

Vitreoretinal Surgical Instrument Tracking in Three Dimensions Using Deep Learning

Pierre F. Baldi^{1-4,*}, Sherif Abdelkarim^{1,2,*}, Junze Liu^{1,2,*}, Josiah K. To^{4,*},
Marialejandra Diaz Ibarra⁵, and Andrew W. Browne³⁻⁵

¹ Department of Computer Science, University of California, Irvine, CA, USA

² Institute for Genomics and Bioinformatics, University of California, Irvine, CA, USA

³ Department of Biomedical Engineering, University of California, Irvine, CA, USA

⁴ Center for Translational Vision Research, Department of Ophthalmology, University of California, Irvine, CA, USA

⁵ Gavin Herbert Eye Institute, Department of Ophthalmology, University of California, Irvine, CA, USA

Correspondence: Andrew W. Browne, Department of Computer Science, University of California, 4038 Bren Hall, Irvine, CA 92697, USA. e-mail: abrowne1@uci.edu
Pierre F. Baldi, Department of Computer Science, University of California, 4038 Bren Hall, Irvine, CA 92697, USA. e-mail: pfbaldi@uci.edu

Received: September 19, 2022

Accepted: December 23, 2022

Published: January 17, 2023

Keywords: artificial intelligence; retina surgery; deep learning

Citation: Baldi PF, Abdelkarim S, Liu J, To JK, Ibarra MD, Browne AW. Vitreoretinal surgical instrument tracking in three dimensions using deep learning. *Transl Vis Sci Technol.* 2023;12(1):20. <https://doi.org/10.1167/tvst.12.1.20>

Purpose: To evaluate the potential for artificial intelligence-based video analysis to determine surgical instrument characteristics when moving in the three-dimensional vitreous space.

Methods: We designed and manufactured a model eye in which we recorded choreographed videos of many surgical instruments moving throughout the eye. We labeled each frame of the videos to describe the surgical tool characteristics: tool type, location, depth, and insertional laterality. We trained two different deep learning models to predict each of the tool characteristics and evaluated model performances on a subset of images.

Results: The accuracy of the classification model on the training set is 84% for the x-y region, 97% for depth, 100% for instrument type, and 100% for laterality of insertion. The accuracy of the classification model on the validation dataset is 83% for the x-y region, 96% for depth, 100% for instrument type, and 100% for laterality of insertion. The close-up detection model performs at 67 frames per second, with precision for most instruments higher than 75%, achieving a mean average precision of 79.3%.

Conclusions: We demonstrated that trained models can track surgical instrument movement in three-dimensional space and determine instrument depth, tip location, instrument insertional laterality, and instrument type. Model performance is nearly instantaneous and justifies further investigation into application to real-world surgical videos.

Translational Relevance: Deep learning offers the potential for software-based safety feedback mechanisms during surgery or the ability to extract metrics of surgical technique that can direct research to optimize surgical outcomes.

Introduction

Retina surgery is performed in the closed confines of the eye and all surgical manipulations are performed manually by a trained surgeon. To achieve the goals of surgery, the surgeon uses microsurgical instruments and a microscope to visualize the intraocular contents through the pupil of the eye. Other than human observers and surgeons, there are no safety mechanisms to prevent error. Further, each

surgeon performs surgery with subjectively optimized ergonomics and mechanics. Therefore, no methodology exists to objectively optimize surgical techniques aimed to restore and optimize human vision. We envision that artificial intelligence (AI) can enhance surgeon performance and extract objective data about surgical techniques. In this work, we briefly review limitations in performing objective surgical research, the use of AI in health care, and the use of deep learning (DL) in surgery. Second, we propose and evaluate a method, using AI, to objectively tract surgical

instruments in the three-dimensional (3D) volume of the eye.

Objective Research in Surgery

Unbiased medical research in human populations is designed around patient randomization, blinding patients and clinicians to intervention, and standardizing interventions and clinical end points. Even in medicine, where pharmacological treatment protocols are more easily standardized than medical procedures, real-world medical outcomes do not match the outcomes achieved in large controlled clinical trials. This mismatch between real-world practice and clinical trials is largely because real-world practice does not ubiquitously recapitulate rigid clinical trial protocols.¹ Real-world experience in medicine does, however, provide valuable guidance to clinical care.^{2,3}

Within medicine, research in surgical fields is uniquely susceptible to bias⁴ because surgeons cannot be blinded to their intervention, and surgical interventions cannot be precisely standardized from one surgeon to the next. Like medical research that studies large cohorts of real-world patients, we believe that valuable information to guide surgical procedures is available for extraction from real-world surgical videos. The major limitation to studying real-world surgery is that no tool exists to data from surgical videos objectively. One method to study videographic data is to annotate videos and train computer vision systems to identify details from the video. A recent review⁵ of challenges facing annotation of surgery videos identified limitations in the skill level of graders (annotators), intergrader variability, and insufficient degrees of objectivity in the assessment of surgical performance. Therefore, there is a need for developing objective and reproducible methods to study surgical videos, and AI offers an avenue for objective surgery evaluation.

Review of AI in Medicine and Ophthalmology

Recently, AI has gained popularity in medical sciences and is a common topic in medical meetings.⁶ Common applications include matching patient symptoms to appropriate physician,⁷ diagnostic methods,⁸ drug discovery,⁹ and patient prognosis.¹⁰ DL,^{11,12} the modern rebranding of neural networks in which networks of simple interconnected neurons are trained to carry various tasks, is currently the dominant component of AI. In biomedical imaging, to analyze images, DL algorithms distinguish themselves by learning relevant features directly from training

data and use them for classification, regression, and other tasks and has been used in many biomedical imaging tasks^{13–19} in our group alone. We further have used DL to assist advanced two-photon imaging²⁰ and predict visible spectrum color images from infrared images.²¹ DL applications in ophthalmology are rapidly expanding across imaging modalities and diseases.^{8,22–29}

DL in Surgery

DL has been applied in surgical fields with great success and has gained popularity in diverse surgical specialties, including gynecology, digestive surgery, cardiology, and urology. Trained models may aid and improve surgical execution with high precision and address some of the technological challenges present in some surgical procedures. A meta-analysis study, combining results from 2289 papers used DL methods to evaluate surgeries, revealed that laparoscopy cholecystectomy was the most consistent surgery, achieving greater than 90% accuracy rates. They summarized that surgical procedures with simple workflow and well-defined automated phase of recognition can be performed with high accuracy, but more complex surgical procedures remained more challenging.³⁰ Additional reviews of DL applied to surgery highlight its potential to optimize preoperative planning and intraoperative performance,³¹ as well as the prediction of postoperative outcomes.³²

The application of AI to ocular surgery is relatively nascent, but remains an attractive option with significant potential. DL was reported for use in cataract surgery to track the pupil, identify the surgical phase, and provide tools to provide real-time visual feedback.^{33,34} The use of DL to aid surgery in the larger 3D volume of the posterior chamber has not been reported. Ophthalmology, for a number of reasons, is uniquely positioned to study surgical techniques from surgical videos. First, the eye is an enclosed space with little mobile anatomy (unlike the abdominal and thoracic cavities). Second, unlike the confines of the cranial vault, gastrointestinal tract, or heart, the eye is filled with a clear media. Third, visualizing anatomic landmarks during ophthalmic surgery is rarely obscured by bleeding. Finally, most ophthalmic surgery is performed using a high-quality surgical microscope that standardizes video recording parameters. Retina surgery is one of the most challenging surgical procedures owing to the limited access to a small surgical volume and sensitivity of tissues to incident light and mechanical forces. Therefore, tools to mitigate human errors and identify

optimal surgical techniques could optimize visual outcomes.

To evaluate the potential of AI/DL for surgical instrument tracking in eye surgery, we engineered an in vitro model eye to collect videographic data of surgical instrument movements in the eye. We then trained an AI system to analyze the videos of surgical instrumentation moving in vitro.

Methods

Computer-aided Design and Model Eye Fabrication

We designed a model eye using Fusion 360 (Autodesk, ver. 1.8) (Fig. 1a). We 3D printed a negative mold of the model eye using a polylactic acid filament (Hatchbox, Omaha, NE) on a Prussa i3 MK3S fused deposition modeling printer (Prussa, Prague, Czech Republic). We then poured a polydimethyl siloxane mixture into the negative mold and allowed to cure overnight at room temperature. We demolded the polydimethyl siloxane from the mold and epoxy glued it to a three-prong 3D printed ring that sits on top of another 3D printed staircase ring used to adjust the height of the model eye over the underlying retinal image. The staircase ring positioned the model eye at incremental distances from the underlying retinal image. We inserted three 23G vitreoretinal surgical cannulas (Alcon, Fort Worth, TX) into the model eye approximately 4 mm posterior to the viewing field.

Illumination Pattern of Surgical Instruments

To choreograph instrument movements in the model eye, an Arduino microcontroller was used to control a 12 light-emitting diode ring light (Adafruit, New York, NY) and a mini audio speaker to produce an audible tone between each choreographed segment. The ring light-emitting diode produced light inside the semitranslucent model eye by sequentially transilluminating each of the four quadrants. This practice yielded images of surgical instruments moving throughout the model eye with illumination from four different angles. Changes in light color from white to red, and sound emitted by the metronome were programmed to occur every 4 seconds to indicate when the surgeon should move the surgical instrument from one region and depth to the next region and depth. The assembled model eye with programmed illumination ring and microcontroller is shown in Figure 1b.

Video Recording

Videos were recorded using a Zeiss Lumera and Resight viewing system (Karl Zeiss Meditech, Dublin, CA) (Fig. 1c). Five 23G instruments were used: vitrector, forceps (in two variants: closed and opened), soft tip, loop, and laser probe (Alcon). Videos of instrument manipulation throughout the surgical field were recorded in nine regions and four depths (Fig. 1d) with instruments inserted through the right- and left-hand cannulas.

Video Processing and Data

After the videos were recorded, each video was trimmed to the exact same starting point in the illumination sequence. Then the videos were processed by a Python script that extracted frame images at a rate of 30 frames per second. Additionally, the script labeled each video frame image with the tool type, x - y region, zone depth, and tool laterality. There are five tool types, nine regions, four depth levels, and two laterality values. For the machine learning experiments described in this article, we used 34 videos, corresponding with a total of 103,437 frames, with an average of 3042 frames per video.

Models

After some experimentation, we designed and trained two models: a classification model and a close-up detection model (Table). The outputs of the classification model are purely categorical and comprise the type of instrument, the instrument's depth estimating the proximity to the retina surface (far, intermediate, near to the retina, and contact with the retina), the x - y location of the instrument's tip quantized to the nine regions outlined in Figure 1d, and the laterality. When the instrument is in the near or contact depth zones, the close-up discrimination model is used to identify the precise location of the instrument's tip. The outputs of the discrimination model include both categorical variables corresponding with a finer grained description of the instruments (e.g., open or closed) and continuous variables corresponding to x - y coordinates.

Classification Model

For the classification model, the data were split equally between training and validation sets. The classification model is used to classify each frame according to the following labels: x - y region, depth zone, tool type, and tool laterality. This model uses the ResNet-18 architecture.³⁵ It is trained for 100 epochs with a

learning rate of 0.01. The region prediction is used when the tool is in the intermediate and far depth zones and the tool $x-y$ location prediction is made for near and contact depth zones. Because ResNet-18 is a deep model, there is a tendency for the model to overfit. To overcome the overfitting, we added weight decay with a lambda of 0.0005, as well as data augmentation methods described elsewhere in this article.

Close-up Detection Model

Images of instruments moving throughout the model eye were randomly selected ($N = 251$ for each instrument). The instrument tip was labeled manually in seven classes (Fig. 2) and then used to train the detection model. The detection model is used when the tool is in the near or contact depth zones, where a precise location of the tip is desirable. We trained a YOLOv5-

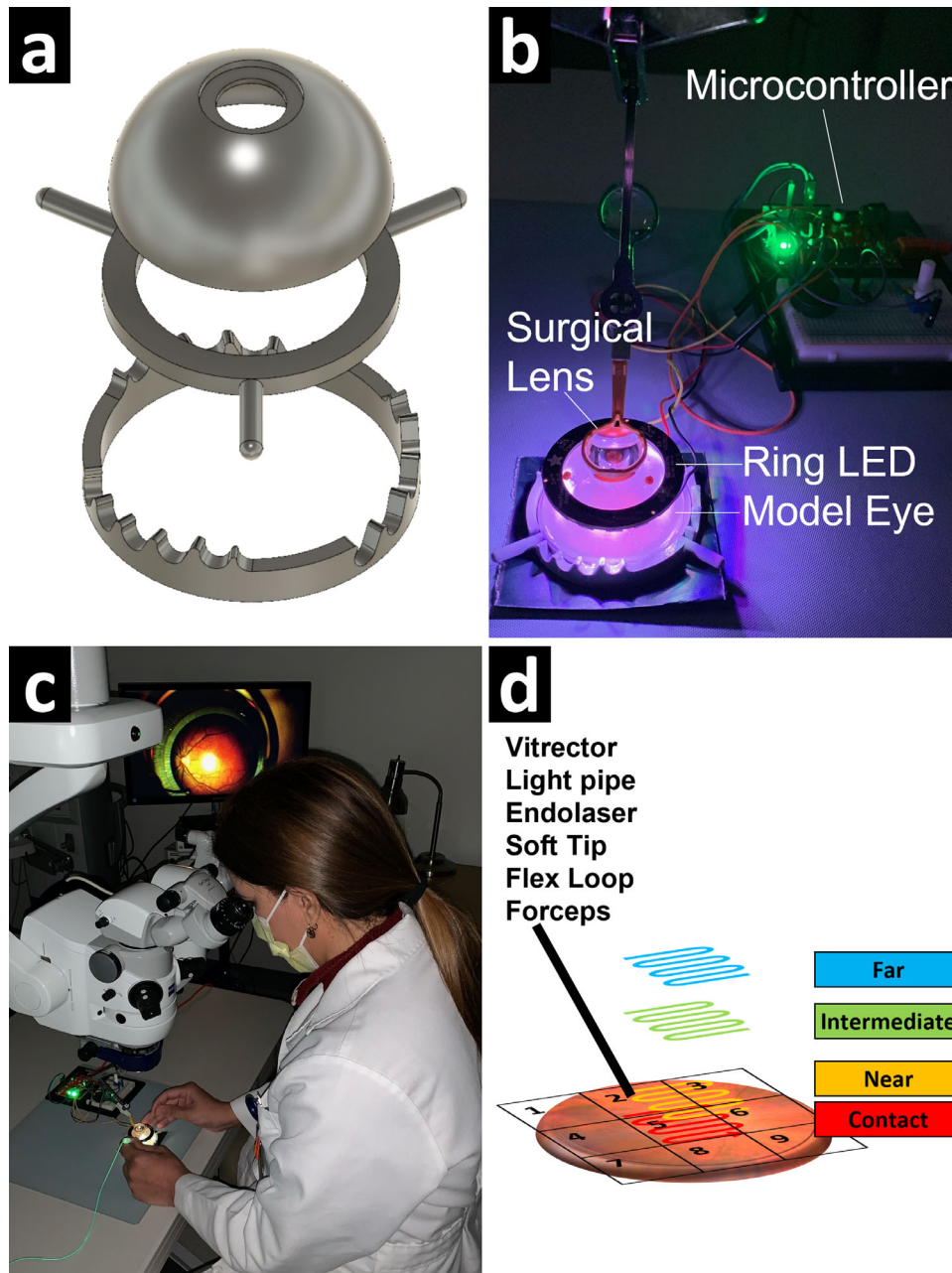
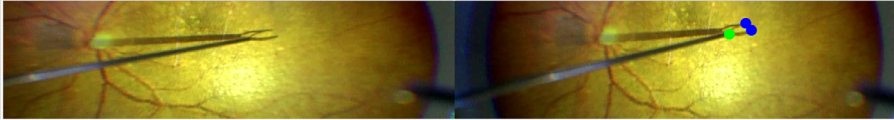


Figure 1. Model eye for data collection. (a) Computer-aided design and dimensions of model eye. (b) Three-dimensional printed model eye positioned beneath surgical microscope with ring light-emitting diode positioned on top of the model and Arduino microcontroller. (c) Vitreoretinal surgeon recording videos of surgical instrument movements inside the model eye. (d) Schema for moving six different surgical instrument types in nine $x-y$ regions and four depth zones.

Table. Summary of Datasets and Models Used to Predict Surgical Tool Characteristics in the Model Eye

| Model Type | Classification ^{ResNet-18} | Detection ^{YOLOv5} |
|---|---|---|
| Description of Data | | |
| Image Example |  | |
| Annotations Details | Descriptors of image frame: Zone Depth Region (X-Y) Instrument Type Instrument Laterality | Manual label of tip on image to provide precise x-y location |
| Data Size | N= 756,000 images total Testing: 50% Validation: 50% | N= 251 images per instrument Testing: 80% Validation: 20% |
| Data use to train each model | | |
| Instrument Depth | X | |
| Instrument X-Y region Depth (near and contact) | | X |
| Instrument X-Y region Depth (far and intermediate) | X | |
| Instrument Type Depth (near and contact) | X | |
| Instrument Type Depth (Far and Intermediate) | | X |

based object detection neural network³⁶ to accomplish the precise tip location detection task. YOLOv5 is a two-stage detector, which is composed of a Cross Stage Partial Network³⁷ as the backbone to initially extract the features, Path Aggregation Network³⁸ plus a Feature Pyramid Network³⁹ as the neck to aggregate the features, and a series of convolutional layers and fully connected layers as the head to output the detection results. Regularly, the backbone of YOLOv5 is pretrained on a large dataset, such as ImageNet,⁴⁰ to allow the network to preliminarily learn common features from general natural images. However, the in vitro dataset of simulated eye surgery images is a special dataset with images that may differ significantly from natural images. Therefore, we trained the YOLOv5-based object detection neural network without using pretrained weights.

Data Augmentation

To improve the generalization capabilities of the models and avoid overfitting, we apply data augmentation strategies to create additional relevant data. The augmentation transformations we applied include brightness jitter, contrast jitter, saturation jitter, hue jitter, random grayscale, random cropping, left-right

flip, translation, scale, and mosaic. Figure 3 provides a sample of comparison of raw images and their augmented counterparts. Data augmentation transformations were applied randomly to each training batch at training time.

Results

We designed and 3D printed a model eye in which we recorded videos using an ophthalmic surgery microscope of vitreoretinal surgery instruments moving throughout the 3D volume of the eye. Each frame from the video was isolated and labeled to specify the location of the surgical tool’s tip in 3D space, the type of surgical tool, and the laterality for inserting the surgical instrument into the model eye. A classification and a close-up detection model were used to predict characteristics for each surgical instruments moving in the model eye.

Classification

The dataset for the classification task consisted of 34 videos with an average of 3043 frames per video. We split the dataset frames between training and testing

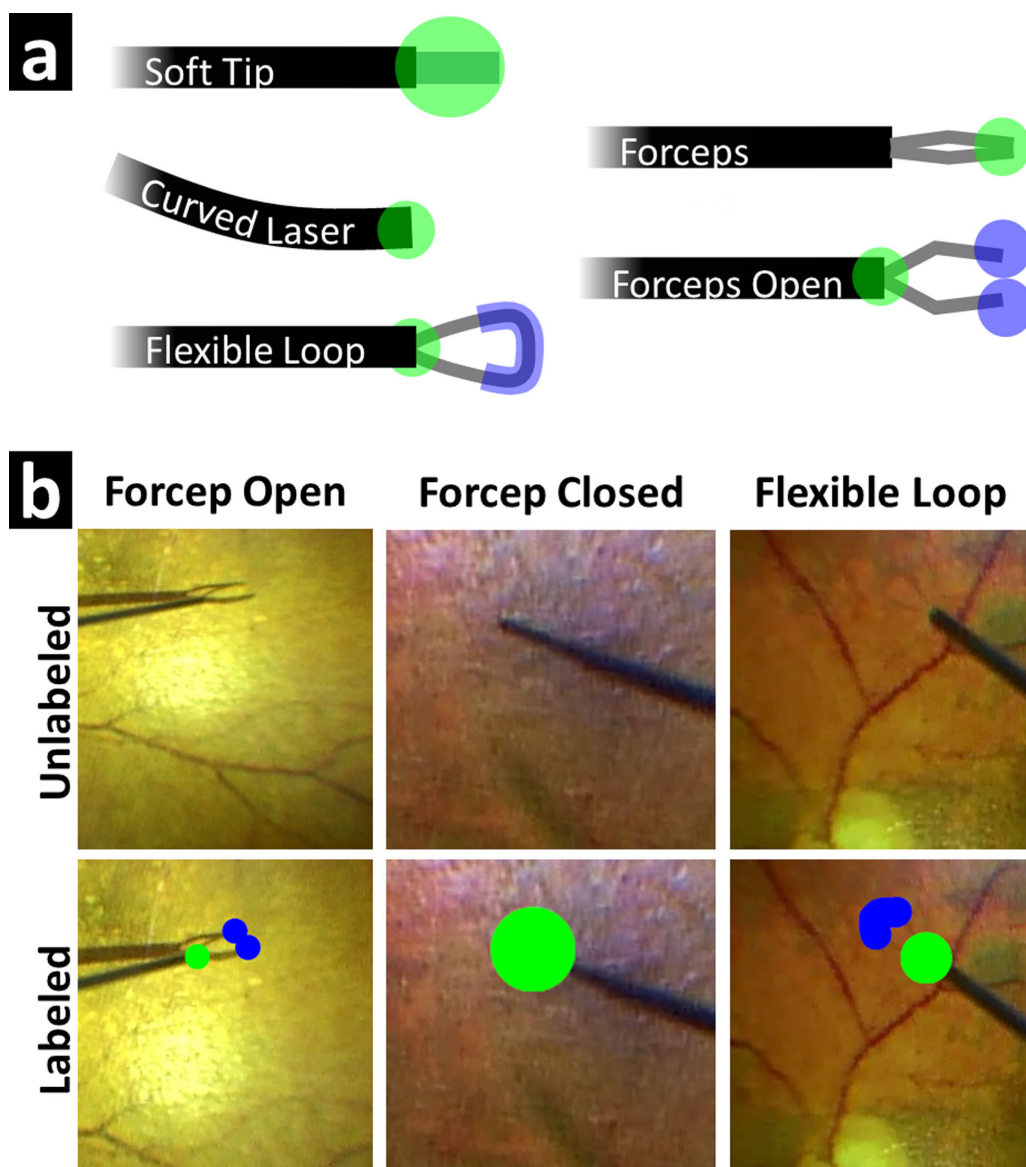


Figure 2. (a) Manual labeling strategy for different surgical tools. *Green* indicates rigid instrument tips. *Blue* indicates instrument tips when the instrument is open (forceps) or extended (loop). (b) Examples of manually labeled instrument tips used to train the AI detection model.

sets with a ratio of 50:50. The original images were resized into a size of 224×224 pixels, and then we applied various augmentation strategies while training the model. We used accuracy as the metric for measuring the models' performance. The accuracy of the classification model on the training set was 84% for x - y region, 97% for depth, 100% for instrument type, and 100% for laterality of insertion. The accuracy of the classification model on the validation dataset was 83% for x - y region, 96% for depth, 100% for instrument type, and 100% for laterality of insertion. Figure 4 shows the confusion matrices for the classification task. The model inference time is 0.9 ms to preprocess each input frame and output its classification results

using one Nvidia TitanX GPU. The trained model can perform real-time classification and process 1111 frames per second.

Close-up Detection

The confusion matrix for the precision in detecting the instrument's tip is presented in Figure 5a. The YOLOv5 model outputs a mean average precision of 79.3% when the threshold is set to 0.5, with a range extending from 62% to 94%. The average precision calculated for most instrument types is greater than 75%, except for the loop tip (62%), which is almost transparent. The model inference time is 14.9 ms to

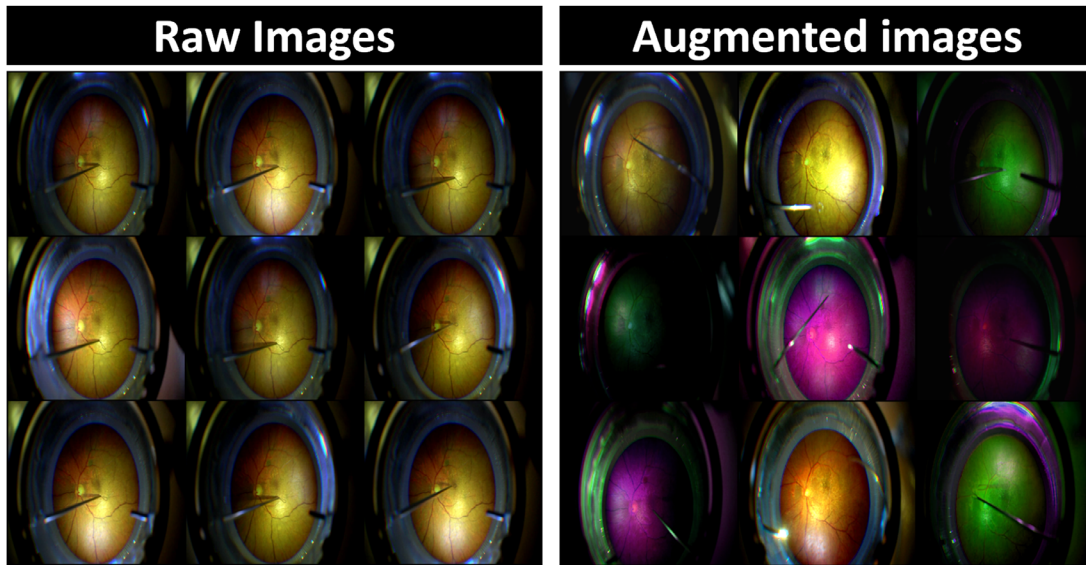


Figure 3. Side-by-side examples of raw and augmented images.

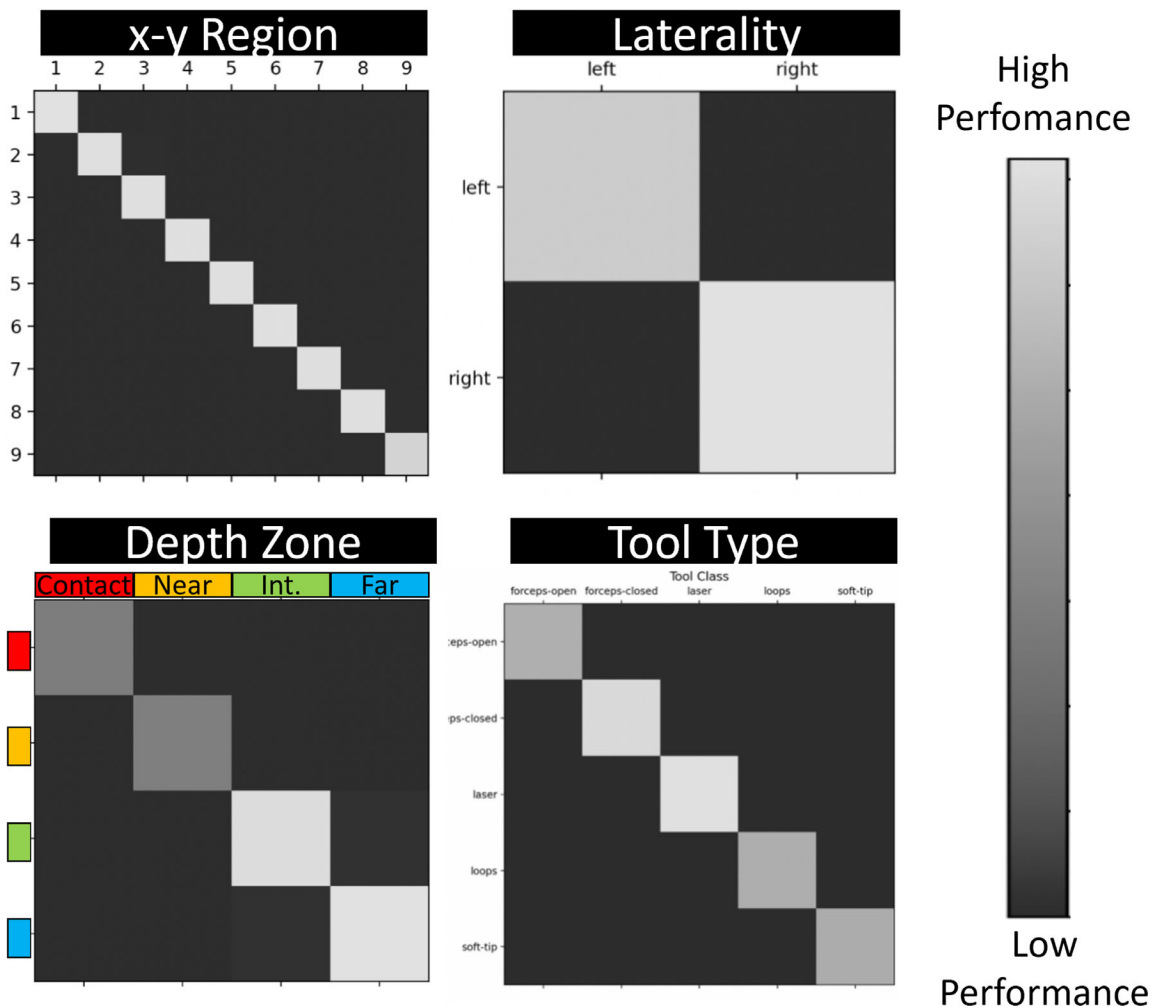


Figure 4. Confusion matrices for the classification model in predicting region, zone, tool type, and tool laterality. Depth zone is color coded as illustrated in Figure 1d. Results computed on validation samples.

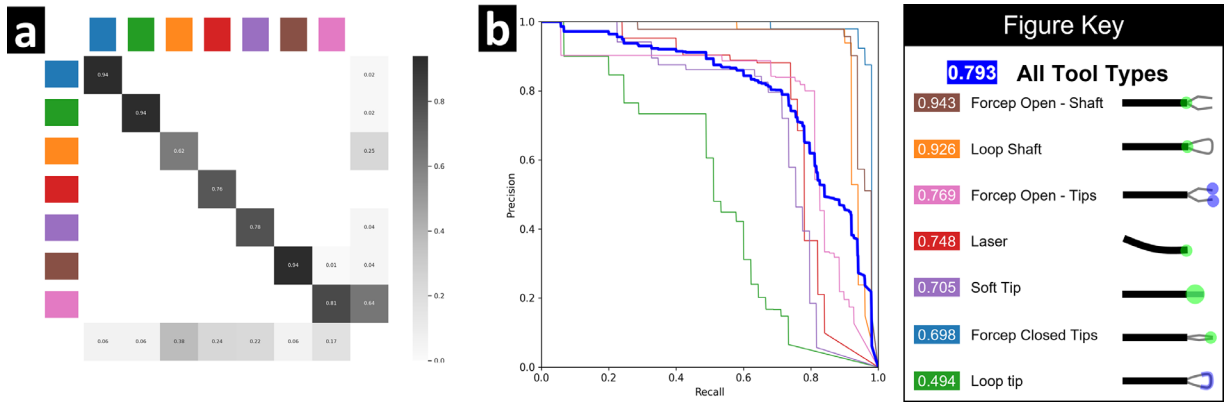


Figure 5. Close-up detection model’s performance for instrument tip detection. (a) Confusion matrix for instrument tip detection computed on validation samples. (b) Precision–recall curve for the instrument tip detection computed on validation samples. Color coding is according to instrument type and instrument status (when applicable), as shown in the figure key (far right). Confidence scores correspond with the white text in each colored box of the figure key.

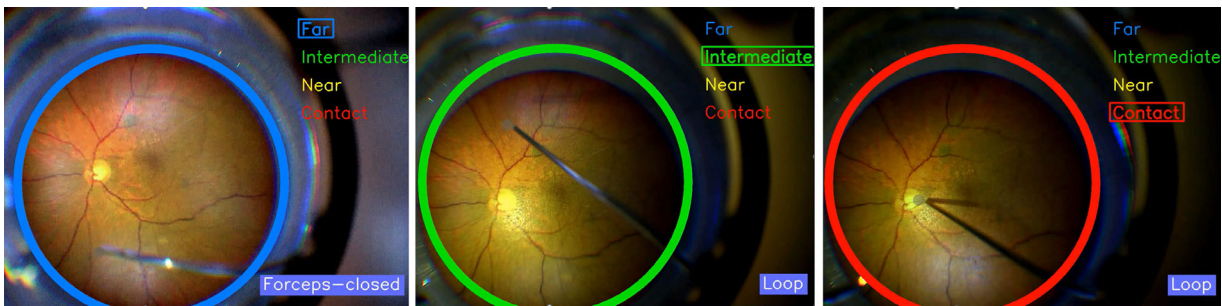


Figure 6. Examples of instrument detection with color-coded depth and a gray circular shadow overlaying the instrument tip.

preprocess each input frame and output its detection results using one Nvidia TitanX GPU. The trained model can perform real-time detection and process 67 frames per second. The precision–recall results for detection are presented in Figure 5b, where the confidence score of each detected object is provided adjacent to the legend labels. A confidence score of 0 means the least confidence and 1 means the most confidence. In short, the trained close-up detection model achieves precise and real-time detection of all given instruments.

Real-time Visual Feedback

To demonstrate one possible real-time rendering to display data about surgical tool manipulations in a 3D space, we evaluated a series of videos and predicted the location and type of surgical tools (Fig. 6). Surgical tool type is displayed in the lower right corner, distance to the retinal surface is outlined in the top right corner, and the rim of the surgical view is highlighted with a color-coded ring indicating predicted tool depth. Finally, a faint gray shadow identifies the instrument

tip location. For all instruments the tip detection was worst for the loop as seen in Figure 6. A video demonstrating real-time model predictions with instrument location feedback rendered in the same way as Figure 6 (Supplementary Movie S1). Regardless of location of illumination direction, the models perform consistently to identify instrument 3D location and instrument type.

Discussion

We sought to evaluate the feasibility for trained AI models to detect surgical instrument characteristics as they moved within the 3D volume of the eye. We developed an artificial eye within which to record videos of surgical instruments moving throughout 3D space (Figs. 1a–d). Using data procured in the artificial eye, we demonstrated that AI can accurately identify surgical instrument type and location moving throughout a simulated 3D space. The AI renders a real-time analysis to enhance visualizing in the artificial surgical field

of view (Fig. 6) and performs very well at predicting multiple instrument characteristics (Figs. 4 and 5).

The classification model accuracy was greatest for predicting tool type and tool laterality and slightly lower for predicting region and depth. Overall, the classification model generalizes well to the data in the validation dataset.

The close-up detection model average precision for instrument tip detection was greater than 75%. The loop tip demonstrated the lowest precision, and this is most likely because the thin loop wire is significantly thinner than most retinal vessels and is, therefore, poorly resolved by the surgical video microscope recording system. The model sometimes predicted the loop tip to be part of the background image of the retina.

When the tools were near or in contact with the retinal surface, we accepted a single output from the prediction model. However, because both models can distinguish instrument depths, future implementation of models predicting instruments in the near and contact depths may compare the confidence of both models and use this information to produce more robust depth inferences. Although the results of this study are promising, our trained model is neither suitable for nor capable of evaluating real-world surgical videos; however, a similar methodology and trained model based on real-world surgical data should be developed. Although our training and testing datasets are influenced by some crossover bias because they are composed of unique and separate frames from the same movie, we believe that this pilot study demonstrates feasibility to train neural networks to detect multiple instrument parameters in a simulated 3D intraocular volume.

Mechanical forces from surgical instruments used in retina surgery are below the tactile perception threshold of the surgeon. Therefore, vitreoretinal surgery is performed based on visual feedback, rather than tactile perception.⁴¹ The only tool to study forces on the retina is a force-sensing micropick that detected forces below tactile sensation and quantified the forces generated during normal maneuvers from those that may cause a surgical complication.⁴² However, specialized force-sensing instruments are not practical for implementation during retina surgery. Because humans rely on visual feedback to infer forces in retina surgery, it may be possible to determine surgical forces using video image analysis to determine rates and directions of instrument movements.

Important contributions using robotic tools in ocular surgery suggest the potential to marry AI with robotic surgical tools in the future. Robots have assisted corneal laceration repair in a porcine eye,⁴³

removal of an epiretinal membrane or inner limiting membrane in a human eye,⁴⁴ and even image-guided robot-assisted 3D navigation of a microsurgical instrument.⁴⁵ However, the only report of DL in vitreoretinal surgery came from a conference report where a pipeline was designed to estimate the relative distance of the vitreoretinal instrument tips from the retina surface using Convolutional Neural Networks and stereo vision recording microscopes with satisfactory performance.⁴⁶ Most surgical microscopes are not equipped with stereo digital video capabilities. However, the stereovision approach would be valuable to train a model to use only one of two stereo channels and make inferences, as described in this work. The new capabilities created by AI in surgery have broad potential applications. Surgical safety may be enhanced in real time by the detection of surgical instrument proximity to the optic nerve and retinal surface. AI trained to track surgical phases and instruments used could be used to automate transcription of an operative report. An additional significant utility of AI extracting data from surgical videos may be as an aid in conducting objective research in surgical techniques across surgical videos from thousands of surgeons.

Conclusions

AI is increasingly being used in medical image interpretation. Now AI is moving into more dynamic imaging modalities including surgical videography. The methods and results in this work demonstrate the feasibility for trained models to accurately detect and classify microsurgical instruments moving within the gross and microanatomic confines of the eye. We demonstrate the strong capabilities for AI to make multiple simultaneous predictions in simulated artificial surgical environment. This data supports our long-term goal to develop software to extract data from real-world surgery videos and use it to investigate and improve intraocular surgery. To transition this technology to real world surgery, creative tools are needed to annotate large volumes of surgical imaging.

Acknowledgments

The authors thank Baruch D. Kuppermann for providing the opportunity to initiate this work, availing the surgical wetlab to collect data and supporting progress at every stage of this project. We thank Anjali Herekar for assistance in recording surgical videos.

Supported by a Research to Prevent Blindness unrestricted grant to UC Irvine Department of Ophthalmology, NIBIB of the National Institute of Health (NIH) – R01EB026705, Arnold and Mabel Beckman Foundation, BrightFocus Foundation

Disclosure: **P.F. Baldi**, None; **S. Abdelkarim**, None; **J. Liu**, None; **J.K. To**, None; **M.D. Ibarra**, None; **A.W. Browne**, None

* PFB, SA, JL, and JKT contributed equally to this article.

References

- Ciulla TA, Huang F, Westby K, Williams DF, Zaveri S, Patel SC. Real-world outcomes of anti-vascular endothelial growth factor therapy in neovascular age-related macular degeneration in the United States. *Ophthalmol Retina*. 2018;2(7):645–653, doi:10.1016/j.oret.2018.01.006.
- Cheema MR, DaCosta J, Talks J. Ten-year real-world outcomes of anti-vascular endothelial growth factor therapy in neovascular age-related macular degeneration. *Clin Ophthalmol*. 2021;15:279–287, doi:10.2147/OPTH.S269162.
- Brynskov T, Munch IC, Larsen TM, Erngaard L, Sorensen TL. Real-world 10-year experiences with intravitreal treatment with ranibizumab and aflibercept for neovascular age-related macular degeneration. *Acta Ophthalmol*. 2020;98(2):132–138, doi:10.1111/aos.14183.
- Paradis C. Bias in surgical research. *Ann Surg*. 2008;248(2):180–188, doi:10.1097/SLA.0b013e318176bf4b.
- Ward TM, Fer DM, Ban Y, Rosman G, Meireles OR, Hashimoto DA. Challenges in surgical video annotation. *Comput Assist Surg (Abingdon)*. 2021;26(1):58–68, doi:10.1080/24699322.2021.1937320.
- Beam AL, Kohane IS. Translating artificial intelligence into clinical care. *JAMA*. 2016;316(22):2368–2369, doi:10.1001/jama.2016.17217.
- Güneş ED, Yaman H, Çekyay B, Verter V. Matching patient and physician preferences in designing a primary care facility network. *J Oper Res Soc*. 2014;65(4):483–496, doi:10.1057/jors.2012.71.
- Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*. 2016;316(22):2402–2410, doi:10.1001/jama.2016.17216.
- Jing Y, Bian Y, Hu Z, Wang L, Xie XQ. Deep learning for drug design: an artificial intelligence paradigm for drug discovery in the big data era. *AAPS J*. 2018;20(3):58, doi:10.1208/s12248-018-0210-0.
- Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542(7639):115–118, doi:10.1038/nature21056.
- Schmidhuber J. Deep learning in neural networks: an overview. *Neural Netw*. 2015;61:85–117, doi:10.1016/j.neunet.2014.09.003.
- Baldi P. *Deep Learning in Science*. Cambridge, UK: Cambridge University Press; 2021.
- Urban G, Tripathi P, Alkayali T, et al. Deep learning localizes and identifies polyps in real time with 96% accuracy in screening colonoscopy. *Gastroenterology*. 2018;155(4):1069–1078.e8, doi:10.1053/j.gastro.2018.06.037.
- Chang P, Grinband J, Weinberg BD, et al. Deep-learning convolutional neural networks accurately classify genetic mutations in gliomas. *AJNR Am J Neuroradiol*. 2018;39(7):1201–1207, doi:10.3174/ajnr.A5667.
- Wang J, Ding H, Bidgoli FA, et al. Detecting cardiovascular disease from mammograms with deep learning. *IEEE Trans Med Imaging*. 2017;36(5):1172–1181, doi:10.1109/TMI.2017.2655486.
- Wang J, Fang Z, Lang N, Yuan H, Su MY, Baldi P. A multi-resolution approach for spinal metastasis detection using deep Siamese neural networks. *Comput Biol Med*. 2017;84:137–146, doi:10.1016/j.combiomed.2017.03.024.
- Urban G, Feil N, Csuka E, et al. Combining deep learning with optical coherence tomography imaging to determine scalp hair and follicle counts. *Lasers Surg Med*. 2021;53:171–178, doi:10.1002/lsm.23324.
- Baldi P, Chauvin Y. Neural networks for fingerprint recognition. *Neural Comput*. 1993;5(3):402–418.
- Urban G, Bache KM, Phan D, et al. Deep learning for drug discovery and cancer research: automated analysis of vascularization images. *IEEE/ACM Trans Comput Biol Bioinform*. 2019;16(3):1029–1035, doi:10.1109/TCBB.2018.2841396.
- McAleer S, Fast A, Xue Y, et al. Deep learning-assisted multiphoton microscopy to reduce light exposure and expedite imaging in tissues with high and low light sensitivity. *Transl Vis Sci Technol*. 2021;10(12):30, doi:10.1167/tvst.10.12.30.
- Browne A, Deyneka E, Ceccarelli F, et al. Deep learning to enable color vision in the dark. *PLoS One*. 2022;17(4):e0265185.

22. Parmar C, Barry JD, Hosny A, Quackenbush J, Aerts H. Data analysis strategies in medical imaging. *Clin Cancer Res*. 2018;24(15):3492–3499, doi:10.1158/1078-0432.CCR-18-0385.
23. Asaoka R, Tanito M, Shibata N, et al. Validation of a deep learning model to screen for glaucoma using images from different fundus cameras and data augmentation. *Ophthalmol Glaucoma*. 2019;2(4):224–231, doi:10.1016/j.ogla.2019.03.008.
24. Phene S, Dunn RC, Hammel N, et al. Deep learning and glaucoma specialists: the relative importance of optic disc features to predict glaucoma referral in fundus photographs. *Ophthalmology*. 2019;126(12):1627–1639, doi:10.1016/j.ophtha.2019.07.024.
25. Abràmoff MD, Lavin PT, Birch M, Shah N, Folk JC. Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. *NPJ Digit Med*. 2018;1:39, doi:10.1038/s41746-018-0040-6.
26. Li Z, Guo C, Nie D, et al. Deep learning for detecting retinal detachment and discerning macular status using ultra-widefield fundus images. *Commun Biol*. 2020;3(1):15, doi:10.1038/s42003-019-0730-x.
27. Lachance A, Godbout M, Antaki F, et al. Predicting visual improvement after macular hole surgery: a combined model using deep learning and clinical features. *Transl Vis Sci Technol*. 2022;11(4):6, doi:10.1167/tvst.11.4.6.
28. Poplin R, Varadarajan AV, Blumer K, et al. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nat Biomed Eng*. 2018;2(3):158–164, doi:10.1038/s41551-018-0195-0.
29. Zhu Z, Shi D, Guankai P, et al. Retinal age gap as a predictive biomarker for mortality risk. *Br J Ophthalmol*. 2022;2(3):158–164, doi:10.1136/bjophthalmol-2021-319807.
30. Garrow CR, Kowalewski KF, Li L, et al. Machine learning for surgical phase recognition: a systematic review. *Ann Surg*. 2021;273(4):684–693, doi:10.1097/sla.0000000000004425.
31. Morris MX, Rajesh A, Asaad M, Hassan A, Saadoun R, Butler CE. Deep learning applications in surgery: Current uses and future directions. *Am Surg*. 2023;89(1):36–42, doi:10.1177/00031348221101490.
32. Lee CK, Hofer I, Gabel E, Baldi P, Cannesson M. Development and validation of a deep neural network model for prediction of postoperative in-hospital mortality. *Anesthesiology*. 2018;129(4):649–662, doi:10.1097/ALN.0000000000002186.
33. Garcia Nespolo R, Yi D, Cole E, Valikodath N, Luciano C, Leiderman YI. Evaluation of artificial intelligence-based intraoperative guidance tools for phacoemulsification cataract surgery. *JAMA Ophthalmol*. 2022;140(2):170–177, doi:10.1001/jamaophthalmol.2021.5742.
34. Yu F, Silva Croso G, Kim TS, et al. Assessment of automated identification of phases in videos of cataract surgery using machine learning and deep learning techniques. *JAMA Netw Open*. 2019;2(4):e191860, doi:10.1001/jamanetworkopen.2019.1860.
35. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Presented at: Proceedings of the IEEE conference on computer vision and pattern recognition; June 27–20, 2016; Las Vegas, NV.
36. Jocher G, Stoken A, Borovec J, et al. NanoCode012, Christopher STAN, Changyu L, ultralytics/yolov5: v3.1 - Bug fixes and performance improvements. version v3.1. Zenodo; 2020. Available at: https://zenodo.org/record/4154370#.Y7cIVajX_9M <http://dx.doi.org/10.5281/zenodo.4154370>.
37. Wang C-Y, Liao H-YM, Wu Y-H, Chen P-Y, Hsieh J-W, Yeh IH. CSPNet: A new backbone that can enhance learning capability of CNN. Presented at: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops; June 14–19, 2020; Virtual.
38. Liu S, Qi L, Qin H, Shi J, Jia J. Path aggregation network for instance segmentation. Presented at: Proceedings of the IEEE conference on computer vision and pattern recognition; June 18–22, 2018; Salt Lake City, Utah.
39. Lin T-Y, Dollár P, Girshick R, He K, Hariharan B, Belongie S. Feature pyramid networks for object detection. Presented at: Proceedings of the IEEE conference on computer vision and pattern recognition; July 21–26, 2017; Honolulu, Hawaii.
40. Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L. Imagenet: A large-scale hierarchical image database. Presented at: 2009 IEEE conference on computer vision and pattern recognition; June 20–25, 2009; Miami, Florida.
41. Jagtap AS, Riviere CN. Applied force during vitreoretinal microsurgery with handheld instruments. *Conf Proc IEEE Eng Med Biol Soc*. 2004;2004:2771–2773, doi:10.1109/iembs.2004.1403792.
42. Sunshine S, Balicki M, He X, et al. A force-sensing microsurgical instrument that detects forces below human tactile sensation. *Retina*. 2013;33(1):200–206, doi:10.1097/IAE.0b013e3182625d2b.

43. Tsirbas A, Mango C, Dutson E. Robotic ocular surgery. *Br J Ophthalmol*. 2007;91(1):18–21, doi:[10.1136/bjo.2006.096040](https://doi.org/10.1136/bjo.2006.096040).
44. Edwards TL, Xue K, Meenink HCM, et al. First-in-human study of the safety and viability of intraocular robotic surgery. *Nat Biomed Eng*. 2018;2:649–656, doi:[10.1038/s41551-018-0248-4](https://doi.org/10.1038/s41551-018-0248-4).
45. Zhou M, Wu J, Ebrahimi A, et al. Spotlight-based 3D instrument guidance for autonomous task in robot-assisted retinal surgery. *IEEE Robot Autom Lett*. 2021;6(4):7750–7757, doi:[10.1109/lra.2021.3100937](https://doi.org/10.1109/lra.2021.3100937).
46. Fatta MD. *Surgical instrument tracking for intraoperative vitrectomy guidance using deep learning*

and stereo vision. Chicago: University of Illinois at Chicago; 2020.

Supplementary Material

Supplementary Movie S1. Movie of different surgical tools moved throughout 3D volume of artificial eye. The direction of illumination is observed to change continuously while the instrument tip is highlighted with a gray circle and the instrument depth is detected and rendered as a color-coded circle around the surgical field of view.