentations are shown overlaid on the B-scans. The outputs of the ERFNet (first column) and the UResNet18 (second column) are compared to the manually annotated ground truth (third column). Both cases show the advantage of the UResNet18 over the ERFNet. The second row shows a challenging B-scan example with the needle inserted into the retina. ERFNet fails to detect most of the needle inside the retina while UResNet18 detects the tip but also generates false positives below. These are filtered out in the subsequent point cloud processing step.

4.2. End-to-end evaluation

As the end result of our pipeline is the distance between the needle tip and the ILM as well as the anterior RPE surface, we evaluate the Euclidean distance errors on 17 iOCT volumes acquired from ex-vivo porcine eyes.

The pipeline output is tested by calculating the error between the estimated and ground truth distances on both test sets. Figure 7 shows the error of estimating the ILM and RPE distances and the influence of the three loss functions specified in section 4.1.1 on the final pipeline output. The best results were achieved using the segmentation model trained with combined focal and Dice loss with an average error of 9.24 $\mu$m, a median error of 10.12 $\mu$m, a standard deviation of 5.44 $\mu$m and a maximum error of 17.03 $\mu$m. Analogously, the distance estimation to the RPE surface boundary with the same model achieves an average error of 8.61 $\mu$m, a median error of 8.78 $\mu$m, a standard deviation of 6.22 $\mu$m and a maximum error of 16.98 $\mu$m.

Furthermore, we separately evaluate the distance errors in our Low Quality and High Quality data sets. Figure 8 shows that the errors of the scans with lower noise levels generally have a lower error variance and less outliers. The UResNet18 trained with the weighted cross entropy loss shows the highest errors as well as the highest variances, while the combination of focal and Dice loss is the most robust to varying noise levels and yields the best overall distance estimates.

Although in Table 1 the model trained on the focal loss function achieves a smaller positional error detecting the ILM and RPE, Fig. 7 and 8 show that the combination of focal and Dice loss leads to overall more robust distance estimates and, hence, is selected for our pipeline. Finally, we evaluate the effect of the point cloud processing by comparing the distance estimates of the full pipeline with the estimates of the pipeline without retina reconstruction and filtering. The results in Fig. 9 show that the described point cloud processing is essential for robust distance estimation.
Fig. 7. Evaluation of the final pipeline distance estimations. We compare the UResNet18 trained on three different loss functions regarding their influence on the final distance errors of our pipeline. We separately evaluate the distance error to the ILM and RPE layer surface boundaries. The errors are given in micrometer.

Fig. 8. Influence of different loss functions on the pipeline performance evaluated on OCT scans with low and high noise levels, separately.

Fig. 9. Influence of the point cloud processing on the final distance estimation evaluated on our Low Quality data set containing challenging segmentation examples. The retina reconstruction and point cloud filtering stages are bypassed to compute the distance error from the unprocessed point clouds. For better illustration, we have omitted one data point of the unprocessed RPE distance error of value of 422 \( \mu m \) from the figure. Outliers are mostly caused by false positive in the segmentation, therefore the noise filtering on the point clouds can effectively improve the robustness of the distance estimation.
4.3. Time profiling

A constraining factor in the design of our pipeline was the requirement to provide update rates suitable for image-guided robotic surgery. The Carl Zeiss Meditec Rescan 700 iOCT system used in our experiments has an acquisition speed of 27000 A-scans per second, which is not suitable for interactive volumetric acquisitions. However, latest advances in OCT technology reach near video-rate volumetric imaging. Carrasco-Zevallos et al. [27] employed a 4D OCT system with an update rate of 15 Hz and achieved to simulate OCT-guided retinal surgery. We believe that such an update rate would meet the requirements of image-guided surgery and would also be sufficient for our pipeline. We use NVIDIA’s TensorRT to optimize our model for inference on the GPU, leveraging the layer fusion and kernel optimization strategies to optimize the model. We did not use any strategies that could potentially compromise segmentation accuracy. The combined optimizations and execution on the GPU decreases the inference time to 13 ms per B-scan. Table 3 shows the average computation times and the standard deviations of the individual pipeline components on our system (Intel Core i9-9920X @3.5GHz and NVidia GeForce RTX 2080 Ti). For this experiment we use the data set *Low Quality*, because it is most representative of the real surgical scenario in terms of needle diameter and orientation. Since for this experiment we are only interested in the time analysis and not the pipeline output, we added 10 iOCT volumes from our training set with similar instrument properties in order to improve the accuracy of the time analysis. By providing distance feedback every 63.82 ms on average, leading to an update rate of 15.66 Hz, we consider our method as suitable for the intraoperative use case. During this experiment we observed a standard deviation of 4.99 ms for the speed of the overall pipeline, with a performance of 67.72 ms in the worst case and 53.20 ms in the best case.

Table 3. Average processing times and standard deviations (SD) in ms for each pipeline step as well as the average update rate [Hz] given input iOCT volumes of 128 B-scans, each 512 × 1024, and 4 segmented B-scans at the tooltip ROI.

|                 | 2D Proj | Tool Seg | ROI | B-scan Seg | PC Gen | Filter | Dist Comp | Total | Update [Hz] |
|-----------------|---------|----------|-----|------------|--------|--------|-----------|-------|-------------|
| **Mean**        | 0.38    | 5.04     | 0.07| 52.60      | 1.68   | 1.66   | 2.37      | 63.82 | 15.66       |
| **SD**          | 0.08    | 0.35     | 0.01| 4.62       | 0.43   | 0.45   | 1.12      | 4.99  | –           |

5. Discussion

In our experiments we have evaluated different networks and loss functions with respect to their segmentation performance of two retinal layer boundaries and the instrument surface in iOCT B-scans. Both, Table 1 and Table 2 show a lower L1 error with respect to the ILM compared to the RPE surface. The difference in the ILM and RPE segmentation performance could be attributed to the more visible and smoother RPE surface compared to the ILM, which often exhibits high surface curvatures due to vessels and deformations, as well as lower intensities at A-scans close to the tool shadow. However, the most important aspect for the application of the pipeline is to minimize the positional error of the segmented tool surface, since it is strongly related to the error between the tool tip and the retinal layers and therefore, also to the final error of our system. In Table 2, compared to the other networks, the UResNet18 shows the lowest tool L1 error. Additionally, fast segmentation networks with less parameters, such as the ERFNet, could not detect the needle tip in some cases of our test set (c.f. Figure 6) which would lead to very high overall distance errors. We favour robustness over fast computation times and use the UResNet18 as the final model for our pipeline showing the best tool segmentation performance.

The speed of our pipeline depends partly on the number of B-scans that have to be segmented to generate the point clouds within the ROI. To minimize this number, we position the OCT scan area such that the B-scans are generated parallel to the tool tip direction. With the small 0.1 mm diameter of 41 gauge subretinal injection needles, on average, four B-scans including
the instrument and neighboring retinal tissue are segmented, assuming a scan area of 5x5 mm
and 128 B-scans per volume. In future work, we will investigate dynamically re-positioning the
OCT scan area by minimizing the angle between the tool insertion direction within the ROI and
the B-scan acquisition direction (Fig. 2(c)) to keep the computational costs at a minimum at
all times. Recent technical advances in OCT systems have pushed A-scan rates to 400 kHz for
microscope-integrated systems [63] and enabled video rate volumetric imaging with updates rates
of 24.2 Hz [28] based on A-scan rates of several GHz. In [27], the authors show that a volumetric
update rate of 15 Hz, acquiring a new OCT volume every 6 ms, is sufficient for 4D OCT guided
surgery and is able to clearly visualize surgical instruments and manipulation of tissue. Our
proposed method can cope well with the fast update rates required for image-guided surgeries by
achieving average computation times of 63.82 ms per volume and can provide distance feedback at
the suggested 15 Hz [27]. The segmentation of the B-scans remains the computational bottleneck
of the pipeline with an average speed of 52.60 ms for four segmented B-scans. The next step
could be to develop faster segmentation methods for tool and retinal layer segmentation in iOCT
B-scans without compromising performance. In this work, we did not leverage all optimization
strategies that TensorRT offers. Borkovkina et al. [51] have reported a speedup of 18x when
using TensorRT optimization methods including reduced precision using INT8. This could be
an avenue to further optimize the proposed pipeline, however careful measures to preserve the
good end-to-end accuracy need to be taken. Compared to the intraoperative pipeline for cornea
surgery presented in [54], the per B-scan segmentation in our application is computationally
more expensive, however, we can effectively reduce the number of segmented B-scans through
the ROI estimation. Further downscaling of the input B-scans to improve the segmentation speed
might result in loss of important instrument information. Also, the segmentation of the cornea
can not easily be compared with the segmentation of retinal layers and tool, since the retinal
layers can exhibit more complex structures, for example introduced through vessels. As our
results have shown, a larger network is important for robust and precise instrument segmentation.

Novel OCT scanning technologies enabled BC-mode [64] imaging, in which multiple sparsely
sampled B-scans are combined to generate a single cross-sectional image with enhanced
instrument and tissue visibility and reduced shadowing artifacts. Such advances have the potential
to improve the segmentation performance of intraoperative OCT by improving the visibility
of surgical tools and retinal structures. The development of dedicated and OCT compatible
instruments for vitreoretinal surgery [65] could additionally improve the visibility of surgical
tools in the B-scans and thus also lead to an improved tool segmentation performance.

The immediate application of the presented pipeline is to precisely and continuously monitor
the distance between tooltip, ILM and RPE, providing data to a robot controller. This information
can guide the robot to reduce the risk associated with subretinal injection. The total processing
time of our pipeline in this scenario is a limiting factor for the robot speed, as the tool tip motion
between two updates cannot be too large when safe motion and clinical grade precision need to
be achieved. Assuming an OCT update rate of 15 Hz, a target area of 25 µm [14] and a distance
estimation accuracy of 10 µm, the needle tip is not allowed to move more than 15 µm in the time
it takes to acquire (~66 ms) and process (~63 ms) the OCT data, in order to avoid accidental
penetration of the RPE once the robot has reached the target area. This results in a maximum
safe speed of ~0.1 mm/s in axial direction regarding the OCT coordinate system. The effective
maximum safe speed of the robot then depends on the incident angle between the needle and
the RPE surface. In an optimal scenario, if the needle starts at an assumed safe distance of 2.5
mm from the retinal surface, the total time to approach the target area is less than 30 seconds,
assuming a retinal thickness of less than 500 µm [13]. In a realistic scenario the robot control
would likely slow down the needle while approaching the target area, however this shows that
in closed-loop robotic targeting, the processing time of our pipeline will not impose strong
limitations on the clinical workflow. However, these estimates will have to be verified once our
system is combined with a closed loop robotic control to form a semi-autonomous injection
system, which we consider the next step for this work.

A possible extension of this work could be to use the generated point clouds to estimate the
current tool motion direction by fitting a line to the tool point cloud (Fig. 10(a)). By combining
the tool motion direction with the segmentations of ILM and RPE, one can estimate the expected
point of contact with both retinal layers and calculate the distance along the trajectory until
the retinal surface is reached. As the tool should be inserted to a defined depth, which can be
determined during surgical planning, the target depth of the needle tip for the injection can be
obtained from the live data as a relative position between ILM and RPE. With the tool point
cloud and the fitted motion direction, the proposed pipeline can estimate when the needle tip will
reach this target layer (Fig. 10(b)).

![Fig. 10. (a) Fitting of the tooltip motion direction (yellow line) to the needle point cloud
and estimation of the contact point of tool tip with the retinal layer boundaries. The ILM
and RPE surfaces are visualized in red and green, respectively. (b) A virtual layer (shown as
the yellow point cloud) is augmented between the ILM and the RPE to visualize the target
insertion depth of the needle. The virtual layer is defined as a relative position between the
ILM and RPE surface boundary and can be obtained during surgery planning.]

6. Conclusion

In this paper, we proposed a pipeline to estimate the distance between the tip of a subretinal
injection needle and two retinal layers, the ILM surface and the anterior surface of the RPE, from
iOCT volumes. First, the tool surface and the two retinal layer boundaries in selected B-scans
around the needle tip are segmented. In an efficient pre-processing step, we propose to reduce
the newly acquired OCT volume to a minimal area around the needle tip, including only a few
B-scans, which allows one to use a model-based tool and layer segmentation of the relevant
volume area at update rates around 15 Hz. Our pipeline achieves an average error of 9.24 \( \mu m \) and
8.61 \( \mu m \) and a standard deviation of 5.44 \( \mu m \) and 6.22 \( \mu m \), for the distance between the needle tip
and the ILM, and RPE surface, respectively. Automatic distance feedback between instrument
tip and retinal layers has many applications for robotic subretinal injection. The distance to the
ILM determines the control strategy, as once the retinal surface is touched, the robot motion
is highly restricted. On the other hand, the distance to the RPE defines the maximum robot
motion before harming significant retinal and retinal support cells. We believe such a pipeline
can deliver important feedback to both surgeon and robot during subretinal injection procedures
and be especially useful for the development of an eventual autonomous robotic approach.

**Funding.** National Institutes of Health (1R01EB025883-01A1).
Acknowledgments. The datasets of this work were collected from Lumera 700 Zeiss ophthalmic microscope with integrated Rescan 700 iOCT engine, Carl Zeiss Meditec AG supported this work by helping to collect the data.

Disclosures. PG: Research to Prevent Blindness, New York, New York, USA (F), J. Willard and Alice S. Marriott Foundation, the Gale Trust, Mr. Herb Ehlers, Mr. Bill Wilbur, Mr. and Mrs. Rajandre Shaw, Ms. Helen Nassif, Ms Mary Ellen Keck, Don and Maggie Feiner, Dick and Gretchen Nielsen, and Mr. Ronald Stiff (R)

References

1. R. Casten, B. W. Rovner, and J. L. Fontenot, “Targeted vision function goals and use of vision resources in ophthalmology patients with age-related macular degeneration and comorbid depressive symptoms,” J. Vis. Impair. & Blind. 110(6), 413–424 (2016).

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 Glob. Heal. 2(2), e106–e116 (2014).

3. R. P. Finger, V. Dainen, B. M. Eldem, J. S. Talks, J.-F. Korobelnik, P. Mitchell, T. Sakamoto, T. Y. Wong, K. Pantiri, and J. Carrasco, “Anti-vascular endothelial growth factor in neovascular age-related macular degeneration—a systematic review of the impact of anti-VEGF on patient outcomes and healthcare systems,” BMC Ophthalmol. 20(1), 294 (2020).

4. M. R. Alexandru and N. M. Alexandra, “Wet age related macular degeneration management and follow-up,” Rom. J. Ophthalmol. 60, 9–13 (2016).

5. Y. Wang, Z. Tang, and P. Gu, “Stem/progenitor cell-based transplantation for retinal degeneration: a review of clinical trials,” Cell Death Disc. 11(9), 793 (2020).

6. K. Xue, M. Gropppe, A. P. Salvetti, and R. E. MacLaren, “Technique of retinal gene therapy: delivery of viral vector into the subretinal space,” Eye 31(9), 1308–1316 (2017).

7. E. P. Rakocy, C.-M. Lai, A. L. Magno, M. E. Wikstrom, M. A. French, C. M. Pierce, S. D. Schwartz, M. S. Blumenkranz, T. W. Chalberg, M. A. Degli-Eposito, and I. J. Constable, “Gene therapy with recombinant adeno-associated vectors for neovascular age-related macular degeneration: 1 year follow-up of a phase 1 randomised clinical trial,” Lancet 386(10011), 2395–2403 (2015).

8. S. J. Gasparini, S. Llonch, O. Borsch, and M. Ader, “Transplantation of photoreceptors into the degenerative retina: Current state and future perspectives,” Prog. Retinal Eye Res. 69, 1–37 (2019).

9. E. J. T. van Zeeburg, K. J. M. Maaijwee, T. O. A. R. Missotten, H. Heimann, and J. C. van Meurs, “A free retinal pigment epithelium-choroid graft in patients with exudative age-related macular degeneration: results up to 7 years,” Am. J. Ophthalmol. 153(1), 120–127.e2 (2012).

10. M. Zarbin, I. Sugino, and E. Townes-Anderson, “Concise review: update on retinal pigment epithelium transplantation for age-related macular degeneration,” Stem Cells Translational Medicine 8(5), 466–477 (2019).

11. S. H. Chung, I. N. Molhoff, U. Nguyen, A. Nguyen, N. Stucka, E. Tieu, S. Manna, R. K. Meleppat, P. Zhang, E. L. Nguyen, J. Fong, R. Zawadzki, and G. Yiu, “Factors impacting efficacy of AAV-mediated CRISPR-based genome editing for treatment of choroidal neovascularization,” Mol. Therapy. Methods and Clinical Development 17, 409–417 (2020).

12. 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 (2000), pp. 59–60 vol.1.

13. Y.-J. Jo, D.-W. Heo, Y.-I. Shin, and J.-Y. Kim, “Diurnal variation of retina thickness measured with time domain and spectral domain optical coherence tomography in healthy subjects,” Invest. Ophthalmol. Vis. Sci. 52(9), 6497–6500 (2011).

14. M. Karampelas, D. A. Sim, P. A. Keane, V. P. Papastefanou, S. R. Sadda, A. Tufail, and J. Dowler, “Evaluation of the IRISS and feasibility studies in an animal model,” Eye 31(9), 2013.

15. R. Taylor, P. Jensen, L. Whitcomb, A. Barnes, R. Kumar, D. Stovinovici, P. Gupta, Z. Wang, E. deJuan, and L. Kavoussi, “A Steady-Hand Robotic System for Microsurgical Augmentation,” in Medical Image Computing and Computer-Assisted Intervention – MICCAI’99, C. Taylor and A. Colchester, eds. (Springer Berlin Heidelberg, 1999), pp. 1031–1041.

16. E. Rahimy, J. Wilson, T.-C. Tsao, S. Schwartz, and J.-P. Hubschman, “Robot-assisted intracocular surgery: development of the IRIS and feasibility studies in an animal model,” Eye 27(8), 972–978 (2013).

17. C. Song, P. L. Gehlbach, and J. U. Kang, “Active tremor cancellation by a Smart handheld vitreoretinal microsurgical tool using swept source optical coherence tomography,” Opt. Express 20(21), 23414–23421 (2012).

18. 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 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), (2013), pp. 5682–5685.

19. A. Gijbels, E. B. Vander Poorten, B. Gorissen, A. Devreker, P. Stalmans, and D. Reynaerts, “Experimental validation of a robotic comanipulation and telemanipulation system for retinal surgery,” in 5th IEEE RAS/EMBS International Conference on Biomedical Robotics and biomechatronics, (2014), pp. 144–150.
20. A. Barthel, D. Trematerra, M. A. Nasseri, D. Zapp, C. P. Lohmann, A. Knoll, and M. Maier, “Haptic interface for robot-assisted ophthalmic surgery,” in 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), (2015), pp. 4906–4909.

21. T. L. Edwards, K. Xue, H. C. M. Meenink, M. J. Beelen, G. J. L. Naus, M. P. Simunovic, M. Latasiewicz, A. D. Farinry, M. D. de Smet, and R. E. MacLaren, “First-in-human study of the safety and viability of intraocular robotic surgery,” Nat. Biomed. Eng. 2(9), 649–656 (2018).

22. N. Rieke, D. J. Tan, C. Amat di San Filippo, F. Tombari, M. Alshekahali, V. Belagiannis, A. Eslami, and N. Navab, “Real-time localization of articulated surgical instruments in retinal microsurgery,” Med. Image Anal. 34, 82–100 (2016).

23. J. P. Ehlers, Y. S. Modi, P. E. Pecen, J. Goshe, W. J. Dupps, A. Rachitskaya, S. Sharma, A. Yuan, R. Singh, P. K. Kaiser, J. L. Reese, C. Calabrisie, A. Watts, and S. K. Srivastava, “The DISCOVER study 3-year results: feasibility and usefulness of microscope-integrated intraoperative OCT during ophthalmic surgery,” Ophthalmology 125(7), 1014–1027 (2018).

24. Y. Li, W. Zhang, V. P. Nguyen, R. Rosen, X. Wang, X. Xia, and Y. M. Paulus, “Real-time OCT guidance and multimodal imaging monitoring of subretinal injection induced choroidal neovascularization in rabbit eyes,” Exp. Eye Res. 186, 107714 (2019).

25. N. Z. Gregori, B. L. Lam, and J. L. Davis, “Intraoperative use of microscope-integrated optical coherence tomography for subretinal gene therapy delivery,” Retina 39(Suppl. 1), S9–S12 (2019).

26. O. Carrasco-Zevallos, B. Keller, C. Viehland, P. Hahn, A. N. Kuo, P. J. DeSouza, C. A. Toth, and J. A. Izatt, “Real-time 4D visualization of surgical maneuvers with 100khz swept-source Microscope Integrated Optical Coherence Tomography (MI OCT) in model eyes,” Investigative Ophthalmol. & Vis. Sci. 55, 1633 (2014).

27. O. M. Carrasco-Zevallos, C. Viehland, B. Keller, R. P. McNab, A. N. Kuo, and J. A. Izatt, “Constant linear velocity spiral scanning for near video rate 4D OCT ophthalmic and surgical imaging with isotropic transverse sampling,” Biomed. Opt. Express 9(10), 5052–5070 (2018).

28. J. P. Kolb, W. Draxinger, J. Klee, T. Pfeiffer, M. Eibl, T. Klein, W. Wieser, and R. Huber, “Live video rate volumetric OCT imaging of the retina with multi-MHz A-scan rates,” PLoS One 14(3), e0213144 (2019).

29. 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, A. L. Martel, P. Abolmaesumi, D. Stoyanov, D. Mateus, M. A. Zuluaga, S. K. Zhou, D. Racocoeau, and L. Joskowicz, eds. (Springer International Publishing, 2020), pp. 267–276.

30. H. Roodaki, K. Filippatos, A. Eslami, and N. Navab, “Introducing augmented reality to optical coherence tomography in ophthalmic microsurgery,” in 2015 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), (2015), pp. 1–6.

31. Y. Li, C. Chen, X. Huang, and J. Huang, “Instrument tracking via online learning in retinal microsurgery,” in Medical Image Computing and Computer-Assisted Intervention – MICCAI 2014, P. Golland, N. Hata, C. Barillot, J. Hornegger, and R. Howe, eds. (Springer International Publishing, 2014), pp. 464–471.

32. N. Rieke, D. J. Tan, M. Alshekahali, F. Tombari, C. A. di San Filippo, V. Belagiannis, A. Eslami, and N. Navab, “Surgical tool tracking and pose estimation in retinal microsurgery,” in Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, N. Navab, J. Hornegger, W. M. Wells, and A. Frangi, eds. (Springer International Publishing, 2015), pp. 266–273.

33. R. Sznitman, K. Ali, R. Richa, R. H. Taylor, G. D. Hager, and P. Fua, “Data-driven visual tracking in retinal microsurgery,” in Medical Image Computing and Computer-Assisted Intervention – MICCAI 2012, N. Ayache, H. Delingette, P. Golland, and K. Mori, eds. (Springer Berlin Heidelberg, 2012), pp. 568–575.

34. M. Zhou, K. Huang, A. Eslami, H. Roodaki, D. Zapp, M. Maier, C. P. Lohmann, A. Knoll, and M. A. Nasseri, “Precision needle tip localization using optical coherence tomography images for subretinal injection,” in 2018 IEEE International Conference on Robotics and Automation (ICRA), (2018), pp. 4033–4040.

35. M. Zhou, H. Roodaki, A. Eslami, G. Chen, K. Huang, M. Maier, C. Lohmann, A. Knoll, and M. A. Nasseri, “Needle segmentation in volumetric optical coherence tomography images for ophthalmic microsurgery,” Appl. Sci. 7(8), 748 (2017).

36. M. Zhou, X. Wang, J. Weiss, A. Eslami, K. Huang, M. Maier, C. P. Lohmann, N. Navab, A. Knoll, and M. A. Nasseri, “Needle localization for robot-assisted subretinal injection based on deep learning,” in 2019 International Conference on Robotics and Automation (ICRA), (2019), pp. 8727–8732.

37. C. Song, P. L. Gehlbach, and J. U. Kang, “Active tremor cancellation by a Smart handheld vitreoretinal microsurgical tool using swept source optical coherence tomography,” Opt. Express 20(21), 23414–23421 (2012).

38. H. Yu, J. Shen, K. M. Joos, and N. Simaan, “Design, calibration and preliminary testing of a robotic telemanipulator for OCT guided retinal surgery,” in 2013 IEEE International Conference on Robotics and Automation, (2013), pp. 225–231.

39. X. Liu, M. Balicki, R. H. Taylor, and J. U. Kang, “Towards automatic calibration of Fourier-Domain OCT for robot-assisted vitreoretinal surgery,” Opt. Express 18(23), 24331–24343 (2010).

40. G.-W. Cheon, Y. Huang, and J. U. Kang, “Active depth-locking handheld micro-injector based on common-path swept source optical coherence tomography,” in Optical Fibers and Sensors for Medical Diagnostics and Treatment Applications XV, vol. 9317 I. Gannot, ed., International Society for Optics and Photonics (SPIE, 2015), pp. 123–127.
41. T. Fabritius, S. Makita, M. Miura, R. Myllylä, and Y. Yasuno, “Automated segmentation of the macula by optical coherence tomography,” Opt. Express 17(18), 15659–15669 (2009).

42. R. Koprowski and Z. Wrobel, Layers Recognition in Tomographic Eye Image Based on Random Contour Analysis (Springer Berlin Heidelberg, 2009), pp. 471–478.

43. S. Lu, C. Y. Cheung, J. Liu, J. H. Lim, C. K. Leung, and T. Y. Wong, “Automated layer segmentation of optical coherence tomography images,” IEEE Trans. Biomed. Eng. 57(10), 2605–2608 (2010).

44. C. Ahlers, C. Simader, W. Geitzenauer, G. Stock, P. Stetson, S. Dastmalchi, and U. Schmidt-Erfurth, “Automatic segmentation in three-dimensional analysis of fibrovascular pigmentepithelial detachment using high-definition optical coherence tomography,” Br. J. Ophthalmol. 92(2), 197–203 (2008).

45. O. Tan, G. Li, A. T.-H. Lu, R. Varma, D. Huang, and A. I. for Glaucoma Study Group, “Mapping of macular substructures with optical coherence tomography for glaucoma diagnosis,” Ophthalmology 115(6), 949–956 (2008).

46. V. Kajic, B. Považay, B. Hermann, B. Hofer, D. Marshall, P. L. Rosin, and W. Drexler, “Robust segmentation of intraretinal layers in the normal human fovea using a novel statistical model based on texture and shape analysis,” Opt. Express 18(14), 14730–14744 (2010).

47. R. Kafieh, H. Rabbani, and S. Kermani, “A review of algorithms for segmentation of optical coherence tomography from retina,” J. Med. Signals Sens. 3(1), 45–60 (2013).

48. A. G. Roy, S. Conjeti, S. P. K. Karri, D. Sheet, C. Wachinger, and N. Navab, “ReLayNet: retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional networks,” Biomed. Opt. Express 8(8), 3627–3642 (2017).

49. J. I. Orlando, P. Seeböck, H. Bogunović, S. Klímscha, 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), pp. 1441–1445.

50. D. Ma, D. Lu, M. Heisler, S. Dabiri, S. Lee, G. W. Ding, M. V. Sarunic, and M. F. Beg, “Cascade dual-branch deep neural networks for retinal layer and fluid segmentation of optical coherence tomography incorporating relative positional map,” (PMLR, Montreal, QC, Canada, 2020), pp. 493–502.

51. S. Borkovkina, A. Camino, W. Janpongsri, M. V. Sarunic, and Y. Jian, “Real-time retinal layer segmentation of OCT volumes with GPU accelerated inferencing using a compressed, low-latency neural network,” Biomed. Opt. Express 11(7), 3968–3984 (2020).

52. A. Shah, L. Zhou, M. D. Abràmoff, and X. Wu, “Multiple surface segmentation using convolution neural nets: application to retinal layer segmentation in OCT images,” Biomed. Opt. Express 9(9), 4509–4526 (2018).

53. A. Tran, J. Weiss, S. Albarqouni, S. Faghi Roohi, and N. Navab, “Retinal layer segmentation reformulated as OCT language processing,” in Medical Image Computing and Computer Assisted Intervention – MICCAI 2020, A. L. Martel, P. Abolmaesumi, D. Stoyanov, D. Mateus, M. A. Zuluaga, S. K. Zhou, D. Racoceanu, and L. Joskowicz, eds. (Springer International Publishing, 2020), pp. 694–703.

54. B. Keller, M. Draelos, G. Tang, S. Farsiu, A. Kuo, K. Hauser, and J. A. Izatt, “Real-time corneal segmentation and 3D needle tracking in intrasurgical OCT,” Biomed. Opt. Express 9(6), 2716–2732 (2018).

55. K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), (2016), pp. 770–778.

56. J. Howard and S. Gugger, “Fastai: a layered API for deep learning,” Information 11(2), 108 (2020).

57. I. Loschchlov and F. Hutter, “Decoupled weight decay regularization,” in International Conference on Learning Representations, (2019).

58. M. Bertalmio, A. L. Bertozzi, and G. Sapiro, “Navier-stokes, fluid dynamics, and image and video inpainting,” in Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, vol. 1 (2001), p. I.

59. T. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, “Focal loss for dense object detection,” in 2017 IEEE International Conference on Computer Vision (ICCV), (2017), pp. 2999–3007.

60. C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. J. Cardoso, “Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations,” in Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, M. J. Cardoso, T. Arbel, G. Carneiro, T. Syeda-Mahmood, J. M. R. S. Tavares, M. Moradi, A. Bradley, H. Greenspan, J. P. Papa, A. Madabhushi, J. C. Nascimento, J. S. Cardoso, V. Bertalmio, A. L. Bertozzi, and G. Sapiro, eds. (Springer International Publishing, 2017), pp. 240–248.

61. 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, N. Navab, J. Hornegger, W. M. Wells, and A. Frangi, eds. (Springer International Publishing, 2015), pp. 234–241.

62. E. Romera, J. M. Álvarez, L. M. Bergasa, and R. Arroyo, “ERFNet: efficient residual factorized convnet for real-time semantic segmentation,” IEEE Trans. on Intell. Transp. Syst. 19(1), 263–272 (2018).

63. C. Vielkind, A.-H. Dhalli, J. Li, M. Jackson-Atogi, L. Vajzovic, A. Kuo, C. Toth, and J. Izatt, “Real time volumetric intrasurgical optical coherence tomography with 4D visualization of surgical maneuvers (conference presentation),” (2020), p. 19.

64. S. Wei, S. Guo, and J. U. Kang, “Analysis and evaluation of BC-mode OCT image visualization for microsurgery guidance,” Biomed. Opt. Express 10(10), 5268–5290 (2019).

65. J. P. Ehlers, A. Uchida, and S. K. Srivastava, “Intraoperative optical coherence tomography-compatible surgical instruments for real-time image-guided ophthalmic surgery,” Br. J. Ophthalmol. 101(10), 1306–1308 (2017).