IMMERSIVE REALITY FOR INTUITIVE SURGERY

CS4605/7470

Project Iris: Report

Atharva Madiwale
Georgia Institute of Technology
atharvamadiwale@gatech.edu

Luke Chen
College of Engineering
Georgia Institute of Technology
lukechen@gatech.edu

Feyi Ogunsanya
College of Computing
Georgia Institute of Technology
fogunsanya3@gatech.edu

Ahmed Klasra
College of Computing
Georgia Institute of Technology
ahmedraufk@gatech.edu

ABSTRACT

This research paper introduces Project Iris, an innovative solution for otolaryngology surgery to be performed with greater technological advancement. An AR headset with surgical magnification improvements, ML segmentation, and a mobile application processing unit was produced from Project Iris. This paper details the device's technological features, gesture control, instance segmentation, and hardware capabilities to help surgeons improve their technical outcomes by reducing physical burden, operation times, and healthcare costs. The four main components of the project include gesture control for magnification, a machine-learning model for the segmentation of anatomical structures, optical flow stabilization, and an Android application for OR integration and data collection. Project Iris also has the potential to be expanded to other surgical fields to provide an innovative solution in surgical practices.

Keywords

Surgical AR; Computer Vision; gesture recognition; Instance segmentation; YOLOV8; otolaryngology

1. INTRODUCTION

In the surgical community, across many subspecialties, there has been increased interest in using head-up displays to enhance visualization and the surgeon's field of view. It has shown numerous benefits thus far in clinical and user studies such as less physical burden and fatigue, less visual scanning, and more time looking at the patients as opposed to an external computer/workstation [1]. It also has shown potential in decreasing operation times, thus enabling more surgeries, and reducing healthcare costs [1].

In the domain of otolaryngology, intricate procedures often involve delicate and critical anatomical structures that demand the utmost precision and comprehensive visualization. Such complexity necessitates innovative solutions to provide surgeons with the necessary tools and information to optimize surgical outcomes.

Iris is a novel surgical head-worn device with built-in computer vision models and gesture recognition that aims to revolutionize surgical procedures by providing enhanced visualization and support to surgeons, thus enhancing surgical outcomes. The project is divided into four sections, each focusing on a crucial aspect of the overall goal.

1) Anatomy Segmentation Using Deep Learning

2) Magnification Control and Visualization

3) Quantitative Data Collection

4) Optical Flow Stabilization

We aim to develop a proof-of-concept device for a surgical based on otolaryngology anatomy that in the future we aim to expand to other surgical domains.

2. BACKGROUND

Every day, thousands of otolaryngology procedures are performed worldwide. Yet, many of these surgeries rely on time-consuming methods such as surgical loupes and surgical microscopes. In time-sensitive procedures, every second counts, and the limitations of traditional equipment can hinder progress. But what if there was a better way? Introducing Project Iris - the efficient solution that's changing the game. Revolutionize otolaryngology with Iris: Immersive Reality for Intuitive Surgery.

Otolaryngology involves the precise surgery of ear, nose, and throat surgery. When treating throat illnesses specifically, such as vocal cord paralysis, it is imperative to be able to zoom into the anatomical regions to enhance the precision of resections.

The current setup for open surgeries involves switching out surgical loupes based on the desired magnification which can increase surgical times by upwards of 30 minutes. Additionally, for endoscopic procedures, large surgical microscopes further enhance visualization but this bulky equipment poses a constraint in the spatially limited operating room.

Project Iris aims to revolutionize surgery by building an AR surgical headset that performs magnification of the surgical field while avoiding the time, space, and cost barriers. At our current stage, we are working on touchless gesture recognition for magnification to accommodate the sterile surgical field, an Android app to improve integration into the OR for parameter configuration, and an intelligent zoom system to accompany manual zoom control.

3. AR HEADSET INTEGRATION FOR ENHANCED MAGNIFICATION

3.1 Gesture Control

Having enhanced magnification integrated into the AR headset empowers surgeons by offering adjustable magnification levels of up to 20.0x for the surgical field. Surgeons have complete control over the magnification level, which can be adjusted through various touch-based and non-touch-based interaction methods. The current surgical setup employs surgical loupes that are used for magnification purposes. These physical lenses need to be swapped every time the magnification of the surgical field needs to be adjusted. This process is tiring and time-consuming, especially in time-sensitive cases where being fast and efficient is imperative. In the next steps, “conceiving a future era of human-computer interaction with the implementations of 3D application where the user may be able to move and rotate objects simply by moving and rotating his hand - all without the help of any input device” [2]. Iris utilizes high-resolution cameras instead of physical lenses with surgeons having intuitive control on the magnification level. This eliminates the need for manual lens swapping as seen with traditional surgical glasses, streamlining the process, and saving valuable time. Surgeons can easily adapt to the magnification levels in real time, allowing them to cater to the specific requirements of each procedure.

Due to the constraints of surgery, research for gesture control included both touch-based and non-touch-based gestures. The primary reason was to maintain sterility within the surgery and operating room. The application incorporates intuitive controls, such as distance sensors (non-touch-based), sliding potentiometers (touch-based), buttons (touch-based), and rotary bezels (touch-based interaction), facilitating seamless adjustment of the magnification level and enhancing the overall visualization experience. Having multiple avenues for magnification control at the moment allows the surgeons to continue trying different gestures to see what is the most intuitive.

3.2 Enhanced Magnification Results

Project Iris incorporates an Epson Moverio BT-40 AR headset for surgical field visualization. This headset houses numerous built-in sensors that capture vital user data, including head tilts, gestures, and taps. These gestures can even trigger magnification actions, combining tactile inputs with improved visual feedback. Utilizing the headset's sensors, our mobile app gains direct access to these inputs, reducing the need for external sensors and hardware and solving gesture problems related to power by creating a “gesture-recognition system that can operate on a range of computing devices” [3]. This streamlined approach maintains system integrity and ensures an efficient user experience.

Currently, the best feedback gestures have been the 2 buttons, one to zoom in and one to zoom out. In addition to the Epson Moverio BT-40 AR headset, the application introduces a wireless neck brace with two tactile buttons that aid in the magnification process for the camera. The neck brace contains a Wi-Fi module that communicates with the onboard computer to signal the camera to zoom in or out. The neck brace will maintain sterility by being encompassed since the buttons will still be able to be pressed. Additionally, this strengthens the requirements of surgery where the surgeon is not allowed to raise their hands over their head.

4. INSTANCE SEGMENTATION

4.1 YOLOV8 Model

The use of artificial intelligence (AI) in medicine and surgery is expanding rapidly, with a focus on computer vision algorithms specifically of great use in segmentation. In the field of otolaryngology, previous papers have explored the segmentation of anatomical structures for its preoperative planning and research use. One such paper segments vocal fold tissue in high-speed video sequences using CNNs with Long Short-Term Memory (LSTM) cells for the temporal context [4]. The state-of-the-art YOLO object detection and instance segmentation model involves the precise identification and segmentation of targeted sections and anatomical structures within images or videos [5]. The YOLOV8 model is the newest version, combining both speed and accuracy. It has demonstrated high performance across different datasets and object categories, due to its training on large-scale datasets like COCO, containing a wide variety of object classes and diverse visual characteristics[6]. It has shown to be a valuable tool in computer vision applications, and to our knowledge, no study has applied the model in the use case of vocal fold segmentation, which we aimed to do in IRIS.

We used the pre-trained YOLOV8 segmentation model and training on a training dataset consisting of 6,228 images of the vocal fold surgical field. Augmentations (gaussian noise, occasions, motion blur, zoom, rotations) were applied using Roboflow, a computer vision developer framework, to increase the variability and generalizability of our model.

A random search of the following hyperparameters was used to optimize our model using the Ray Tune Python library:

- Batch size: Batches of size 4, 8, and 16 were considered.

- Dropout: Dropout values in a continuous space between 0.1 and 0.5 were considered.
- Weight decay: Weight decay was considered in the log space between 1e-4 and 1e-2.
- Initial learning rate: Initial learning rate was considered in log space between 1e-5 and 1e-2
- Final Learning rate: Final learning rate was considered in log space between 1e-4 and 1e-2

4.2 Training Results

During training, bordering box loss, segmentation loss, and object classification loss were selected. To prevent overfitting, an early stopping technique was implemented. The training was performed in parallel on a machine with a GPU and 16 CPU cores.

Our best-performing model was selected based on the most optimal loss values as well as a qualitative assessment of the model's performance on a test dataset. A summary of accuracy plots can be depicted in Figure 1. The selected hyperparameters are depicted in Table 1.

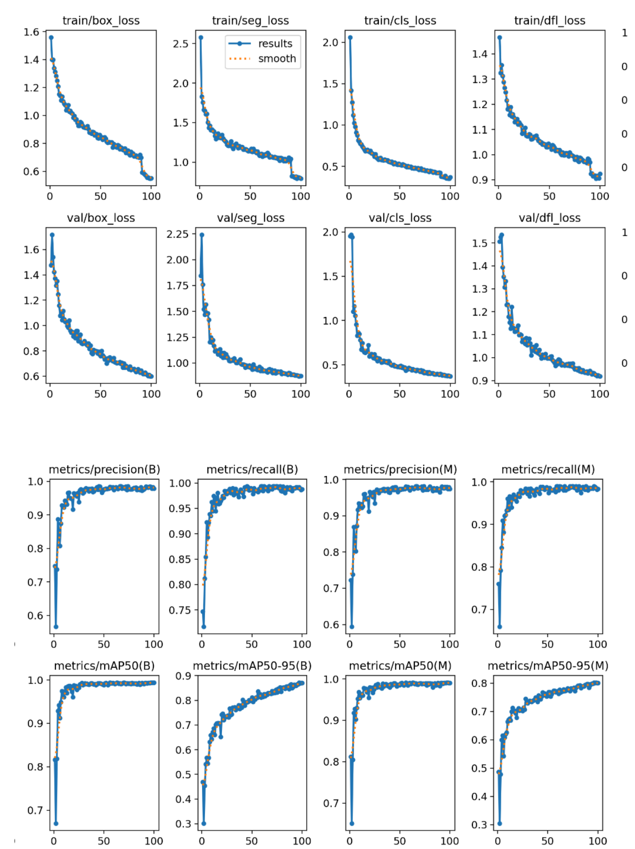

Figure 1. Metrics for the YOLOv8 Instance Segmentation model trained to segment Vocal Folds

Batch Size	Dropout	Weight Decay	Initial Learning Rate	Final Learning Rate
16	0.2	0.0005	0.01	0.01

Table 1. Summary of selected hyperparameters in best performing YOLOv8 model.

A visual evaluation of the model's performance can be seen in Figure 2. Through this process, we have achieved remarkable accuracy and robust performance across multiple images and videos. The segmentation of vocal folds using our application significantly enhances the visual perception of surgeons, enabling them to effortlessly identify and visualize the vocal folds even under challenging conditions. This functionality proves especially beneficial during training sessions for new surgeons and doctors, as it offers clear and precise visualization, aiding in their education and skill development.

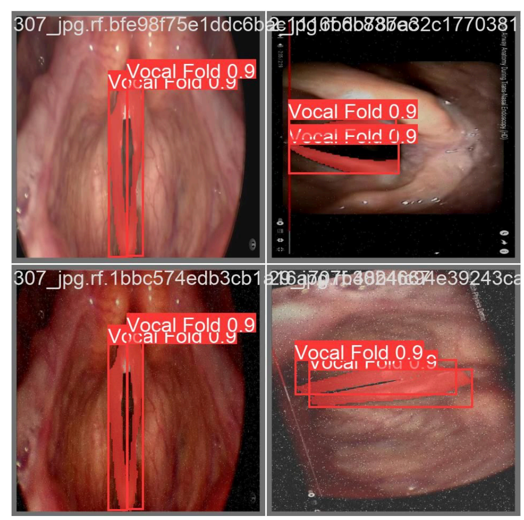

Figure 2. Instance Segmentation of Vocal Folds using best-performing YOLOv8 Model

5. ANDROID MOBILE APPLICATION INTEGRATION AND PROTOTYPE

One of the major motivations of our project is to make sure that we create a standalone device that not only doesn't require any cloud computing for processing, but is also highly accessible to people with varying budgets and needs. The Epson Moverio BT-40's compatibility with android allows us to tap into the sensors for gesture recognition and make the device portable. However it also leads to other issues such as lack of computing power for the required machine learning tasks and the intricate requirements for the Epson Moverio BT-40 to be compatible with any android device. For the purpose of development and testing, we used a small board computer called the ROK-3588s which has the required computing power, comes in a suitable size and weight and is cheaper than most smartphones compatible with the Epson Moverio BT-40. The current iteration of the application only includes the zoom-in and zoom-out function by connecting to the module on the neck brace, and displays footage from a USB camera attached to the device that is placed on the prototype. Future priorities and implementations include displaying the YOLOV8 segmentation on the device using the onboard NPU, and then iterating on the app to make it functional on consumer phones with the prototype.

5.B PROTOTYPE

Taking into consideration the ergonomic aspects of surgeons, the prototype and Proof of Concept (POC) has been designed in a way that surgeons can visualize the surgical field while maintaining a straight posture. An adjustable pitch camera bracket has been placed on top of the headset that allows the surgeons to adjust the angle of the camera to better suit their preferences. This adjustable camera bracket allows surgeons to visualize the surgical field in front of them.

Fig 3: HWD being used to perform a simple suturing application

The prototype was tested during a user study conducted at Grady Memorial Hospital / Emory University School of Medicine during a surgical training session. The user study focussed on studying the integration of HWD in surgeries and comparing their usability and integration with surgical microscopes and loupes. The user study was divided in 3 stages that tested the surgeons ability to perform a simple suturing application while wearing the HWD and the surgical microscope separately. Overall, approximately 20 surgeons tested our headset and 13 independent user studies were conducted. The feedback was extremely positive and allowed us to understand the different aspects we need to focus on for the next stage of the project.

6. CONCLUSION

With our work, we were able to prove that not only is it possible to replace surgical loupes with AR technology in a real surgical environment, but that this technology can be affordable and applicable to a multitude of fields of medicine. One of the major challenges of this project included some of the design aspects such as what gestures to use for the zoom function, which became a bigger problem than expected due to the strict rules and procedures of an operating room. This was a problem which required alot of feedback from the end-users to solve, and is still open-ended should we come up with a better solution than the tactile buttons, however we were quite satisfied with the feedback of our current implementation. Project Iris aims to make this device portable, affordable and be able to perform complex machine learning tasks. There are not many devices compatible that fulfill all these requirements while also being compatible with the Epson Moverio BT-40, making development quite difficult. This includes the implementation of the YOLOV8 model on such a device. We were able to work around these issues, and are confident that we have shown this technology is feasible to implement. We hope that this will lead to a new line of medical equipment in the operating room, and that our work has a tangible impact on the many surgeons that would use it, making their work easier by making the magnification process simpler, and then amplifying their abilities through the use of the YOLOV8 segmentation. In the future we hope to be able to develop our prototype to the point where it can be tested in a real surgery procedure, and then combine those test results with the results of the user study that we are conducting currently so that we have a better idea of how to move forward.

7. REFERENCES

[1] Yoon, J. W., Chen, R. E., Kim, E. J., Akinduro, O. O., Kerezoudis, P., Han, P. K., Si, P., Freeman, W. D., Diaz, R. J., Komotar, R. J., Pirris, S. M., Brown, B. L., Bydon, M., Wang, M. Y., Wharen, R. E., & Quinones-Hinojosa, A. (2018). Augmented reality for the surgeon: Systematic review. The International Journal of Medical Robotics and Computer Assisted Surgery, 14(4). https://doi.org/10.1002/rcs.1914

[2] Rautaray, Siddharth S., Real Time Hand Gesture Recognition System for Dynamic Applications (2012). International Journal of UbiComp (IJU), Vol.3, No.1, January 2012, Available at SSRN: https://ssrn.com/abstract=3702844

[3] "Bringing Gesture Recognition to All Devices." USENIX | The Advanced Computing Systems Association, www.usenix.org/conference/nsdi14/technical-sessions/presentation/kellogg.

[4] Fully automatic segmentation of glottis and vocal folds in endoscopic laryngeal high-speed videos using a deep Convolutional LSTM Network Fehling MK, Grosch F, Schuster ME, Schick B, Lohscheller J (2020) Fully automatic segmentation of glottis and vocal folds in endoscopic laryngeal high-speed videos using a deep Convolutional LSTM Network. PLOS ONE 15(2): e0227791. https://doi.org/10.1371/journal.pone.0227791

[5] Redmon, J.; Farhadi, A. YOLO9000: Better, Faster, Stronger. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (Cvpr 2017), Honolulu, HI, USA, 21–26 July 2017; pp. 6517–6525.

[6] Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014). Microsoft Coco: Common Objects in Context. Computer Vision – ECCV 2014, 740–755. https://doi.org/10.1007/978-3-319-10602-1_4