

# Semi-Automated Extraction of Lens Fragments Via a Surgical Robot Using Semantic Segmentation of OCT Images With Deep Learning - Experimental Results in *Ex Vivo* Animal Model

Changyeob Shin <sup>1</sup>, Matthew J. Gerber <sup>1</sup>, Yu-Hsiu Lee <sup>1</sup>, Mercedes Rodriguez, Sahba Aghajani Pedram <sup>1</sup>, Jean-Pierre Hubschman, Tsu-Chin Tsao <sup>1</sup>, and Jacob Rosen

**Abstract**—The overarching goal of this letter is to demonstrate the feasibility of using optical coherence tomography (OCT) to guide a robotic system to extract lens fragments from *ex vivo* pig eyes. A convolutional neural network (CNN) was developed to semantically segment four intraocular structures (lens material, capsule, cornea, and iris) from OCT images. The neural network was trained on images from ten pig eyes, validated on images from eight different eyes, and tested on images from another ten eyes. This segmentation algorithm was incorporated into the Intraocular Robotic Interventional Surgical System (IRISS) to realize semi-automated detection and extraction of lens material. To demonstrate the system, the semi-automated detection and extraction task was performed on seven separate *ex vivo* pig eyes. The developed neural network exhibited 78.20% for the validation set and 83.89% for the test set in mean intersection over union metrics. Successful implementation and efficacy of the developed method were confirmed by comparing the preoperative and postoperative OCT volume scans from the seven experiments.

**Index Terms**—Cataract surgery, computer vision for medical robotics, deep learning, medical robots and systems, surgical robotics: planning.

## I. INTRODUCTION

**C**ATARACTS are the progressive clouding of the natural lens of the eye and represent the leading cause of blindness

Manuscript received November 29, 2020; accepted March 25, 2021. Date of publication April 12, 2021; date of current version April 29, 2021. This letter was recommended for publication by Associate Editor L. Fichera and Editor P. Valdastrì upon evaluation of the reviewer's comments. This work was supported in part by funds from the UCLA Stein Eye Institute; The Hess Foundation, New York, NY, USA; The Earl and Doris Peterson Fund, Los Angeles, CA, USA; an unrestricted institutional grant from Research to Prevent Blindness (RPB), New York, NY, USA; unrestricted gifts that support Tsao's research program from various donors; the National Institutes of Health under Grant No. R01 EY030595-01; and the Public Health Services under Grant T32-EY7026-43. (Corresponding author: Changyeob Shin.)

Changyeob Shin, Yu-Hsiu Lee, Sahba Aghajani Pedram, Tsu-Chin Tsao, and Jacob Rosen are with the Department of Mechanical and Aerospace Engineering, University of California, Los Angeles, CA 90025 USA (e-mail: shinhujune@ucla.edu; yuhsiulee@ucla.edu; sahbaap@ucla.edu; ttsao@ucla.edu; jacobrosen@ucla.edu).

Matthew J. Gerber, Mercedes Rodriguez, and Jean-Pierre Hubschman are with the Stein Eye Institute, University of California, Los Angeles, CA 90025 USA (e-mail: gerber211@ucla.edu; mercedesrodriguez@mednet.ucla.edu; hubschman@jsei.ucla.edu).

Digital Object Identifier 10.1109/LRA.2021.3072574

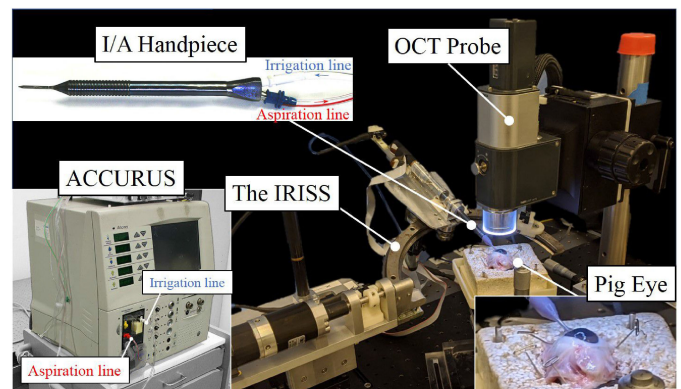


Fig. 1. Experiment setup with the IRISS and OCT system.

and visual impairment in the world [1]. Cataracts can be treated by removal of the opaque lens through cataract surgery, which is the most frequently performed surgical procedure in the United States, totaling approximately three million operations per year [2]. In cataract surgery, the opaque lens is extracted and replaced with an intraocular lens implant through several surgical steps including corneal incision, capsulorhexis, nucleus removal, cortical material removal, capsular bag polishing, and implant injection.

While technologies such as femtosecond laser systems can improve the eye-preparation steps, lens extraction—the most delicate and dangerous step—continues to be manually performed. Safe, effective lens removal is challenged by the physiological limitations of a human surgeon including hand tremor [3] and limited resolution of depth sensing [4]. In particular, the posterior capsule (PC) is a delicate and thin (approximately 4–9  $\mu\text{m}$ ) membrane which is optically translucent and difficult to visualize [5]. A surgeon is liable to misinterpret shadows and other indirect visual indications of the surgical instrument position, thereby increasing risk of PC rupture, one of the most common complications of cataract surgery [6].

However, the motion and stability requirements of cataract surgery are not prohibitive to the application of robotic surgical systems. The incorporation of robotic systems has recently

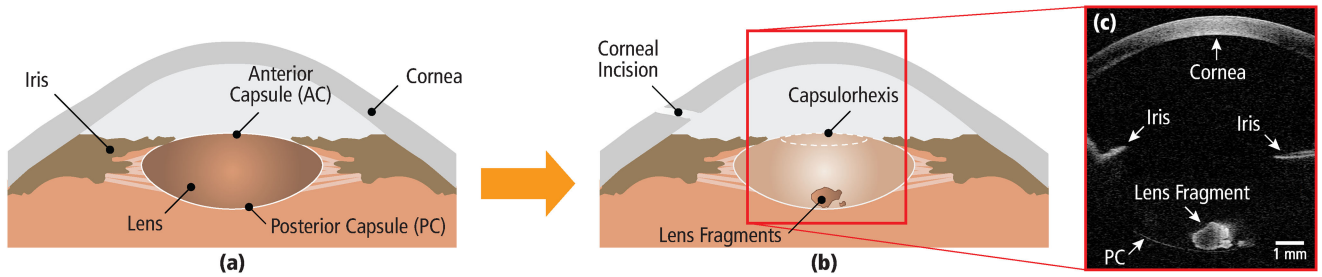


Fig. 2. Relevant eye anatomy and the corresponding OCT B-scan. (a) the normal eye before any procedure (b) the eye status during the lens removal step (c) OCT B-scan corresponding to (b). It was generated by merging two B-scans in different depth of scan for the tutorial purpose.

found widespread use throughout many fields such as urology, gynecology, and general surgery with the development of systems such as the da Vinci Surgical System [7]. In the field of ophthalmology, several teleoperated robotic systems have been developed and tested on *in vivo* models including human patients. Examples include the Preceyes Surgical System [8] from Preceyes BV as well as the Mynutia intraocular surgical system [9]. Both systems have demonstrated the capability of performing a range of teleoperated vitreoretinal surgical procedures including membrane peeling, subretinal injection, and retinal vein cannulation. However, intraocular robotic systems which have focused on performing procedures specific to cataract surgery are rare, and none have been demonstrated in an automated fashion. In contrast to vitreoretinal procedures, cataract extraction presents a less-structured and more dynamic workspace presenting unique challenges for an automated robotic system. Furthermore, the procedure is complicated by the difficulty of locating lens fragments due to ambient lighting and lack of depth sensing. The Intraocular Robotic Interventional Surgical System (IRISS) developed at UCLA is one system which has been used to demonstrate semi-automated lens extraction on *ex vivo* pig eyes using optical coherence tomography (OCT) as visual feedback to guide the robotic system [10].

To overcome the limited sensing capability of a surgeon during cataract surgery, OCT has been incorporated into the IRISS to localize the tool and the surrounding anatomy [10]. Through OCT-based visualization (Fig. 2), the tool position relative to intraocular anatomical structures can be understood during surgical procedures and maneuvers can be more safely executed. In [10], an automatic tool insertion method and a trajectory generation algorithm were developed using the parametric model of the eye and fitting the model to OCT B-scan and volume scan data. Although this study introduced automation into important steps of the lens extraction, specifically nucleus, the remaining lens fragments at the end of the procedure were manually localized in the camera and OCT frames in order to remove them. The data acquired by OCT suffers from a low signal-to-noise ratio and is corrupted by granular interference inherent to the acquisition process (commonly referred to as speckle noise). Conventional image-processing techniques suffer from the presence of speckle noise, and while methods to reduce the speckle noise have been investigated (e.g., [11]), the challenge to extract useful information from OCT data remains.

Recent advancements in deep learning have demonstrated success in computer-vision problems. Among deep-learning architectures, convolutional neural networks (CNNs), motivated by how the brain processes visual information, have shown promising performance in many image classification problems [12]. CNNs extract visual features from a given dataset using multiple channels and layers of convolution layers. As the structure of a CNN gets deeper, the higher level visual features are learned. These visual features representing the dataset are used to classify an image into an object or to generate a pixel-level segmentation map.

Deep learning has been employed in a few aspects of eye surgery for segmentation of OCT images. In [13], a fully convolutional neural network was used to localize the cornea and needle in OCT images. The segmentation algorithm was proposed to be used in deep anterior lamellar keratoplasty surgery and porcine eyes were used to train and validate the performance. Corneal interface segmentation network (Cornet) was introduced to segment three corneal interfaces for anterior segment interventions [14]. A method which uses OCT volumetric images and a CNN for localizing a tool under the retina was presented in [15]. Other works on segmentation of OCT images mostly focused on retinal layers. Retinal layer segmentation network (ReLayNet) was developed to segment retinal layers and fluid for monitoring the degradation of vision quality caused by diabetics [16]. Work on segmenting Bruch's membrane and choroid layer in OCT images to generate the thickness map was presented [17]. Although these works addressed segmentation problems of OCT images for eye, they are not applicable for the lens extraction task because the area of the eye where OCT scanned in those works is not appropriate.

For application to cataract surgery, existing work in image segmentation has been limited to tracking of a surgical instrument in camera images using CNN [18] and an attention-based neural network [19]. However, to achieve safe automated removal of lens material during cataract surgery, the ability to localize relevant intraocular anatomy in real-time will be required. For this reason, we present a method to localize intraocular anatomy in OCT images. Furthermore, a framework to incorporate the developed OCT-segmentation algorithm into the intraocular robotic system is presented. The efficacy of the framework was demonstrated by experiments with *ex vivo* pig eyes. To the best of our knowledge, this is the first demonstration of automated segmentation of OCT images for the removal of

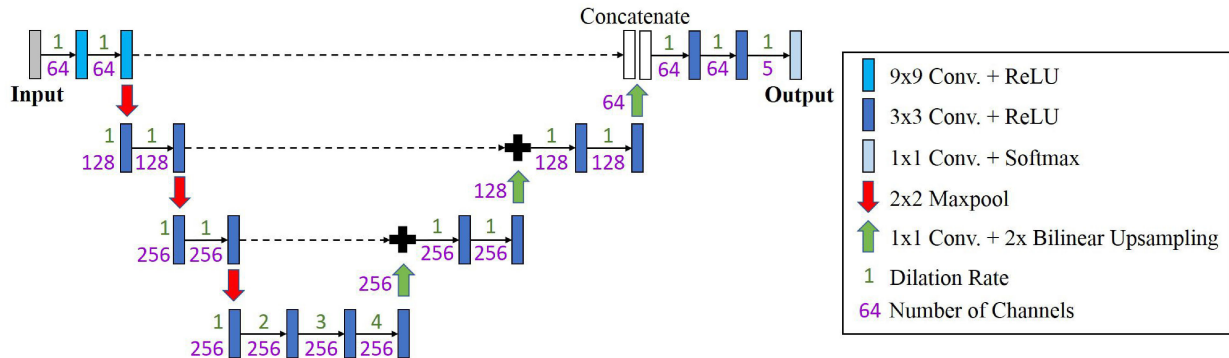


Fig. 3. Shown is the structure of the convolutional neural network used in this work.

lens material and the first to apply this to guidance of a robotic surgical system.

The main contributions of this paper are summarized as follows:

#### Contributions:

- Development and verification of a deep-learning framework to segment intraocular anatomy (lens material, capsule, cornea, and iris) in OCT images.
- Development of a framework for semi-automated robotic extraction of lens fragments from pig eyes.
- Demonstration and evaluation of the integrated solution on seven *ex vivo* pig eyes.

## II. MATERIALS AND METHODS

### A. Pig Eye Preparation

As the eye model, *ex vivo* pig eyes were used (Sioux-Preme Packing, Sioux City, Iowa, USA). The unscalded, enucleated eyes were shipped on ice overnight from pigs butchered the previous day. The eyes were secured by pinning their excess skin into a custom polystyrene holder. Preparation of each eye was performed under a surgical microscope (M840, Leica Microsystems, GmbH). A temporal corneal multiplanar incision was created with a 2.8 mm keratome blade to ensure a watertight wound. Sterile lubricating jelly (MDS032290H, Medline) was injected into the anterior chamber to protect the corneal endothelium. The jelly has proven to be a good alternative to the expensive ophthalmic viscoelastic gel and exhibits similar optical and material properties. A cystotome was used to create a central linear cut in the anterior capsule and then pushed to generate a flap, which was manipulated with forceps to create a 6–8 mm diameter continuous curvilinear capsulorhexis. Some of the jelly was removed from the anterior chamber in order to proceed with the hydrodissection. Balanced saline solution was then injected with a cannula attached to a syringe between the outer part of the lens and the capsular bag to achieve their separation. Lens removal was accomplished by slowly aspirating the lens using an I/A handpiece. Lens fragments of various sizes (Section II-B5 and Table II) were intentionally left in the capsular bag and pushed onto the PC with sterile jelly, which also helped provide a smoother concave shape to the PC.

### B. Deep Learning-Based Segmentation

1) *Data Labelling*: In this study, we approached the problem of localizing lens material as a semantic segmentation problem. Specifically, given an OCT image  $\mathcal{I}$ , we wished to find a function  $\mathcal{F} : \mathcal{I} \rightarrow \mathcal{L}$  that mapped every pixel in  $\mathcal{I}$  to a label  $\mathcal{L} \in \{1, \dots, n_c\}$ , where  $n_c$  is the number of classes. The segmentation task was a  $n_c = 5$  class-classification problem where the classes were the (1) lens, (2) capsule, (3) cornea, (4) iris, and (5) background. The I/A handpiece was included in the lens class because during cortical-material clean-up, the I/A handpiece is commonly occluded by the lens material. In addition, the location of the I/A handpiece can be identified using the forward kinematics of the robot and the known robot-to-OCT registration and therefore can be differentiated from the lens material if necessary.

For the training and validation of the algorithm, the OCT-acquired data was manually labeled (Fig. 4). Specifically, the cornea appears as a thick, transparent curve along the anterior segment; the iris appears as hyper-reflective regions on either or both sides of the scan; the lens material appears as amorphous forms within the capsular bag; and the PC appears as a thin, reflective curve. The PC location is important to know to reduce the likelihood of PC rupture—a serious surgical complication. While the cornea and iris are relatively static throughout the procedure, the lens material and flexible PC will change location; their localization is essential for performing safe, effective robotic cataract surgery.

2) *Deep Neural Network*: The shape and location of intraocular tissue in the acquired data can vary significantly between frames due to fluid turbulence force and deformation of tissue. These challenges decreased our confidence that a model-based approach could produce accurate results. Instead, a deep-learning approach for segmenting the OCT images was applied. Among the various types of deep convolution neural network structures, a fully convolutional neural network was exploited to provide pixel-level segmentation (Fig. 3). Its design was inspired by the U-net and FCN-8 structures [20], [21], with an encoder part to extract high-level features and a decoder part to recover the feature channel dimensions to the original input size.

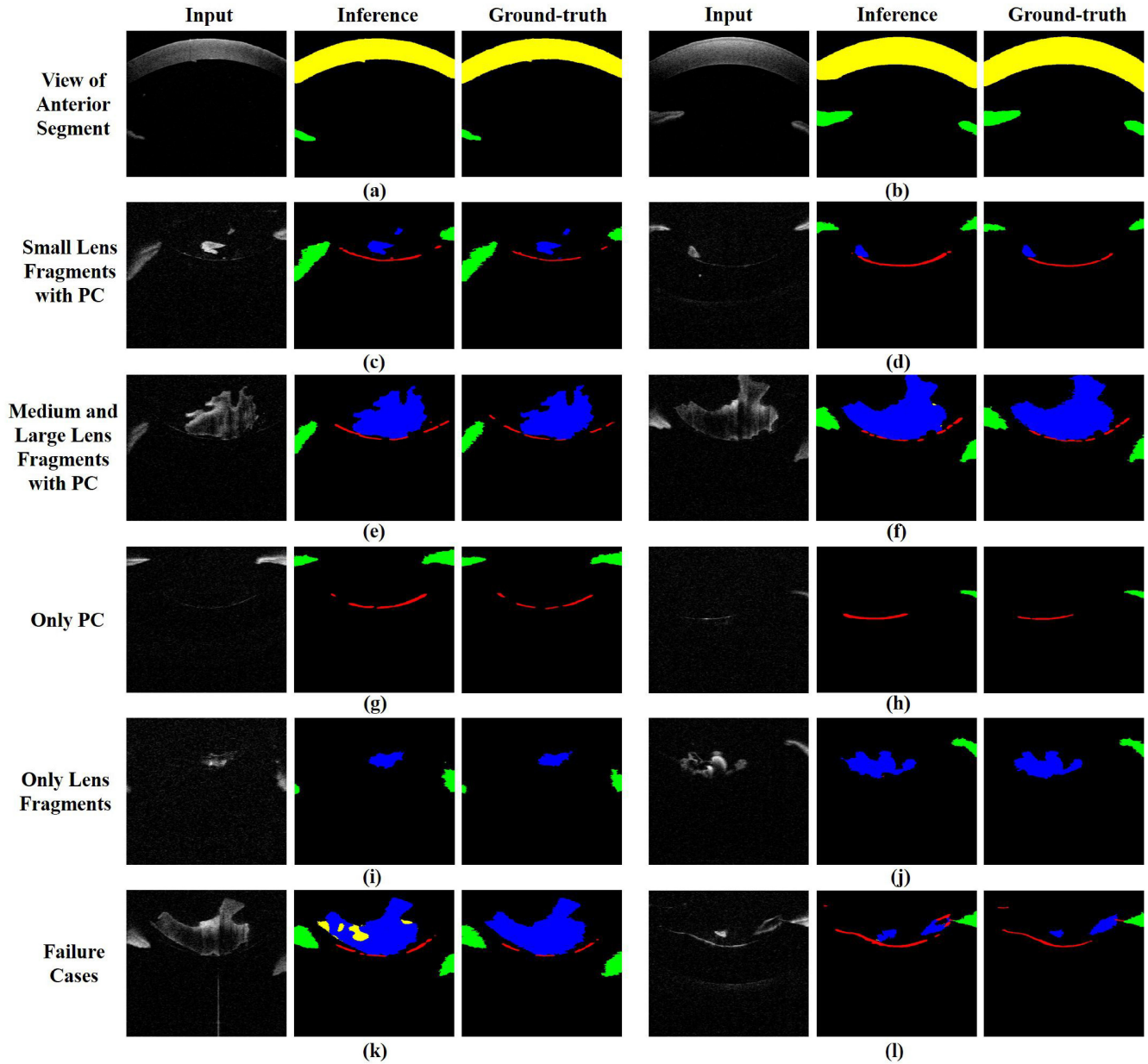


Fig. 4. Illustration of OCT image segmentation results. Each set includes input (left), inference result from DNN (middle), and ground truth (right). Images in (f) and (k) are from the validation set; all other images are from the test set. In the inference and ground-truth images, the segmented intraocular anatomical structures are colored yellow (cornea), green (iris), blue (lens), and red (posterior capsule). (a) anterior view with cornea and iris, other side of iris is not captured, (b) anterior view, (c,d) small lens fragments with posterior capsule, (e) medium lens fragment with posterior capsule, (f) large fragment of lens with posterior capsule, (g,h) posterior capsule without lens, (i,j) only lens, posterior capsule is not visible, (k) failure case I, a portion of a large lens fragment is classified as cornea, (l) failure case II, inside of the lens fragment is not scanned and the part of outline of the lens is classified as the capsule.

In each level of abstractions, there is a bridge which links the features map to a decoded channel to incorporate different levels of information into the segmentation. We also used a dilated convolution in the highest level of features to enlarge the receptive field of the filters [22]. The dilated rate was increased by one in each feature-extraction level. The filter size of each convolution layer was  $3 \times 3$  except for the first two convolutions. The batch normalization was followed with a rectified linear unit (ReLU) activation layer for each convolution layer except the last  $1 \times 1$  layer. The first two convolutions have  $9 \times 9$  kernels to extract denser features with a large receptive field. Upsampling

in the decoder part consisted of  $1 \times 1$  convolutions to match the channel numbers in the following layers and the dimension was upsampled with bilinear interpolation.

3) *Training*: Our dataset is split into training data (809 images) and validation data (111 images). The OCT images for training were collected from ten eyes and the images for the validation were from a different set of eight eyes.

The developed network was trained using the dice coefficient loss (DCL) and the focal loss (FL). These loss functions were selected to help address the class imbalance between the size of each intraocular structure (Fig. 4). In particular, the capsule

typically occupies a small portion in the dataset. DCL and FL are defined as:

$$\text{DCL} = 1 - \frac{1}{n_c} \sum_{c=1}^{n_c} \frac{2|P_c \circ G_c|}{2|P_c \circ G_c| + |P'_c \circ G_c| + |P_c \circ G'_c|} \quad (1)$$

$$\text{FL} = - \sum_{c=1}^{n_c} \alpha |(1 - P_c)^\gamma \circ G_c \circ \log(P_c)| \quad (2)$$

where  $P_c$  is a prediction map from the neural network and  $G_c$  is ground-truth for channel  $c$ . Both are  $\mathbb{R}^{H \times W}$  when the dimension of the input image is  $\in \mathbb{R}^{H \times W}$ . The symbol  $\circ$  represents an element-wise matrix multiplication (the Hadamard product) and  $|\cdot| : \mathbb{R}^{H \times W} \rightarrow \mathbb{R}$  is the summation of all elements.  $P'_c$  is  $1 - P_c$  and  $G'_c$  is  $1 - G_c$  where  $\mathbf{1} \in \mathbb{R}^{H \times W}$  is a matrix of ones. The DCL minimizes the ratio between the intersection and union of the prediction and the ground truth. The summation term is unity when the predictions perfectly match the ground truth and is subtracted from one to make the minimum value of the DCL zero. The exponential and log operations are element-wise. The FL adds a factor  $(1 - P_c)^\gamma$  to the cross-entropy loss, to address problems of data imbalance [23]. The added term reduces the loss for well-classified classes.

Dimension of the input image was reduced in two steps. First, the top-most 29 rows of data are removed from the image due to the presence of acquisition artifacts. Second, the data is sampled by half by removing every other column of data. These steps reduce the dimension of the input image from  $1024 \times 400$  to  $498 \times 200$  px.

An Adam optimizer [24] with a learning rate of  $1 \times 10^{-4}$  was used to train the network. The learning rate was reduced by half if the validation accuracy did not increase for three consecutive epochs. The training was stopped if the performance of the network did not improve for ten epochs. For the focal loss,  $\gamma \in \{0, 0.5, 1.0, 2.0, 3.0, 4.0, 5.0, 10.0\}$  and  $\alpha = 0.2$  was tested.

The training data was augmented by horizontal reflection, vertical translation, and variation in intensity values using multiplication of scale factors sampled uniformly between 0.5 and 1. To maintain the geometrical meaning of the intraocular structures, vertical reflection was not used.

4) *Evaluation*: The validation and test set consisted of images across a range of lens-fragment sizes and scan depths (anterior and posterior views). The performance of the trained neural network was evaluated with two metrics: accuracy and inference time. Intersection over Union (IoU) was used to measure accuracy, defined as:

$$\text{Mean IoU} = \frac{1}{n_c} \sum_{i=1}^{n_c} \frac{\mathcal{C}_{ii}}{\mathcal{G}_i + \mathcal{P}_i - \mathcal{C}_{ii}} \quad (3)$$

where  $\mathcal{C}_{ii}$  is the number of pixels whose ground truth is labeled as class  $i$  with inference result  $i$ ,  $\mathcal{G}_i$  is the number of pixels labeled as the class  $i$  in the ground truth, and  $\mathcal{P}_i$  is the number of pixels predicted as class  $i$ . Inference time was an important metric because it determined the feasibility of incorporating the framework into a robotic system as feedback. Inference time was calculated as the mean time to obtain an inference

TABLE I  
EVALUATION METRICS OF NEURAL NETWORKS

Dataset	Validation set		Test set
Loss functions	DCL	Focal	DCL
Mean IoU (%)	78.20	75.42	83.89
Lens IoU (%)	75.03	72.71	81.18
Capsule IoU (%)	46.63	36.66	59.81
Inference time	31.28 ms		

probability map of 1000 images from the network. The network was implemented in Keras/Tensorflow. All experiments were performed on an Nvidia Geforce GTX 1080 Ti with 11 GB of memory.

5) *Segmentation Accuracy*: Among the trained models, the ones which exhibited the highest accuracy on the validation set were selected (Table I). For the focal loss, the model that was trained with  $\gamma = 4.0$  exhibited the highest accuracy. The model trained with the DCL exhibited higher accuracy than the model trained with the FL. This difference was most pronounced in the accuracy of capsule detection (46.63% cf. 36.66%). Based on this result, we opted to use the model trained with DCL. The performance of the CNN was evaluated on a separate test set consisting of 200 images from an additional ten pig eyes (Table I). In the test set, seven pig eyes were from the eyes used in the robot experiment.

Figure 4 illustrates examples of inference results from the developed network. Shown are 12 sets of images consisting of input image (OCT B-scan), the inference result, and the ground truth. The images represent cases with different sizes of lens fragments and scan depths. In the validation and test data, the total area of the lens fragment varies from 139 px ( $0.13 \text{ mm}^2$ ) to 20,884 px ( $19.17 \text{ mm}^2$ ). Areas less than 7000 pixels ( $6.43 \text{ mm}^2$ ) were defined as “small,” areas greater than 14,000 pixels ( $12.86 \text{ mm}^2$ ) were defined as “large,” and areas between these two values were defined as “medium.”

6) *Phase Ambiguity in the OCT System*: Like many OCT systems, the data acquired by the OCT device in this work suffers from phase ambiguity. The phase ambiguity results in image inversion of anatomical structures which are physically closer to the probe than the scanning depth (Fig. 5(a)). For the lens-extraction task, where the scanning depth is focused near the PC, the cornea and iris will appear inverted in B-scan images. This inversion introduces difficulty for correct labeling, especially when the inverted structures appear deeper in the scan where the signal-to-noise ratio is lower. To address this problem, we defined an area mask as shown in Fig. 5. This masked area was not considered in the accuracy tests. We use the knowledge that the cornea is at least several millimeters anterior to the capsular bag and therefore can be safely ignored.

During the operation of the supervised lens extraction, the scan depth can be adjusted by physically moving the OCT probe closer or further from the eye. Doing so shifts where the inverted cornea and iris will appear in the B-scan image relative to the PC and lens material. With larger pig eyes, the depth could be adjusted such that these structures were not visible.

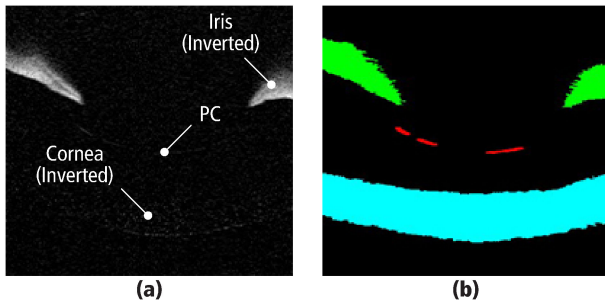


Fig. 5. (a) A B-scan containing non-inverted PC and inverted iris and cornea. The inverted cornea complicates the learning algorithm and is ignored. (b) Pixel labels for the three anatomical structures in (a). Green: iris, red: PC, and cyan: cornea (ignored area).

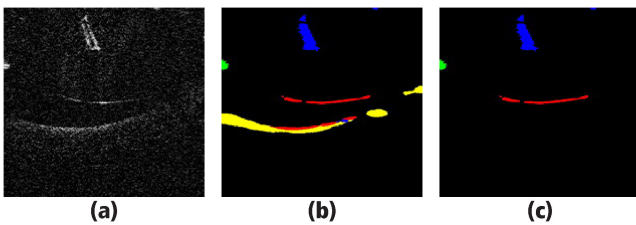


Fig. 6. (a) B-scan data showing a fragment of lens, the PC, and the inverted cornea. Note the inverted cornea appears similar to the PC. (b) Segmentation result without compensation for the inverted cornea. Note the incorrect labels. (c) Segmentation results after cornea removal. Color labels are blue: lens, green: iris, red: PC, and yellow: cornea.

However, in the case of smaller eyes, this may not be an option. This problem could be solved in two ways. First, the operator can select the ignored area in the image and the segmentation algorithm neglects the chosen area. Second, the location of the inverted cornea is found using the pixel information classified as the cornea. Then, the ignored region could be defined using the predefined two convex shape kernels and user-defined cornea thickness. This is possible because the cornea shape is convex when inverted and does not significantly deform. An example of this case is shown in Fig. 6. It is seen that the outline of the cornea is classified as the PC due to the majority of the cornea not being captured because of low signal-to-noise ratio. However, using the cornea information, the inverted cornea can be successfully removed using the post-processing algorithm.

### C. OCT-Guided Robot Control

To extract a lens fragment, the segmentation algorithm was incorporated into the control of the robotic system as follows. An OCT B-scan was first acquired of the eye and then segmented using the CNN. Next, the segmented B-scan data was processed to identify the centroid of the largest binary blob of a lens fragment. Through an offline registration of Robot-to-OCT coordinates presented in [25], the location of any point detected in the OCT coordinate frame can be known in the robot coordinate frame. This registration algorithm uses a volume scan of the surgical tool, a shape-fitting algorithm, and robot kinematics to find the spatial relation between the frames. The robot was then

commanded to move towards the lens fragment with a prescribed 1 mm/s tool-tip speed by following a joint-space trajectory from its initial position to the detected centroid location. The software architecture allowed for motion abort and rerouting of the path of the I/A handpiece to help account for the dynamic nature of the process.

## III. EXPERIMENT AND RESULTS

To assess the developed framework for semi-automated extraction of lens fragments, we performed lens extraction experiments on seven *ex vivo* pig eyes.

### A. Robotic System

The robotic system used in this work was the IRISS. The IRISS has been used to perform a range of teleoperated intraocular surgical procedures on *ex vivo* pig eyes [26] as well as demonstrate partially automated lens removal on *ex vivo* pig eyes with OCT feedback [10]. Detailed description of the IRISS mechanism and kinematics are provided in previous work [10], [26]. The intraocular pressure is maintained through the use of an Alcon I/A handpiece, which is connected to an Alcon Accurus vitrectomy system and controlled via the robotic system. The IRISS sends pulse-width signals to the Accurus to mimic standard foot pedal commands and the system regulates the irrigation pressure as a function of the aspiration force.

### B. OCT System

The OCT system used in this work (Telesto II-1060LR, Thorlabs) is capable of acquiring two-dimensional, cross-sectional images (B-scans) and three-dimensional volume scans. The axial resolution of the system was  $9.18 \mu\text{m}/\text{px}$  and lateral resolution of  $25 \mu\text{m}/\text{px}$ . B-scans were acquired at a width of 10 mm and depth of 9.4 mm while the volume scans were acquired with a volume of  $10 \times 10 \times 9.4 \text{mm}^3$ . The automated lens-extraction portion of this work relied on the B-scan data as feedback while the volume scans were used only for evaluation purposes. For this work, the IRISS was mounted with a straight-tip, side port, I/A handpiece (8172 UltraFLOW, Alcon). The I/A handpiece was registered to the robotic workspace according to the calibration process developed in previous work [25], but other tool-localization methods could be used (e.g., [27]).

### C. Protocol

To begin a trial, a pig eye was manually prepared (Section II-A), fixed to a Styrofoam holder, and placed within the physical workspace of the robotic system (Fig. 1). Next, the IRISS was teleoperated to align the tip of the I/A handpiece to the corneal incision, the tool was inserted approximately 1–2 mm into the eye (past the corneal endothelium), and the irrigation pressure set to 60 mmHg to maintain intraocular pressure. At this point, the operator acquired a single B-scan image of lens material by shifting the OCT scanning plane through the eye based on the camera image. This scheme was used to avoid the I/A handpiece occluding the lens material.

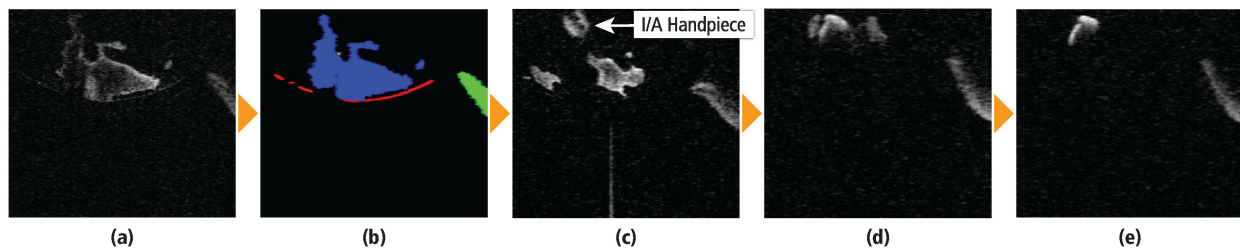


Fig. 7. OCT B-scans of an exemplary robotic experiment. (a) the selected scan by the operator, (b) segmentation result of (a), (c) the tool is placed on the cortical material, (d) the cortical material is gradually removed by the tool (e) the lens material has been removed.

TABLE II  
EVALUATION METRICS OF LENS EXTRACTION

Eye Number	1	2	3	4	5	6	7
Lens Volume [mm <sup>3</sup> ]	24.18	7.66	8.64	8.78	13.78	41.16	2.90
Time to Aspirate [s]	48.09	2.90	101.42	44.06	290.77	397.18	45.96

The selected B-scan was sent to the segmentation algorithm and the surgical robot was controlled using the method explained in Section II-C. Once the I/A handpiece was positioned at the lens fragment, the aspiration force was stepped up to 200 mmHg and the lens fragment aspirated (Table II). The aspiration was stopped once the operator deemed the fragment had been removed based on feedback from the continuously acquired B-scan and camera images (Fig. 7). The sequences in Fig. 7 demonstrate the I/A handpiece, which appears as an ellipse sectioned by the B-scan plane, was moved to the targeted lens fragment and then successfully aspirated it. After the lens fragment extraction, complete removal was confirmed by a trained fellow.

For post-trial evaluation, OCT volume scans were acquired before and after the lens-extraction operation, outside the scope of the automated procedures (Fig. 8). The volume of the fragment of lens material (Table II) was calculated by manual segmentation of the OCT volume scan: the number of lens-material voxels were summed and then multiplied by the known voxel volume (Section III-B).

#### IV. DISCUSSION AND CONCLUSION

The proposed model successfully segmented the target anatomical structures across different scenarios, including depth of scan (Fig. 4 a–j). Furthermore, the success of the developed method suggests its ability to handle the speckle noise that is common in OCT images.

However, there are at least two failure cases which should be addressed in future work. First, the lens fragment can sometimes appear similar (shape and intensity) to the inverted cornea (Fig. 4 k). This misclassification can occur because OCT data is grayscale and the lens material can present as any arbitrary form (including a cornea-like curve). Second, the internal intensity of lens fragments varies for unknown reasons, sometimes causing the fragment to appear solid and at other times hollow (Fig. 4 l). In the hollow appearance, the edge of the lens fragment may

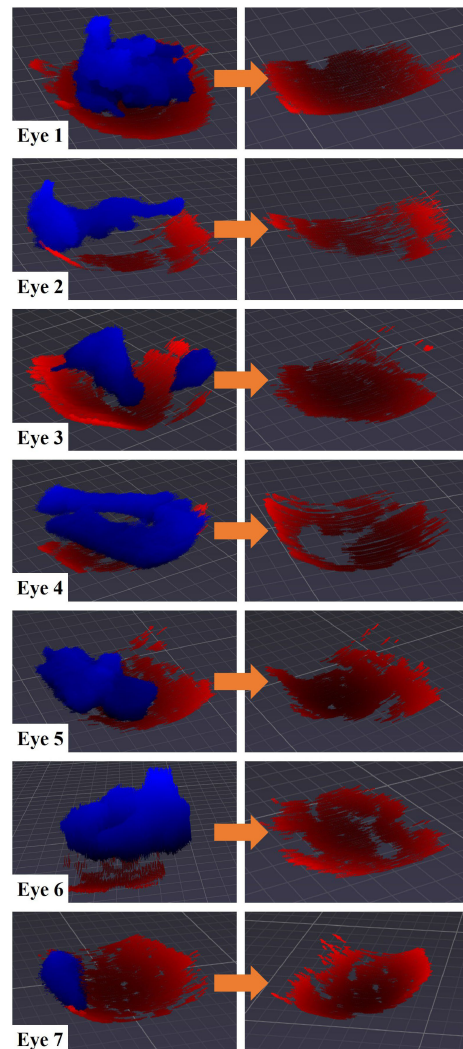


Fig. 8. Preoperative and postoperative OCT volume scans that illustrate the successful removal of the lens material from each pig eye. Lens and capsule are colored blue and red, respectively. The square grid has an edge length of  $90.6 \mu\text{m}$ .

appear as a thin, hyper-reflective line, which the segmentation algorithm oftentimes mislabels as capsule. Three explanations for this inconsistency in appearance are (1) the ratio of jelly to water during eye preparation affects the lens-fragment intensity since water attenuates OCT signal to a greater extent

than the jelly, (2) the lens fragment may be nucleus material rather than cortical material, and (3) the overall size of the lens fragment may affect its internal intensity. Future experimentation with the eye preparation is expected to reveal the reasoning behind the inconsistency and suggest potential mitigation strategies.

It was observed that the time to aspirate the cortical material varied between each operation (Table II), but no clear correlation was found between the volume of the lens material and the time required to aspirate. However, it was observed that the majority of the reported time was spent aspirating the fragment as it occluded the vacuum port of the I/A handpiece, and this observation offers some explanation: the tool-tip would often-times become clogged and would require a long time to clear itself. Therefore, with a tool better suited for removal of the soft pig-eye lens material (e.g., phacoemulsification probe), it is expected that the time to aspirate would correlate with the volume of lens material.

In this work, an integrated framework was developed for semi-automated detection and extraction of lens fragments in *ex vivo* pig eyes. It was shown that OCT images, which suffer from low signal-to-noise ratio, can be automatically segmented to localize intraocular structures using the developed CNN. Furthermore, with the successful experimental results on seven pig eyes, it was demonstrated that segmentation results from the CNN can be used to guide a surgical robot to extract lens fragments. To enable the full automation of lens extraction using OCT feedback, future work from this study includes improving the segmentation algorithm to cope with the failure cases observed in the experiment. Furthermore, a path-planning algorithm of the surgical robot that exploits the dynamics of intraocular structures during the lens-extraction procedure should be developed to perform the task more effectively and avoid posterior capsule rupture. In reality, the lens material changes shape and location as a function of the aspiration and irrigation forces, and a means to account for these changes will be necessary using real-time OCT feedback.

#### ACKNOWLEDGMENT

The authors would like to thank Nicholas Iafe, MD for his surgical expertise and assisting with the physical experiments.

#### REFERENCES

- [1] S. Resnikoff *et al.*, "Global data on visual impairment in the year 2002," *Bull. world health org.*, vol. 82, pp. 844–851, 2004.
- [2] N. G. Congdon, D. S. Friedman, and T. Lietman, "Important causes of visual impairment in the world today," *Jama*, vol. 290, no. 15, pp. 2057–2060, 2003.
- [3] C. N. Riviere, R. S. Rader, and P. K. Khosla, "Characteristics of hand motion of eye surgeons," *Proc. 19th Ann. Int. Conf. IEEE Eng. Med. Biol. Soc. 'Magnificent Milestones and Emerg. Opportunities Med. Eng.'* (Cat. No 97CH36136), vol. 4, pp. 1690–1693, 1997.
- [4] P. B. Hibbard, A. E. Haines, and R. L. Hornsey, "Magnitude, precision, and realism of depth perception in stereoscopic vision," *Cogn. Res.: Princ. Implic.*, vol. 2, no. 1, p. 25, 2017.
- [5] S. Krag and T. T. Andreassen, "Mechanical properties of the human posterior lens capsule," *Invest. ophthalmol. Vis. Sci.*, vol. 44, no. 2, pp. 691–696, 2003.
- [6] M. Zare, M.-A. Javadi, B. Einollahi, A.-R. Baradaran-Rafii, S. Feizi, and V. Kiavash, "Risk factors for posterior capsule rupture and vitreous loss during phacoemulsification," *J. ophthalmic Vis. Res.*, vol. 4, no. 4, p. 208, 2009.
- [7] G. S. Guthart and J. K. Salisbury, "The intuitive/sup tnm/telesurgery System: Overview and application," *Proc. ICRA. Millennium Conf. IEEE Int. Conf. Robot. Automat. Symposia Proc. (Cat. No 00CH37065)*, vol. 1, pp. 618–621, 2000.
- [8] M. D. de Smet *et al.*, "Release of experimental retinal vein occlusions by direct intraluminal injection of ocriplasmin," *Brit. J. Ophthalmol.*, vol. 100, no. 12, pp. 1742–1746, 2016.
- [9] K. Willekens *et al.*, "Robot-assisted retinal vein cannulation in an in vivo porcine retinal vein occlusion model," *Acta Ophthalmol.*, vol. 95, no. 3, pp. 270–275, 2017.
- [10] C.-W. Chen *et al.*, "Intraocular robotic interventional surgical system (iriss): Semi-automated oct-guided cataract removal," *Int. J. Med. Robot. Comput. Assist. Surg.*, vol. 14, no. 6, 2018, Art. no. e 1949.
- [11] Y. Ma, X. Chen, W. Zhu, X. Cheng, D. Xiang, and F. Shi, "Speckle noise reduction in optical coherence tomography images based on edge-sensitive cgan," *Biomed. opt. exp.*, vol. 9, no. 11, pp. 5129–5146, 2018.
- [12] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [13] I. Park, H. K. Kim, W. K. Chung, and K. Kim, "Deep learning based real-time oct image segmentation and correction for robotic needle insertion systems," *IEEE Robot. Automat. Lett.*, vol. 5, no. 3, pp. 4517–4524, 2020.
- [14] T. S. Mathai, K. L. Lathrop, and J. Galeotti, "Learning to segment corneal tissue interfaces in oct images," in *16th Int. Symp. Biomed. Imag. (ISBI 2019)*, 2019, pp. 1432–1436.
- [15] M. Zhou *et al.*, "Needle localization for robot-assisted subretinal injection based on deep learning," in *Proc. Int. Conf. Robot. Automat. (ICRA)*, 2019, pp. 8727–8732.
- [16] A. G. Roy *et al.*, "Relaynet: Retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional networks," *Biomed. Opt. Exp.*, vol. 8, no. 8, pp. 3627–3642, 2017.
- [17] S. Masood *et al.*, "Automatic choroid layer segmentation from optical coherence tomography images using deep learning," *Sci. Rep.*, vol. 9, no. 1, pp. 1–18, 2019.
- [18] D. Zang, G.-B. Bian, Y. Wang, and Z. Li, "An extremely fast and precise convolutional neural network for recognition and localization of cataract surgical tools," in *Int. Conf. Med. Image Comput. Comput.-Assist. Interv.* Springer, 2019, pp. 56–64.
- [19] Z.-L. Ni *et al.*, "Raunet: Residual attention u-net for semantic segmentation of cataract surgical instruments," in *Proc. Int. Conf. Neural Inf. Process.* Springer, 2019, pp. 139–149.
- [20] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. image comput. comput.-assist. interv.* Springer, 2015, pp. 234–241.
- [21] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 3431–3440.
- [22] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," in *Proc. Int. Conf. Learn. Representations (ICLR)*, May 2016.
- [23] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2980–2988.
- [24] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Representations (ICLR)*, May 2014.
- [25] M. J. Gerber, J.-P. Hubschman, and T.-C. Tsao, "Robotic posterior capsule polishing by optical coherence tomography image guidance," *Int. J. Med. Robot. Comput. Assist. Surg.*, Art. no. eRCS 2248, 2021.
- [26] J. T. Wilson *et al.*, "Intraocular robotic interventional surgical system (iriss): Mechanical design, evaluation, and master-slave manipulation," in *Proc. Int. J. Med. Robot. Comput. Assist. Surg.*, vol. 14, no. 1, 2018, Art. no. e 1842.
- [27] M. Zhou *et al.*, "6DOF needle pose estimation for robot-assisted vitreoretinal surgery," *IEEE Access*, vol. 7, pp. 63 113–63 122, 2019.