2019 International Conference on Robotics and Automation (ICRA)
Palais des congres de Montreal, Montreal, Canada, May 20-24, 2019

# Needle Localization for Robot-assisted Subretinal Injection based on Deep Learning

Mingchuan Zhou[1], Xijia Wang[1], Jakob Weiss[2], Abouzar Eslami[3], Kai Huang[4], Mathias Maier[5], Chris P. Lohmann[5], Nassir Navab[2], Alois Knoll[1] and M. Ali Nasseri[5]

***Abstract*— Subretinal injection is known to be a complicated task for ophthalmologists to perform, the main sources of difficulties are the fine anatomy of the retina, insufficient visual feedback, and high surgical precision. Image guided robot-assisted surgery is one of the promising solutions that bring significant surgical enhancement in treatment outcome and reduces the physical limitations of human surgeons. In this paper, we demonstrate a robust framework for needle detection and localization in subretinal injection using microscope-integrated Optical Coherence Tomography (MI-OCT) based on deep learning. The proposed method consists of two main steps: a) the preprocessing of OCT volumetric images; b) needle localization in the processed images. The first step is to coarsely localize the needle position based on the needle information above the retinal surface and crop the original image into a small region of interest (ROI). Afterward, the cropped small image is fed into a well trained network for detection and localization of the needle segment. The entire framework is extensively validated in ex-vivo pig eye experiments with robotic subretinal injection. The results show that the proposed method can localize the needle accurately with a confidence of 99.2%.**

## I. Introduction

Subretinal injection is a typical vitreoretinal surgery with a delicate and complex workflow, which needs critical surgical skills and considerations. Despite its difficulties, the subretinal injection is an important ophthalmic procedure since it enables delivery of the drug directly into the retinal layers to provide more effective treatments [1]. As an example, Fig. 1 shows a subretinal injection of Tissue Plasminogen Activator (tPA) to dissolve the clotted blood under the retina. This injection is performed for a subretinal hemorrhage case. Subretinal hemorrhage results from various eye diseases such as Retinal Vessel Occlusion (RVO); Aged-related Macular Degeneration (AMD); and Diabetic Retinopathy (DR). Recently, subretinal injection has also been proposed to deliver the stem cell and gene cargo under the retina for curing AMD which is one of the leading causes of blindness in developed countries (15% incidence rate with people age over 65 years old).

The main challenges of manual subretinal injection are the low position ability of human hand and limited visual feedback from conventional microscope images. The reported hand tremor of an ophthalmic surgeon is 182 µm RMS in amplitude [2] while the acceptable precision for subretinal injection is 25 µm with 250 µm average thickness of the retina. Furthermore, the en-face view from conventional microscope is not able to provide position information of the needle tip under the translucent retina. Consequently, subretinal intervention is a demanding operation requiring a surgeon with excellent physical condition and abundant clinical experience, only a limited number of surgeons around the world fulfill these capabilities.

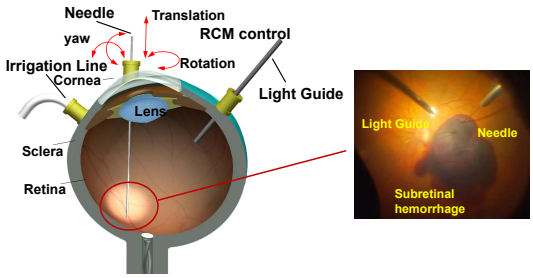



Fig. 1. The setup for subretinal injection in a subretinal hemorrhage case.

Recently, robot-assisted surgery (RAS) setups are known as the solution for reducing the work intensity, increasing the surgical outcomes and prolonging the service time of experienced surgeons in ophthalmic surgery [3]–[8]. Over the years, these setups are getting closer to clinical trials. In September 2016, surgeons at Oxford's John Radcliffe Hospital performed the world's first robotic eye surgery. The eye surgical robot named Robotic Retinal Dissection Device (R2D2) with 10 µm accuracy was used in the clinical trials for subretinal injection [1]. They used the microscope-integrated Optical Coherence Tomography (MI-OCT) (RESCAN 700, Carl Zeiss Meditec AG., Germany) to enhance the visual feedback during needle insertion under the retina. Although intraoperative MI-OCT setups show several benefits in ophthalmic applications for operations both by human and robot, all the current intraoperative MI-OCT setups are adjusted manually and do not have the automatic needle detection and positioning function which extends the surgery time and also distracts the surgeon's attention. In parts of the procedure where OCT is needed the most, it is distracting the most. For instance, during the needle insertion, the surgeon needs to pay attention to a lot of information e.g. the tool pose and position information, tool shadow information for estimation the distance of the tool tip to retina, and also the B-scan image from X or Z direction for estimation the needle tip depth under the retina (see Fig. 2). During the injection, another assistance may need to manually localize the needle tip by adjusting the scan window position.

To localize the needle tip for subretinal injection, Zhou et al. [9] proposed an algorithm to predict the needle tip under retina using the geometrical information of the needle. The

[1]Chair for Robotics and Embedded Systems, Technische Universität München, Germany. {zhoum, xijia.wang, knoll}@tum.de
[2]Chair for Computer Aided Medical Procedures and Augmented Reality, Technische Universität München, Germany. {jakob.weiss, navab}@tum.de
[3]Carl Zeiss Meditec, Germany. abouzar.eslami@zeiss.com
[4]Key Laboratory of Machine Intelligence and Advanced Computing (Sun Yat-sen University), Ministry of Education, China School of Data and Computer Science, Sun Yat-sen University, China huangk36@mail.sysu.edu.cn
[5]Augenklinik und Poliklinik, Klinikum rechts der Isar der Technische Universität München, Germany. {mathias.maier, chris.lohmann,ali.nasseri}@mri.tum.de

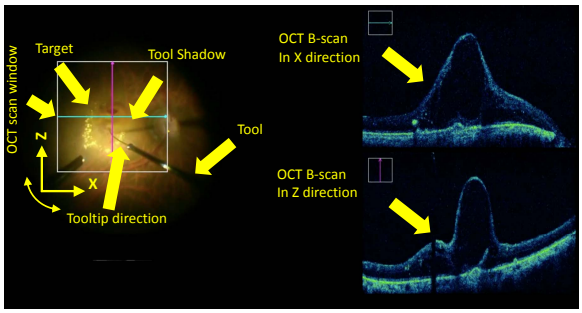



Fig. 2. The current OCT scan setup. The surgeon needs to pay attention to several information points labeled with the yellow arrow.

main premise is that the needle does not have large deformation during the insertion. However, the thin needle can bring better clinic performance for reducing injection trauma to the retina. The latest clinical subretinal microcannula used in clinic operation is 40-41 gauge (Ø=0.16-0.15mm) while previously available cannulas were only 32G (Ø=0.235mm). Furthermore, soft material for the needle tip part e.g. teflon is used instead of metal. Therefore, the premise does not meet very well with the thinner needle used in the current clinical operation which introduces significant deformation during puncture and retinal insertion.

In order to have a more universal framework with the variety of needles compatibility, and to improve the efficiency of the robot-assisted subretinal injection and furthermore, to give the surgeon enhanced information about the needle tip under the retina, we propose a novel framework to robustly localize needle segment under the retina using volumetric OCT images. The contributions of this paper are in two aspects: first and foremost, we take advantage of the needle information above the retina in B-scan images to coarsely predict the needle position under the retina and crop the original high-resolution image into a small region of interest (ROI). A convolutional neural network (CNN) is adopted to train on the cropped ROI images with the benefit of foreground and background balance, moreover, cropped images significantly reduce noise and reflection from the original B-scan images to enable facilitating very high accuracy results. Secondly, we create original and cropped datasets with 3811 OCT B-scan images in several ex-vivo pig eyes. The original image trained model and cropped image trained model are compared and evaluated. The results show that the cropped image trained model can localize the needle accurately with a confidence of 99.2%.

The remainder of the paper is organized as follows: in the next section, we briefly present related work. The proposed method is described in Section III. In Section IV, the performance of the proposed method is evaluated and discussed. Finally, Section V concludes this paper and presents the future work.

## II. Related Work

Having the benefit of suitable resolution and radiationless imaging mechanism, the OCT imaging modality now is popular not only in the ophthalmologic diagnostics but also in the intraoperative operation to enhance the visual feedback for the surgeon. Edwards et al. [1] performed the first subretinal injection with the robotic assistance under MI-OCT guidance which focuses on the human clinical trials and safety assessment. However, the OCT images could be processed to localize the needle segment instead of continues manually tracking the needle tip, which can further improve the system accuracy and reduce the surgical time without creating distraction for the surgeon. To track and localize the needle in vitreoretinal surgery, the possible method currently is based on either microscope or MI-OCT [9] providing appropriate precision. The single microscope camera was first introduced to track the needle pose in ophthalmic surgery using either color-based or geometry-based features [10]. However, localization of the needle in 3D with only a single view is not practically possible. Probst et al. [11] introduced a deep learning based method to localize the forceps in stereo microscopic images. This method has the advantage of low device cost, however, it faces challenges including the illumination varies and difficulty of detection instrument underlying tissue. Here we analyze the related work in two aspects which are the instrument pose estimation in ophthalmic surgery and the state of art for object detection.

The research on instrument localization in OCT images has attracted the attention of researchers. Zhou et al. [12] introduced a fully conventional neural network to segment the needle in volumetric OCT images when the needle is above the tissue. Weiss et al. [13] introduced a method to estimate the 5 DoF needle pose for navigation of subretinal injection. Gessert et al. [14] introduced a 3D convolutional neural network to directly estimate the 6D pose of the marker from the OCT volume. They used a marker with obvious geometrical features instead of a surgical instrument to simplify the problem. All of these methods focus on the needle localization above the tissue. For cases where the needle is below the retina, Zhou et al. [9] proposed a geometrical based method to calibrate the needle before insertion and then predict the needle tip position under retina during the injection. This method could be used when the needle has the geometrical feature to be tracked and the deformation of the needle is ignored. However, with thin needle diameter and soft material, which are becoming popular in ophthalmic applications, this method will be limited.

In order to have a versatile method to localize the needle under the retina, we propose detection and localization of the needle directly in B-scan images from OCT volume. Therefore, the problem is transformed into the object detection task. The object detection is rapidly developing specifically because of the advancements in deep learning technology and improvement in computational power. The two-stage detectors and one-stage detectors are the dominant object detectors in modern objection detection [15]. Recently, Lin et al. proposed a one-stage detector RetinaNet [16] to surpass the two-stage detector in accuracy and remain the speed advantage of one-stage detector with focal loss principle. Based on these advancements, we propose a robust needle localization framework for subretinal injection. The prosperity of our method is having the capability to localize the deformed needle under retina without geometrical feature which can be more versatile and feasible in the current and future clinic application.

## III. Method

The overall framework contains two parts: the first part is the robust ROI crop; and the second part is the deep learning based needle localization as shown in Fig. 3. The volumetric images are captured by the MI-OCT by scanning the injection area. The original B-scan images are processed with the ROI crop method to significantly reduce the size

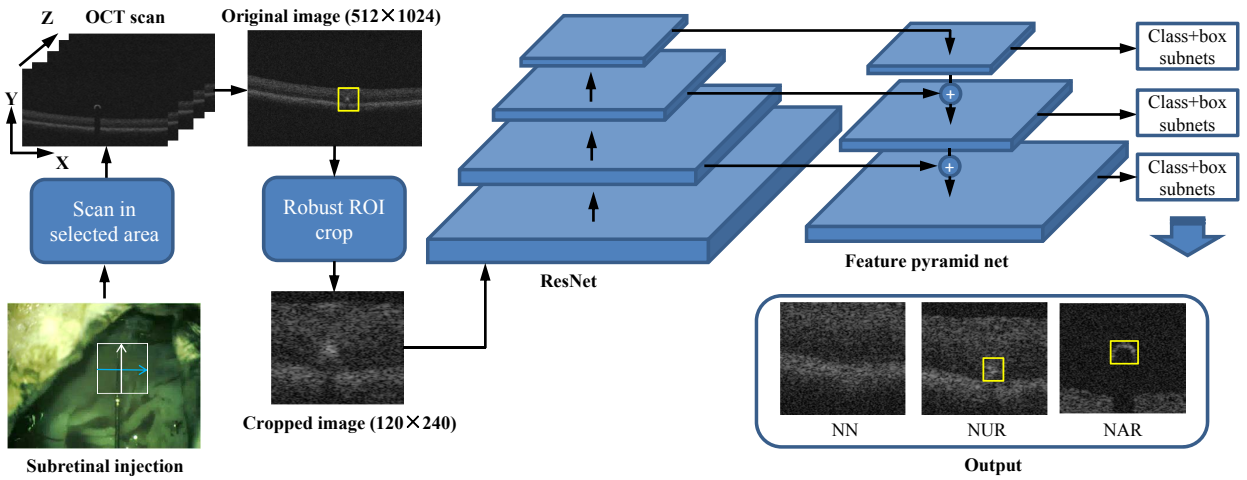



Fig. 3. The proposed framework. ResNet, feature pyramid net, and class+box subnets constitute RetinaNet. The output has three categories, no needle (NN), needle above retina (NAR) which means needle has no interaction with retinal tissue, and needle under retina (NUR) which means needle has interaction with retinal tissue.

of the image from 512×1024 to 120×240. Furthermore, cropped images are fed into the RetinaNet for needle segment detection and localization.

### A. Robust ROI Crop

The microscope used in this work is an OPMI LUMERA 700 with integrated RESCAN 700 OCT. The OCT engine has a wavelength of 840 nm and a scanning speed of 54000 A-scans per second. Due to the fact that the intended surgical area for subretinal injection is usually very small, we set the OCT engine to obtain the scan area 3 mm×3 mm×2 mm with the highest resolution of 128 B-scans, each with 512 A-scans. The benefit of high resolution is that it can reserve as much information as possible. Each of the B-scan images can potentially contain a needle segment. The needle bevel part cross-section is a spot of pixels while the needle body part cross-section is a half ellipse. When the needle is placed above the retina, the needle segment is isolated from the tissue and most of the time it creates a shadow on the retina. When the needle is inserted into the retina, some of the needle segments are under the surface layer reducing the clarity of the needle shape and making its position hard to be distinguished in B-scan image, especially with the presence of reflection and noise. Fig. 4 shows some examples of B-scan image with needle or without needle segment. The following points are taken into considerations in the ROI crop algorithm design: (1) the needle is a continuous object in the OCT volumetric image. (2) the very thin needle could be bent during the injection. (3) the needle tip is the most important segment to be localized. (4) most of the needle segments in OCT B-scans are above the retina since the injection depth is relatively shallow (less than 250 μm) compared to the OCT imaging range in depth. Based on point (4) we can apply the needle and retinal surface detection algorithms based on ellipse detection algorithms described in [17], the points in each B-scan that are used to fit the ellipse can be covered with a bounding box $B_i$. Moreover, we use the middle point on the upper bottom edge $M_i = (Mx_i, My_i, Mz_i)$ to fit the space polynomial curve which predicts the needle segment location instead of using the ellipse center because the center of the fitted ellipse will not be accurate. A second order polynomial is chosen because higher orders result in rippling

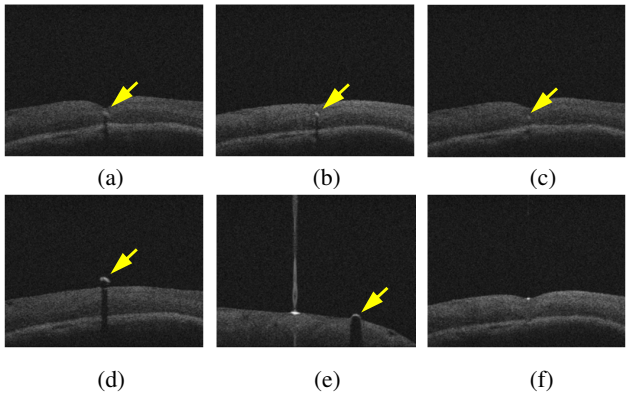



Fig. 4. (a), (b), (c), (d), (e) are examples with needle segment above and under retina (yellow arrow is used to localize the needle segment position). (f) is the one of confusion examples without needle but noise or reflection.

which effects uncharacteristic of the actual needle deflection. The needle model $N(x, y, z)$ is shown as,

$$N(x,y,z)\begin{cases} x(z) = a_2z^2 + a_1z + a_0 \\ y(z) = b_2z^2 + b_1z + b_0 \end{cases} \quad (1)$$

where $a_2$, $a_1$, $a_0$ and $b_2$, $b_1$, $b_0$ represent the parameters to be identified for the $N(x, y, z)$. The inlier dataset can be defined as,

$$inliers = \{e_i < \varepsilon\} \quad (2)$$

where $\varepsilon$ is the threshold for inlier tolerance, $e_i$ is the distance between point $M_i$ and point $N$ in Euclidean space which can be calculated as,

$$e_i = \sqrt{(Mx_i - x(Mz_i))^2 + (My_i - x(My_i))^2} \quad (3)$$

The point in $M_i$ will be categorized as outliers if $e_i$ is larger than $\varepsilon$. The rest points are treated as consensus set and a cost function $C$ will be calculated for all of these points in $M_i$ shown as follows,

$$C = \sum_{i=1}^{K} f(e_i) \quad (4)$$

where $f(e_i)$ equals $e_i$ when $e_i < \varepsilon$, otherwise equals $\varepsilon$. The M-Estimator Sample Consensus (MSAC) technique developed by Torr et al. [18] is used to obtain the optimal consensus set with minimization of the cost function. With the output of the parameters for needle model $N$, the needle location in all B-scan images can be predicted as shown in Fig. 5. Thus, the ROI bounding box $L_i = [Lx_i, Ly_i, w, h]$ ($Lx_i$ and $Ly_i$ are coordinated for the top left corner of the bounding box, $w$ and $h$ are the width and height of the bounding box) in $i$-th B-scan image is represented as,

$$L_i = \begin{cases} NaN, \text{if } y(Mz_i) < \frac{h}{2}. \\ [x(Mz_i) - \frac{w}{2}, y_{max} - h, w, y_{max}], \\ \text{if } y(Mz_i) > y_{max} - \frac{h}{2}. \\ [x(Mz_i) - \frac{w}{2}, y(Mz_i) - \frac{h}{2}, w, h], \\ \text{if } \frac{h}{2} < y(Mz_i) < y_{max} - \frac{h}{2}. \end{cases} \quad (5)$$

where $NaN$ means the bounding box does not exist, $y_{max}$ is maximum imaging depth. In this way, we could crop bounding box with lots of candidates where the needle may appear under retina. The needle reaching out of image will be ignored since needle segment in these candidates are always too far away above the retina which are not important information. $w$ and $h$ are decided by the needle diameter which ensures the needle to be covered by the cropped bounding box. Here with the needle of 40 gauge (Ø=0.16 mm), we set $w = 120$ and $h = 240$ where the larger $h$ can help to include the shadow information.

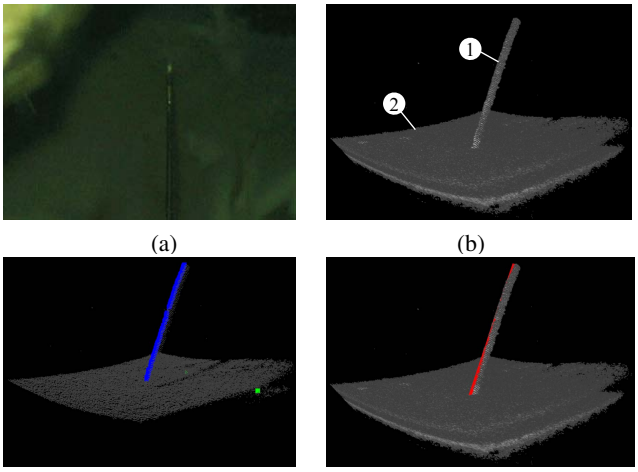



Fig. 5. (a) The microscope image with needle inserted inside retina. (b) The rendered OCT volume in oblique view. ① represents needle and ② represents retina. (c) $M_i$ points in OCT volume. The blue point are inliers and green points are outliers. (d) Needle model $N(x, y, z)$ in OCT volume. The red line is the RANSAC fitted needle model. The image crop operation are carried out based on $N(x, y, z)$ and Eqn. 5.

### B. Automatic Needle Localization Under Retina

For needle point classification and detection, RetinaNet [16] model is used. This network is a single unified network composed of backbone network and task specific subnetwork as shown in Fig. 3. Considering that we feed the image with the cropped size, the backbone network used here is Resnet18 for providing convolutional feature map and subnetwork perform the classification and bounding box regression task. The RetinaNet is one detector which can achieve comparable accuracy as two-stage detector while have the real time processing speed. The detectors normally have the problem of class imbalance which is even more serious in our needle detection scenario. Based on the subretinal injection tests on the ex-vivo pig eyes, the bounding box area of the needle segment is ranging from 352 to 1131 pixels which is a small region compared to the cropped image with 28,800 pixels and the original image with 524,288 pixels. This means that even though we have already cropped the image into a significantly smaller size compared to original image, the imbalance of the foreground and background problem still exits. Thus the focal loss concept is introduced to evaluate the unbalance of the foreground and background [16] as follows:

$$FL(p_t) = -\alpha(1 - p_t)^{\gamma} log(p_t) \quad (6)$$

where $p$ is the model's estimated probability for the specified class; $\alpha \in [0, 1]$ is the weighting factor for the specified class; and $\gamma \in [0, +\infty]$ is the tuneable focusing parameter. As previously introduced, the needle above the retinal surface can be detected and automatic annotated with the ellipse fitting algorithm. Therefore, the main difficulty comes when the needle detection under the surface. Thus we firstly manually annotate the needle which has the interaction with the retina (adhered to the retinal surface or under retinal surface), and then train the RetinaNet model with the annotated images. Moreover, we evaluate the trained model to localize the needle segment with our test dataset. The evaluation performance of the network will be introduced in the next section.

## IV. Experiments and Results

In this section, we present dataset preparation with our robot-assisted subretinal injection setup as well as the results related to the network performance with different parameters.

### A. Dataset Preparation

The dataset is collected on the robot-assisted subretinal injection platform with ex-vivo fresh pig eyes as shown in Fig. 6. The fresh ex-vivo pig eye (prepared for experiments within 2 hours after removal) has very similar structure with the human eye. The iRAM!S robot is mounted on an adjustment bracket. The OPMI LUMERA 700 with integrated RESCAN 700 intraoperative OCT engine is fixed on the optical table to reduce the influence of ambient vibration. OCT scan area can be easily relocated by the control panel. The iRAM!S eye surgical robot with 5DoF, with Pirezo motor technology, is utilized to perform the injection experiments. Piezo motor (SmarACT GmbH) provides 1 µm accuracy by using PID controller with integrated incremental optical encoder. The robot is adjusted manually by the adjustment bracket to make the needle tip approach to the retina. During the injection, only the motor that is holding the syringe is enabled to control the needle position. Thereafter, the motor is controlled to advance an exact 20 µm for each step and the OCT volume is captured after each movement. We stop the movement with enough insertion depth and the same procedure is repeated on the several ex-vivo pig eyes. Finally, we get 150 cubes with 19,200 images.

All images are processed by the ROI crop method and are divided into three categories, no needle (NN), needle above retina (NAR) which means needle has no interaction with retinal tissue, needle under retina (NUR) which means needle has interaction with retinal tissue. Since most of the images

present the needle above the retina, we randomly select 1000 images NN, 1000 images NAR, and 1000 images NUR from 140 cubes to balance the different categories. The rest 10 cubes with 811 images (287 images NN, 440 images NAR, and 84 images NUR) are selected as test dataset. All the selected and processed images are put into two datasets, one is original image dataset where the image size is 512×1024 and the other one is the cropped image dataset where the image size is 120×240.

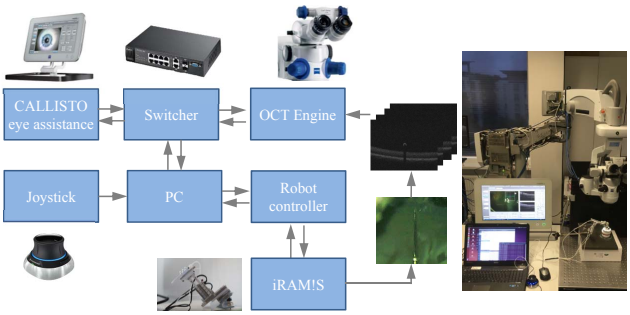



Fig. 6. The robot-assisted subretinal injection setup.

### B. Training and Evaluation

*1) Metrics:* Since the needle position is critical to be known in the future application, for example, detecting the needle insertion depth, intersection over union (IoU) for detected bounding box which represents the accuracy of needle position is used. The IoU is calculated as,

$$\text{IoU} = A_o/A_u \tag{7}$$

where $A_o$ is the area of the overlap for the detected bounding box and ground truth bounding box, $A_u$ is the area of the union for the detected bounding box and ground truth bounding box. To determine the performance of the needle with respect to the overall performance, we use the mean average precision (MAP) to calculate the mean value of average precision (AP) for each individual query $Q_i$. MAP can be calculated as,

$$\text{MAP} = \frac{1}{|Q|}\sum_{i=1}^{|Q|}\text{AP}(Q_i) \tag{8}$$

where $|Q|$ denotes the number of the categories. By introducing both metrics, we could analyze the performance of category accuracy and also the needle localization precision.

*2) Training:* Both datasets are trained with the same RetinaNet network for 100 epochs from scratch with around 5 hours in NVidia Titan-X GPU with i7-7700K CPU and 16 GB RAM. In order to determine the optimized $\gamma$ and $\alpha$, different combinations of $\gamma$ and $\alpha$ are tested and tuned with the objective function of AP and MAP. The output of the network contains many bounding boxes with the probability of which category it is. We only output the one with the maximum probability. For tuning the parameters, the IoU threshold is set to 0.5 which is considered as a good location predictor for filtering out the inaccurate predicted needle position. The trained model evaluation results are listed in Table I and Table II for the original dataset and cropped dataset, respectively. From these results, we can find that the cropped dataset trained model is significantly better than the original dataset trained model overall with the best-tuned result mAP of 0.97 and 0.81, respectively. Moreover, the interface time of cropped dataset trained model is furthermore reduced by 37% compared to the original dataset trained model. From these results, it can be seen that $\gamma$ and $\alpha$ have the ability to tune the performance to a certain degree. However, the original data trained model introduces not only larger foreground-background unbalance, but also more noise and reflection reducing the information entropy.

TABLE I
THE ORIGINAL DATASET TRAINED MODEL.

| $\gamma$ | $\alpha$ | AP_NN | AP_NAR | AP_NUR | MAP | time(ms) |
|---|---|---|---|---|---|---|
| 1.0 | 0.25 | 0.3539 | 0.0 | 0.0 | 0.1769 | 50.6 |
| 1.5 | 0.25 | 0.9663 | 0.8490 | 0.2639 | 0.7614 | 54.7 |
| 2.5 | 0.25 | 0.9503 | 0.8823 | 0.4367 | 0.8049 | 52.3 |
| 5.0 | 0.25 | 0.8750 | 0.8272 | 0.2560 | 0.6959 | 54.8 |
| 1.0 | 0.5 | 0.3539 | 0.8945 | 0.2938 | 0.4740 | 54.6 |
| 1.5 | 0.5 | 0.9535 | 0.8579 | 0.2595 | 0.7561 | 54.5 |
| 2.5 | 0.5 | 0.9696 | 0.8544 | 0.4194 | 0.8033 | 53.4 |
| 5.0 | 0.5 | 0.9025 | 0.8715 | 0.2671 | 0.7359 | 54.0 |
| 1.0 | 0.75 | 0.9631 | 0.8821 | 0.3148 | 0.7808 | 53.3 |
| 1.5 | 0.75 | 0.9795 | 0.8890 | 0.2828 | 0.7827 | 52.6 |
| **2.5** | **0.75** | **0.9729** | **0.8602** | **0.4552** | **0.8153** | **53.7** |
| 5.0 | 0.75 | 0.9535 | 0.8706 | 0.3184 | 0.7740 | 128.4 |

TABLE II
THE CROPPED DATASET TRAINED MODEL.

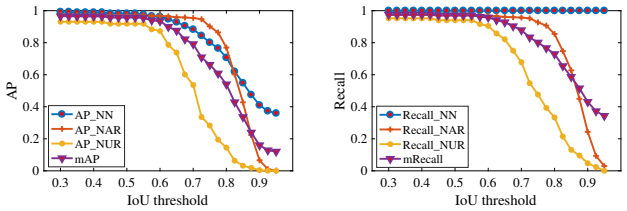
| $\gamma$ | $\alpha$ | AP_NN | AP_NAR | AP_NUR | MAP | time(ms) |
|---|---|---|---|---|---|---|
| 1.0 | 0.25 | 0.9897 | 0.9702 | 0.9011 | 0.9627 | 33.6 |
| 1.5 | 0.25 | 0.9931 | 0.9701 | 0.9180 | 0.9686 | 33.5 |
| 2.5 | 0.25 | 0.9863 | 0.9701 | 0.8991 | 0.9604 | 33.6 |
| 5.0 | 0.25 | 0.9729 | 0.9680 | 0.8470 | 0.9402 | 36.5 |
| 1.0 | 0.5 | 0.9897 | 0.9740 | 0.9057 | 0.9648 | 33.3 |
| 1.5 | 0.5 | 0.9897 | 0.9655 | 0.9221 | 0.9667 | 35.9 |
| 2.5 | 0.5 | 0.9931 | 0.9725 | 0.9125 | 0.9678 | 34.4 |
| 5.0 | 0.5 | 0.9897 | 0.9680 | 0.8732 | 0.9551 | 33.5 |
| 1.0 | 0.75 | 0.9931 | 0.9772 | 0.9080 | 0.9678 | 36.1 |
| 1.5 | 0.75 | 0.9931 | 0.9698 | 0.9006 | 0.9641 | 36.1 |
| **2.5** | **0.75** | **0.9931** | **0.9770** | **0.9201** | **0.9708** | **33.8** |
| 5.0 | 0.75 | 0.9897 | 0.9627 | 0.9043 | 0.9616 | 36.1 |

*3) Evaluation:* In order to give the sensitivity analysis of the IoU threshold value, we evaluate the AP and recall performance with different IoU threshold on the cropped dataset trained model($\gamma$=2.5, $\alpha$=0.75) shown in Fig. 7. The AP and recall value show steady until IoU threshold is more than 0.55 and then the AP and recall for needle detection (NAR and NUR) decrease to 0 when the IoU threshold is more than 0.9. Table III shows the evaluation result of annotation and detection with the IoU threshold of 0.55 in the best model. By integrating NAR and NUR into one category, meaning that the needle existing in the image regardless of the interaction situation of needle and retina, only 5 images are misclassified to NN. This indication shows that the method has the ability to localize the needle with the accuracy of 0.55 (IoU value) in the confidence of 99.2%. Fig. 8 shows some of the detected examples providing the method is working properly in most of the cases. It only fails to detect when needle tip is very small (usually the first needle tip in the B-scan). This will not reduce the operational quality since missing one B-scan needle tip detection will

TABLE III
THE EVALUATION OF THE THREE CATEGORIES.

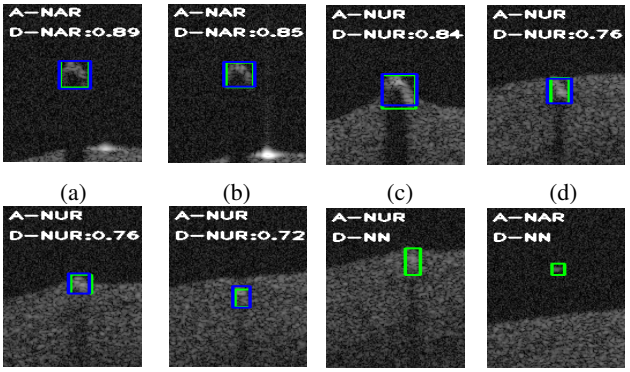
| | NN | NAR | NUR | Detection |
|---|---|---|---|---|
| NN | 287 | 2 | 3 | 292 |
| NAR | 0 | 427 | 2 | 429 |
| NUR | 0 | 11 | 79 | 90 |
| Annotation | 287 | 440 | 84 | |

only cause 23 μm (the scan resolution in Z direction) at most for needle localization error in one direction. This error is within the tolerance of needle localization requirement in subretinal injection which is 25 μm.




(a) AP vs IoU threshold (b) Recall vs IoU threshold

Fig. 7. (a) The function of IoU threshold with AP in three categories (AP_NN, AP_NAR, and AP_NUR). (b) The function of IoU threshold with AP in three categories (Recall_NN, Recall_NAR, and Recall_NUR), where mRecall denotes the mean recall value of three category.




(a) (b) (c) (d)

(e) (f) (g) (h)

Fig. 8. (a), (b), (c), (d), (e) and (f) are examples with correct detection. (g) and (f) are examples of incorrect detection when needle tip is very small. The yellow bounding box is the annotation-category (A-NAR, A-NUR, and A-NN) and blue bounding box is the detection-category (D-NAR, D-NUR, and D-NN). The number following is the IoU value.

## V. CONCLUSION

In this paper, we presented a flexible and robust framework for needle detection and localization in subretinal injection using MI-OCT images based on deep learning. Different from feeding data directly into the network, we take advantage of the needle geometrical features in volumetric OCT images to design a robust ROI crop method, thus the image size is significantly reduced. Afterward, the state of art one-stage detector named RetinaNet is applied to train the cropped images for the needle detection and localization. The evaluation results on ex-vivo pig eyes show that the performance of cropped image trained model is significant better than the original image trained model. The cropped image trained model can localize the needle accurately with a confidence of 99.2%. The future work will focus on integrating the proposed method with the intraocular path planning for vessel avoidance under the retina and safety assessment for the OCT servoing injection.

## ACKNOWLEDGMENT

We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan Xp GPU used for this research. We also thank Carl Zeiss Meditec AG for providing imaging device.

## REFERENCES

[1] T. Edwards, K. Xue, H. Meenink, M. Beelen, G. Naus, M. Simunovic, M. Latasiewicz, A. Farmery, M. de Smet, and R. MacLaren, "First-in-human study of the safety and viability of intraocular robotic surgery," *Nat. Biomed. Eng.*, p. 1, 2018.

[2] C. N. Riviere and P. S. Jensen, "A study of instrument motion in retinal microsurgery," in *Eng. Med. Biol. Soc. (EMBC), 2000 22th Annu. Int. Conf. IEEE*, vol. 1. IEEE, 2000, pp. 59–60.

[3] R. Taylor, P. Jensen, L. Whitcomb, A. Barnes, R. Kumar, D. Stoianovici, P. Gupta, Z. Wang, E. Dejuan, and L. Kavoussi, "A steady-hand robotic system for microsurgical augmentation," *Int. J. Rob. Res.*, vol. 18, no. 12, pp. 1201–1210, 1999.

[4] E. Rahimy, J. Wilson, T. C. Tsao, S. Schwartz, and J. P. Hubschman, "Robot-assisted intraocular surgery: development of the IRISS and feasibility studies in an animal model," *Eye*, vol. 27, no. 8, pp. 972–978, 2013.

[5] 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*, vol. 20, no. 21, pp. 23 414–23 421, 2012.

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

[7] A. Gijbels, E. Vander Poorten, B. Gorissen, A. Devreker, P. Stalmans, and D. Reynaerts, "Experimental validation of a robotic comanipulation and telemanipulation system for retinal surgery," in *2014 5th IEEE RAS EMBS Int. Conf. Biomed. Robot. Biomechatronics*. IEEE, 2014, pp. 144–150.

[8] 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 *Eng. Med. Biol. Soc. (EMBC), 2015 37th Annu. Int. Conf. IEEE*. IEEE, 2015, pp. 4906–4909.

[9] 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, to be appear." in *IEEE Int. Conf. on Rob. and Auto.*, 2018.

[10] N. Rieke, D. J. Tan, M. Alsheakhali, F. Tombari, C. A. di San Filippo, V. Belagiannis, A. Eslami, and N. Navab, "Surgical tool tracking and pose estimation in retinal microsurgery," in *Int. Conf. Med. Image Comput. Comput. Interv.* Springer, 2015, pp. 266–273.

[11] T. Probst, K.-K. Maninis, A. Chhatkuli, M. Ourak, E. Vander Poorten, and L. Van Gool, "Automatic tool landmark detection for stereo vision in robot-assisted retinal surgery," *IEEE Robot. Autom. Lett.*, vol. 3, no. 1, pp. 612–619, 2018.

[12] M. Zhou, H. Roodaki, A. Eslami, G. Chen, K. Huang, M. Maier, C. P. Lohmann, A. Knoll, and M. A. Nasseri, "Needle segmentation in volumetric optical coherence tomography images for ophthalmic microsurgery," *Applied Sciences*, vol. 7, no. 8, p. 748, 2017.

[13] J. Weiss, N. Rieke, M. A. Nasseri, M. Maier, A. Eslami, and N. Navab, "Fast 5dof needle tracking in ioct," *Int. J. Comput Assist. Radiol. Surg.*, vol. 13, no. 6, pp. 787–796, 2018.

[14] N. Gessert, M. Schlüter, and A. Schlaefer, "A deep learning approach for pose estimation from volumetric oct data," *Med. Image Anal.*, vol. 46, pp. 162–179, 2018.

[15] J. R. Uijlings, K. E. Van De Sande, T. Gevers, and A. W. Smeulders, "Selective search for object recognition," *International journal of computer vision*, vol. 104, no. 2, pp. 154–171, 2013.

[16] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," *IEEE transactions on pattern analysis and machine intelligence*, 2018.

[17] M. Zhou, M. Hamad, J. Weiss, A. Eslami, K. Huang, M. Maier, C. P. Lohmann, N. Navab, A. Knoll, and M. A. Nasseri, "Towards robotic eye surgery: Marker-free, online hand-eye calibration using optical coherence tomography images," *IEEE Robot. Autom. Lett.*, 2018.

[18] P. H. Torr and A. Zisserman, "Mlesac: A new robust estimator with application to estimating image geometry," *Computer vision and image understanding*, vol. 78, no. 1, pp. 138–156, 2000.