Published in final edited form as:

Achiev Solut Mech Eng II (2019). 2020 February ; 896: 183–194. doi:10.4028/www.scientific.net/AMM.896.183.

Real Time Prediction of Sclera Force with LSTM Neural Networks in Robot-Assisted Retinal Surgery

Changyan He^{1,a}, Niravkumar Patel^{1,b}, Marin Kobilarov^{1,c}, Iulian Iordachita^{1,d}

¹LCSR at the Johns Hopkins University, Baltimore, MD, 21218, USA

Abstract

Retinal microsurgery is one of the most technically demanding surgeries, during which the surgical tool needs to be inserted into the eyeball and is constantly constrained by the sclerotomy port. During the surgery, any unexpected manipulation could cause extreme tool-sclera contact force leading to sclera damage. Although, a robot assistant could reduce hand tremor and improve the tool positioning accuracy, it cannot prevent or alarm the surgeon about the upcoming danger caused by surgeon's misoperations, i.e., applying excessive force on the sclera. In this paper, we present a new method based on a Long Short Term Memory recurrent neural network for predicting the user behavior, i.e., the contact force between the tool and sclera (sclera force) and the insertion depth of the tool from sclera contact point (insertion depth) in real time (40Hz). The predicted force information is provided to the user through auditory feedback to alarm any unexpected sclera force. The user behavior data is collected in a mock retinal surgical operation on a dry eye phantom with Steady Hand Eye Robot and a novel multi-function sensing tool. The Long Short Term Memory recurrent neural network is trained on the collected time series of sclera force and insertion depth. The network can predict the sclera force and insertion depth 100 milliseconds in the future with 95.29% and 96.57% accuracy, respectively, and can help reduce the fraction of unsafe sclera forces from 40.19% to 15.43%.

Keywords

Neural network; scleral force; robot-assisted retinal surgery

Introduction

Safe manipulation is essential in retinal surgery due to its high precision requirement, small and constrained workspace, and delicate eye tissue. Factors such as physiological hand tremor, fatigue, poor kinesthetic feedback, patient movement, and the absence of force sensing could potentially lead to surgeon's misoperations, and subsequently iatrogenic injury. During retinal surgery, a surgical microscope is placed above the patient to provide the magnified visualization of the interior of the eye. The surgeon inserts small instruments (e.g. 25Ga) through the sclerotomy port ($\phi < 1\text{mm}$) on the sclera, the white part of the eye, to perform delicate tissue manipulations in the posterior of the eye as shown in Fig.1. One example of such challenging retinal surgical tasks is retinal vein cannulation (RVC),

^a changyanhe@jhu.edu . ^b snpatel89@jhu.edu . ^c marin@jhu.edu . ^d iordachita@jhu.edu .

which is a most promising treatment for the retinal vein occlusion. During this operation, surgeon needs to pass and hold the needle or micropipette through the sclerotomy port and carefully insert it into the occluded retinal vein. After insertion, the surgeon needs to hold the tool steadily for as long as two minutes for clot-dissolving drug injection. Steady and precise motion is required, because the diameter of the retina vein can be as tiny as 80 μm , while the human hand tremor is over 100 μm on average [1]. Additionally, the tool-tissue interaction force is well below the human sensory threshold [2], thus retinal surgical tasks demand advanced surgical skills that are near or beyond natural human capabilities. The success of retinal surgeries still strongly depends on the surgeon's level of experience and skills, a requirement that could be relaxed with the help of more advanced robotic assistive technology.

Continuing efforts are being devoted to the development of the surgical robotic systems to enhance and expand the capabilities of surgeons during retinal surgeries. Two robot-assisted retinal surgeries, RVC and epiretinal membrane peeling (EMP), have been performed successfully on human patients recently [3], [4], demonstrating the clinical feasibility of the robotic technology for retinal microsurgery. Major enabling technologies for robot-assisted retinal surgeries include teleoperative manipulation systems [5]-[7], handheld robotic devices [8], untethered micro-robots [9], and flexible micro-manipulators [10].

Our team has proposed the cooperative control approach and developed the Steady Hand Eye Robot (SHER) [11]. SHER allows the user to directly hold the tool mounted on the robot end-effector. The velocity of SHER follows the user input force applied on the tool handle, while the user's hand tremor is damped by the robot stiff mechanical structure. SHER provides not only the steady and precise motion, but also the transparent manipulation. Additionally, our group previously designed and developed a family of "smart" tools based on Fiber Brag Grating (FBG) sensor [12], in which the multi-function sensing tool [13] can measure the contact force between tool tip and retina (tip force), tool shaft and sclerotomy port (sclera force), and tool depth inserted inside the eyeball (insertion depth). The force sensed from the "smart" tools can be used to give auditory feedback and thus help improve the surgeon's awareness [14].

The aforementioned robotic devices can give "passive support" to the surgeons, i.e. filtering the hand tremor, improving the tool location accuracy, and feeding the operation force information back to surgeon. However, the unsafe manipulations caused by the lack of skills or fatigue cannot be "actively" filtered or prevented by the robotic system, due to the lack of the information about the surgeon's behavior in advance.

Making prediction in advance for the surgeon's next movement (e.g. 100 millisecond in the future) based on his/her operation history can potentially resolve the above mentioned problems. The predicted information can be used to identify the unsafe manipulations that might occur in the near future, e.g. the extreme operation force exerted on the eye tissue. Then the robotic system can actively provide auditory alarm, interrupt or rectify these mis-manipulations. The surgeon's performance in a surgical task is time based, i.e., time series, on which Recurrent Neural Networks (RNNs) can effectively handle. Especially, the Long Short Term Memory (LSTM) RNNs [15] can capture long range dependencies and nonlinear

dynamics and model varying-length sequential data, achieving state-of-the-art results for problems spanning clinical diagnosis [16], image segmentation [17], and language modeling [18]. Other standard sequential models, such as Kalman filters or fixed-lag smoothers and predictors, are ill-equipped to learn long-range dependencies [16].

In this paper, we report initial results for predicting possibly unsafe user behavior immediately before it is about to occur, to provide enough reaction time for preventing tissue damage during robot-assisted retinal surgery. An LSTM model is adopted and trained to predict the user behavior, which is defined through two correlated time-series: the sclera force and the insertion depth as shown in Fig. 2 (a). The predicted information is compared with known safe operation thresholds to give auditory feedback, and to alarm users about any upcoming misoperations. Experiments are conducted on a dry eye phantom to examine the feasibility of the proposed method.

The goal of this work is to establish the user's behavior library, and achieve a safety-aware cooperative control in robot-assisted surgical systems. While there are a number of other signals useful for behavior prediction, such as tool inclination and eye tracking and imaging, this initial investigation is limited to the two basic signals: sclera force and insertion depth.

Robot control method design

The surgeon manipulates the robotic assistance through the admittance control law. Surgeon's behavior data is collected by the sensing system, and fed into the LSTM neural network. The network predicts the sclera force 100 milliseconds in the future and provides it to the auditory feedback module. This module can generate three different sounds corresponding to safe, warning, and danger status based on the three force thresholds set in advance. Once the predicted sclera force is beyond the threshold, the user will get appropriate sound alarm. Besides auditory substitution, the predicted information can be further used to implement the multi-variable admittance control [13] to enable the robot actively react to the user's misoperations.

A. Admittance Control Law

During operation both the user and the robot hold the tool, the user's manipulation force is applied on the robot handle and fed as an input into the admittance control law as shown in Eq. (2)

$$\dot{x}_{hh} = \alpha F_{hh} \quad (1)$$

$$\dot{x}_{rh} = Ad_{grh} \dot{x}_{hh} \quad (2)$$

where $\dot{x}_{hh}$ and $\dot{x}_{rh}$ are the desired robot handle velocities in the handle frame and in the robot frame, respectively, F_{hh} is the user's manipulation force input measured in the robot handle frame, α is the admittance gain tuned by the robot pedal, Ad_{grh} is the adjoint transformation.

B. Force calculation algorithm

The multi-function sensing tool is designed and fabricated as shown in Fig. 2. The tool shaft is made of a stainless steel wire with a diameter of 0.635 mm and machined to contain three longitudinal V-shape grooves. Each groove is filled and glued with one optical fiber containing three FBG sensors (Technica S.A, Beijing, China), i.e., nine FBG sensors are embedded in the tool shaft in total. The tool dimensions are shown in Fig. 2.

The multi-function sensing tool can measure the sclera force and the insertion depth. Considering ambient temperature and noise, we use the sensor reading s to calculate the force defined as follows:

$$\Delta s_{ij} = \Delta \lambda_{ij} - \frac{1}{3} \sum_{j=1}^3 \Delta \lambda_{ij}, \quad (3)$$

where λ_{ij} is the wavelength shift of the FBG, $i=I, II$ is the FBG segment, $j=1, 2, 3$ denotes the FBG sensors on the same segment. The sclera force exerted on the tool shaft generates strain in the FBG sensors, thus it contributes to the sensor reading:

$$\Delta S_i = K_i M_i = K_i F_s d_i, \quad (4)$$

where $S_i = [s_{i1}, s_{i2}, s_{i3}]^T$ denotes the sensor reading of FBG sensors in segment i , $F_s = [F_x, F_y]^T$ denotes the sclera force applied at sclerotomy port, d_i denotes the distance from the sclerotomy port to FBG at segment i along the tool shaft, $M_i = [M_x, M_y]^T$ denotes the moment attributed to F_s on FBG sensors of segment i , $K_i (i=I, II)$ is 3x2 constant coefficient matrices, which is obtained through the tool calibration procedures.

Then the sclera force can be calculated using Eq. (5):

$$F_s = \frac{M_{II} - M_I}{\Delta l} = \frac{K_{II}^* \Delta S_{II} - K_I^* \Delta S_I}{\Delta l} \quad (5)$$

C. RNN architecture design

We assume that the surgeon's behavior can be captured through a history of time series of sensor measurements. An LSTM network was constructed in order to make future predictions based on such history. An LSTM [15] is based on memory cells composed of four main elements: one input gate, one forget gate, one output gate, and one neuron with a self-recurrent connection (a connection to itself). The gates serve to modulate the interactions between the memory cell itself and its environment. The input gate determines the importance of new information, while the output gate determines the part of the current information to push out. The forget gate can modulate the memory cells self-recurrent connection, allowing the cell to remember or forget its previous state, as needed. In this work, the proposed LSTM neural network uses memory cells with forget gates but without peephole connections. To perform sensor reading predictions, a fully connected layer is used as the network output after the LSTM as shown in Fig. 3. The mean absolute error (MAE) is used as the loss function at each output.

The following standard equations [19] describe recurrent algebraic relationship in the LSTM:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f), \quad (6)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \quad (7)$$

$$\tilde{C}_t = \phi(W_C \cdot [h_{t-1}, x_t] + b_C), \quad (8)$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t, \quad (9)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o), \quad (10)$$

$$h_t = o_t \odot \phi(C_t), \quad (11)$$

where h_{t-1} stands for the memory cells at the previous sequence step, σ stands for an element-wise application of the sigmoid (logistic) function, ϕ stands for an element-wise application of the tanh function. The input, output, and forget gates are denoted by i , o , and f respectively, while C is the cell state. W and b are the weight and bias, respectively.

After updating the memory cell at current time, the output of the network is calculated with the dense layer:

$$Y_{t+p} = W_d \cdot h_t + b_d \quad (12)$$

Experiments and results

A. Experimental setup

The experimental setup is shown as Fig. 4(a), including SHER, the multi-function sensing tool, a microscope, a FBG interrogator, an eye phantom, a monitor and a speaker. SHER is used as robot assistance to provide the “co-operation” for the user. It has 5 degrees of freedom and high accuracy (translation resolution is less than $3\mu\text{m}$, and the rotation resolution is 0.0005°). The multi-function sensing tool is mounted on SHER with a quick release mechanism and used to collect the user behavior data. It was calibrated based on Eq. (4) using a precision scale (Sartorius ED224S Extend Analytical Balance, Goettingen Germany) with resolution of 1 mg.

The tool validation experiment was carried out with the same scale to test the calibration results. The validation results of the RMS errors for sclera force and insertion depth are 2.3 mN and 0.6 mm, respectively. The sm 130-700 optical sensor interrogator (Micron Optics, Atlanta, GA) is used to monitor the FBG sensors within the spectrum from 1525 nm to 1565 nm at 2 kHz refresh rate.

A dry eye phantom made of silicon rubber placed into a 3D-printed socket was used for validation. A printed paper with several curves representing the retinal vessels in different colors is glued on the eyeball surface as shown in Fig.4 (b). In addition, a microscope (ZEISS, Germany) is utilized to provide the magnified view for users to observe inside the eyeball, and a Point Grey camera (FLIR Systems, Inc.) is attached on the microscope for recording the user's interaction with the eyeball.

B. Surgical task design

The mock retinal surgery operation for collecting the user behavior data is designed as follows:

- (1) insert the tool into the eyeball through the sclerotomy port;
- (2) adjust the eyeball's position and orientation for best view under the microscope using the inserted tool;
- (3) insert the tool to the home point, and then following one vessel to the end without touching the retina surface;
- (4) retract the tool axially until the tip reaches the eyeball boundary.

It should be noted that the data recording begins after the tool is inserted into the eyeball around 10mm, to make the tool contact with the sclerotomy port, thereby to get the reliable insertion depth since the Eq. (6) is only valid when the sclera force exceeds 10 mN as discussed in II-B.

C. STM neural network training

Successful training critically depends on proper choice of network hyper-parameters [20]. To find a suitable set of the hyper-parameters, i.e., network size and depth, several validation trials were carried out as shown in Table I. One LSTM layer with 500 neural units and two dense layers with 200 neural units in first layer and 1 neural unit in second layer are selected eventually, since the model with this set of parameters resulted in the smallest validation errors. The *tanh* and *hard sigmoid* were employed as activation function and recurrent activation function, respectively. The input of the network are the sclera force and the insertion depth with 150 timesteps history, and the output of the network are the same quantities 20 timesteps in the future, as shown in following equation:

$$[F_s(t+p), d(t+p)] = f[F_s(t-1, t-2, \dots, t-n), d(t-1, t-2, \dots, t-n)]$$

where F_s is sclera force, d is insertion depth, $n = 150$ denotes the input data history, and $p = 30$ denotes the timestep predicted ahead.

One timestep represents 5 milliseconds as the data collection frequency is 200Hz, therefore, the past 750 milliseconds user behavior data is collected and fed to the network to produce the one in 100 milliseconds in the future. Data for 60 repetitions of the surgery task described in section III-B is collected for network training and testing, and 70% of the data

is used for training and the remaining 30% is used for validation. The training dataset was divided into mini-batches of sequences of size 200. The Adam optimizer [21] and MAE is applied in this process. The network was implemented using Keras [22] and the training was performed on a system with Nvidia Titan Black GPUs, a 20-cores CPU, and 128GB RAM. Single-GPU training took 90 minutes on the integral data. The loss of the training process decreases from 15.54 at the beginning to 4.32 in the end as shown in Fig. 5.

D. Sclera Force Prediction and Feedback in Real Time

The experiments of real time sclera force prediction and signaling are carried out to validate the prediction accuracy of the proposed LSTM neural network and the effectiveness of the auditory feedback on safe sclera force manipulation. The network predictor runs in PYTHON at 40 Hz, while the sensing system runs in C++ at 200 Hz. A "shared memory" architecture is used to transfer the actual user operation data and network predictions between PYTHON and C++.

- 1). Prediction accuracy validation: Real time prediction experiments were conducted to compare the actual user behavior data and the predicted one, and to validate the prediction accuracy. The user was asked to hold the multi-function sensing tool mounted on SHER to perform the mock retinal surgery (described in III-B) 50 times. The actual and the predicted user behavior data was collected and analyzed, where the MAE for sclera force and insertion depth are 12.2 ± 16.6 mN and 1.03 ± 1.4 mm, respectively. Considering the maximum sclera force (261.8 mN) and insertion depth (30.2 mm) during the operation, the prediction accuracy of the sclera force and insertion depth are 95.29% and 96.57%, respectively. Result from one of the 50 trials is shown in Fig. 6, as well as the prediction MAE in all trials.
- 2). Auditory feedback implementation: Experiments under three conditions were carried out:
 - with auditory feedback based on predicted sclera force (AFPF);
 - with auditory feedback based on instantaneous operative sclera force (AFIF);
 - without auditory feedback (No AF)

The experiments under the last two conditions are considered as the comparison experiments. In AFPF and AFIF, the user gets three different auditory signals representing safe, warning, and danger when the predicted sclera force (in AFPF) or actual sclera force (in AFIF) exceeds three force thresholds: 100 mN, 120 mN, and 140 mN, respectively. In this paper, the sclera force of 140 mN is assumed as the maximum safety force threshold based on our previous study [23]. The user is expected to react to the auditory feedback signal and adjust his/her manipulation to decrease the sclera force. In each condition, the user was asked to repeat the procedures described in section III-B 20 times. The overall sclera force in each condition for all the trials is depicted in Fig. 7, as well as the average value at a time interval of 50 milliseconds. Both the average value curves of the AFPF and AFIF are below the force threshold, while the average value curve of the No AF exceeds the force threshold. As shown in Table II, we analyzed the elapsed time of the unsafe force (force larger than 140 mN) in each condition, the time percentage with unsafe force in AFPF is the smallest among the three conditions.

Discussion and conclusion

The experiment results of the real time prediction shows that the prediction MAE of the sclera force is 12.2 ± 16.6 mN. Considering the maximum measured sclera force of 261.8 mN during typical operation, such predicted values can be regarded as statistically useful data. One of the factors affecting the prediction accuracy is the force measurement accuracy of the tool itself (the tool validation RMS error is 2.3 mN). In the auditory feedback experiments, both the AFPF and AFIF can help decrease the manipulation force comparing to No AF, while AFPF enhances the performance slightly compared to AFIF; the reason is that the admittance controlled robot prolongs the user's reacting time; therefore the forecasting time of 100 milliseconds may not be long enough for user to practically react. In future, we will focus on improving the network's behavior to provide the prediction information further into the future without significant loss of accuracy. Also, we noticed that the experiment time in No AF is smallest among the three conditions, since the user needs to adjust his/her manipulation once he/she obtains the warning sound in auditory feedback conditions.

Currently the scleral force is predicted using the LSTM- based regression method, though it can provide accurate prediction information, the prediction is limited to a short future time. The future work will investigate the classification method using deep learning method to procure the prediction. The preliminary study of that can be found in [24]. In the same time only the predicted sclera force is utilized and fed back directly to the user through auditory substitution. Future work will incorporate the predicted insertion depth together with the predicted sclera force to implement multiple variable admittance control, thereby to enable robotic assistance itself to react to imminent tissue damage during operation. The preliminary study of the admittance control is presented in [25]. In addition, as the first step in this research, we collected the manipulation data with one user, and on a specific operation task to test the feasibility of the proposed method. Future work will involve multiple users and different surgical tasks. The multi-function sensing tool can also measure the force at the tool tip, which will also be incorporated into the network training in future work to provide more information for robot control.

Retinal surgery requires delicate skill of the surgeon, any misoperations, i.e. large sclera manipulation force, could cause the tissue damage. Many researchers have investigated robotic devices to assist surgeons with the operation, which can improve the manipulation accuracy and cancel the hand tremor, but cannot help to prevent the unintentional misoperations. The method to provide the surgeons with auditory feedback based on the predicted operative force can signal the upcoming danger and enhance the safety of the retinal surgery. This method can also be applied to other minimally invasive surgeries, to provide additional useful information to improve the surgical robot control performance.

Our research sits at the intersection of three fields: robotics, neural network learning, and retinal surgery safe manipulation. Although each of them is an established field, it is the first time to apply LSTM networks to robot-assisted retinal surgery to enhance safety. We designed and implemented the network to learn the user operative behavior in a mock retinal surgery, and developed a new multi-function sensing tool to collect the user manipulation information including the sclera force and insertion depth. The auditory feedback based

on predicted sclera force produced by the trained network model was implemented to signal the upcoming dangerous manipulation force. This system can potentially provide safe micromanipulation that can improve the outcome of the retinal microsurgery, also to establish the user behavior library and achieve the safety-aware cooperative control in robot-assisted surgical systems. In the future, we will further investigate this system in phantoms and in vivo experiments.

Acknowledgment

This work was supported by U.S. National Institutes of Health under grant number 1R01EB023943-01. The work of C. He was supported in part by the China Scholarship Council under Grant 201706020074, the National Natural Science Foundation of China under Grant 51875011, and the National Hi-tech Research and Development Program of China with grant 2017YFB1302702

References

- [1]. Singhy S and Riviere C, "Physiological tremor amplitude during retinal microsurgery," in Bioengineering Conference, 2002. Proceedings of the IEEE 28th Annual Northeast. IEEE, 2002, pp. 171–172.
- [2]. Gupta PK, Jensen PS, and de Juan E, "Surgical forces and tactile perception during retinal microsurgery," in International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 1999, pp. 1218–1225.
- [3]. Walsh F, Robot operates inside eye in world first, 2016. [Online]. Available: <http://www.bbc.com/news/health-37246995>
- [4]. Donald T, "Auditory force feedback substitution improves surgical precision during simulated ophthalmic surgery," Retina Today, vol. 4, pp. 59–61, 2017.
- [5]. He C, Huang L, Yang Y, Liang Q, and Li Y, "Research and realization of a master-slave robotic system for retinal vascular bypass surgery," Chinese Journal of Mechanical Engineering, vol. 31, no. 1, p. 78, 2018.
- [6]. Gijbels A, Wouters N, Stalmans P, Van Brussel H, Reynaerts D, and Vander Poorten E, "Design and realisation of a novel robotic manipulator for retinal surgery," in Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on. IEEE, 2013, pp. 3598–3603.
- [7]. de Smet MD, Meenink TC, Janssens T, Vanheukelom V, Naus GJ, Beelen MJ, Meers C, Jonckx B, and Stassen J-M, "Robotic assisted cannulation of occluded retinal veins," PloS one, vol. 11, no. 9, p. e0162037, 2016. [PubMed: 27676261]
- [8]. MacLachlan RA, Becker BC, Tabar e's JC, Podnar GW, Lobes LA Jr, and Riviere CN, "Micron: an actively stabilized handheld tool for microsurgery," IEEE Transactions on Robotics, vol. 28, no. 1, pp. 195–212, 2012. [PubMed: 23028266]
- [9]. Kummer MP, Abbott JJ, Kratochvil BE, Borer R, Sengul A, and Nelson BJ, "Octomag: An electromagnetic system for 5-dof wireless micromanipulation," IEEE Transactions on Robotics, vol. 26, no. 6, pp. 1006–1017, 2010.
- [10]. Hubschman J, Bourges J, Choi W, Mozayan A, Tsirbas A, Kim C, and Schwartz S, "the microhand: a new concept of micro-forceps for ocular robotic surgery," Eye, vol. 24, no. 2, p. 364, 2010. [PubMed: 19300461]
- [11]. Uneri A, Balicki MA, Handa J, Gehlbach P, Taylor RH, and Iordachita I, "New steady-hand eye robot with micro-force sensing for vitreoretinal surgery," in Biomedical Robotics and Biomechatronics (BioRob), 2010 3rd IEEE RAS and EMBS International Conference on. IEEE, 2010, pp. 814–819.
- [12]. Iordachita I, Sun Z, Balicki M, Kang JU, Phee SJ, Handa J, Gehlbach P, and Taylor R, "A sub-millimetric, 0.25 mm resolution fully integrated fiber-optic force-sensing tool for retinal microsurgery," International journal of computer assisted radiology and surgery, vol. 4, no. 4, pp. 383–390, 2009. [PubMed: 20033585]

- [13]. He X, Balicki M, Gehlbach P, Handa J, Taylor R, and Iordachita I, "A multi-function force sensing instrument for variable admittance robot control in retinal microsurgery," in Robotics and Automation (ICRA), 2014 IEEE International Conference on. IEEE, 2014, pp. 1411–1418.
- [14]. Cutler N, Balicki M, Finkelstein M, Wang J, Gehlbach P, Mc-Gready J, Iordachita I, Taylor R, and Handa JT, "Auditory force feedback substitution improves surgical precision during simulated ophthalmic surgery," *Investigative ophthalmology & visual science*, vol. 54, no. 2, pp. 1316–1324, 2013. [PubMed: 23329663]
- [15]. Hochreiter S and Schmidhuber J, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997. [PubMed: 9377276]
- [16]. Lipton ZC, Kale DC, Elkan C, and Wetzel R, "Learning to diagnose with lstm recurrent neural networks," *arXiv preprint arXiv:1511.03677*, 2015.
- [17]. Stollenga MF, Byeon W, Liwicki M, and Schmidhuber J, "Parallel multi-dimensional lstm, with application to fast biomedical volumetric image segmentation," in *Advances in neural information processing systems*, 2015, pp. 2998–3006.
- [18]. Sundermeyer M, Schluter R, and Ney H, "Lstm neural networks for language modeling," in *Thirteenth Annual Conference of the International Speech Communication Association*, 2012.
- [19]. Graves A, Mohamed A, and Hinton G, "Speech recognition with deep recurrent neural networks," in *Acoustics, speech and signal processing (ICASSP), 2013 IEEE international conference on. IEEE*, 2013, pp. 6645–6649.
- [20]. Greff K, Srivastava RK, Koutn ík J, Steunebrink BR, and Schmidhuber J, "Lstm: A search space odyssey," *IEEE transactions on neural networks and learning systems*, vol. 28, no. 10, pp. 2222–2232, 2017. [PubMed: 27411231]
- [21]. Kingma DP and Ba J, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [22]. Chollet Fet al., "Keras," <https://github.com/keras-team/keras>, 2015.
- [23]. Horise Y, He X, Gehlbach P, Taylor R, and Iordachita I, "FBG-based sensorized light pipe for robotic intraocular illumination facilitates bi-manual retinal microsurgery," in *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE. IEEE*, 2015, pp. 13–16.
- [24]. He C, Patel N, Iordachita I, and Kobilarov M, "Enabling technology for safe robot-assisted retinal surgery: Early warning for unsafe scleral force," in *Robotics and Automation (ICRA), 2019 IEEE International Conference on.*, 2019.
- [25]. He C, Patel N, Ebrahimi A, Kobilarov M, and Iordachita I, "Preliminary study of an rnn-based active interventional robotic system (airs) in retinal microsurgery," *International journal of computer assisted radiology and surgery*, pp. 1–10, 2019.

Fig. 1. Retinal microsurgery: (a) the surgeon performs the procedure under the microscope. (b) two tools are inserted into eyeball, one is cannulation instrument to inject medicine into the retinal vein in the case of RVC, the other is a lightpipe to illuminate the posterior of the eye.

Fig. 2.

Multi-function sensing tool. (a). The depiction of the sclera force and the insertion depth, the sclera force is the contact force between the sclerotomy port and the tool shaft, the insertion depth is the distance between the tool tip and the sclerotomy port when tool is inside of the eyeball. (b). the tool dimensions.

Fig. 3.

The proposed LSTM neural network. The recurrent layer is unrolled. The network gets input from data history with n timestep in past, and output the future one at time $t+p$.

Fig. 4.
Experimental setup of the mock retinal surgery with (a). robotic assistance and (b). the dry eye model

Fig. 5.
Network loss performance during training.

Fig. 6.

The real time prediction results: (a). one sample of the real time prediction result, (b). the magnified part of the prediction result of (a) from 5.5s to 6.5s, (c). the prediction MAE of the 50 trials.

Fig. 7.

The overall sclera force performance in three different conditions. The sclera force data of 20 trials in AFPF, AFIF, and No AF are shown in (a), (b) and (c), respectively, the red line shows the safe force threshold of 140mN. The percentage between the portion of unsafe sclera force (force larger than the threshold) and the whole data in AFPF, AFIF and No AF is 15.43%, 17.13%, and 40.19%, respectively.

TABLE I

VALIDATION RESULTS FOR DIFFERENT HYPER-PARAMETERS

RNN model	Neuron Numbers		Validation RMS error	
	LSTM	Dense	Sclera force (mN)	Insertion depth (mm)
LSTM with single dense	300	1	12.3±12.4	2.08±2.07
	500	1	10.8±10.4	1.2±0.8
Layer LSTM	300	200, 1	14.9±11.7	1.3±0.7
	300	300, 1	13.7±11.1	1.2±0.8
	500	200, 1	10.4±8.2	1.1±0.8
Multi-LSTMs	200, 100	1	23.5±18.2	6.8±5.7
	200, 200	1	42.5±35.4	8.2±7.3

TABLE II

The unsafe force percentage

	Time of unsafe force(s)	Total elapsed time(s)	Unsafe force percentage
AFPF	59.96	388.49	15.43%
AFIF	47.96	279.95	17.13%
No AF	68.76	171.1	40.19%