# Real-Time Framewise Semantic Understanding of Instruments and Tissues Via Deep Learning in Vitreoretinal Surgery

**Rogerio G. Nespolo[1, 2]; Darvin Yi, PhD[1, 2]; Yannek I. Leiderman, MD, PhD[1, 2]**

*[1]Department of Ophthalmology and Visual Sciences - Illinois Eye and Ear Infirmary, University of Illinois at Chicago*

*[2]Richard and Loan Hill Department of Biomedical Engineering, University of Illinois at Chicago*

## INTRODUCTION

Real-time imaging in vitreoretinal surgery directly impacts the handling of surgical instrumentation and the surgeon's decision-making process. This study employed deep learning neural networks to concurrently track and segment instruments and target tissue landmarks within the retina when attached to a surgical microscope, providing a platform to potentially improve surgical workflow via surgical guidance tools.

## METHODS

**DATASET**

A hundred and one (101) vitreoretinal procedures consisting of core vitrectomy, membranectomy, and endolaser application were employed to train and validate an instance segmentation convolutional neural network. In addition, three vitreoretinal surgeons manually annotated the following features from six-hundred and six (606) fundus frames: vitrector, forceps, and endolaser tooltips, optic disc, fovea, retinal tear and detachment, fibrovascular proliferation, endolaser spot, the area where endolaser was applied, and macular hole.

**DEEP LEARNING MODEL**

A fully convolutional instance segmentation model based on YOLACT++[1] was trained and implemented for this study. The model combines a ResNet-50 convolutional neural network[2] (CNN) with a feature pyramid network (FPN)[3], a prediction head network, and a prototype generator network. The ResNet-50 network weights were pre-trained on the ImageNet[4] dataset and retrained in our dataset via transfer learning. The training was performed during 2191 epochs in batches of 10 images, employing the stochastic gradient descent (SGD) optimizer with an initial learning rate of 1e-3, a momentum of 0.9, and a weight decay of 5e-4. Five-fold cross-validation was employed to assess the performance of the model. Figure 1 displays the schematic of the model's data pipeline.

**INTEGRATION WITH SURGICAL MICROSCOPE**

The final platform was integrated into a 3D visualization system employed in ophthalmic procedures. Pre-recorded videos were reproduced by the visualization system, and the post-processed image was displayed using the picture-in-picture function to demonstrate the integration of the platform with the operating room (Figure 2).

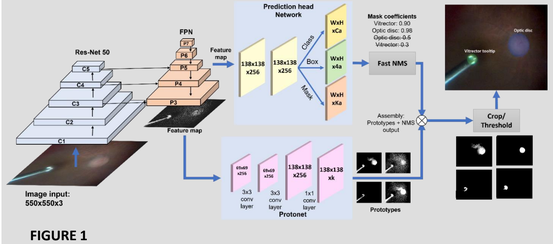


**FIGURE 1**
Architecture and data pipeline of our YOLACT++ model to detect and segment instruments and retinal features

## RESULTS

**MODEL'S PERFORMANCE ASSESSMENT**

| | Detection | Segmentation |
|---|---|---|
| **Feature** | **AUPR (mean ± SD)** | **AUPR (mean ± SD)** |
| Vitrector tooltip | 0.972±0.009 | 0.888±0.028 |
| Membrane peeling tooltip | 0.967±0.011 | 0.773±0.098 |
| Endolaser tooltip | 0.773±0.098 | 0.224±0.079 |
| Optic disc | 0.939±0.016 | 0.928±0.013 |
| Fovea | 0.863±0.035 | 0.844±0.039 |
| Retinal tear | 0.969±0.018 | 0.925±0.032 |
| Retinal detachment | 0.944±0.039 | 0.935±0.055 |
| Fibrovascular proliferation | 0.809±0.039 | 0.788±0.056 |
| Endolaser spot | 0.888±0.023 | 0.82±0.033 |
| Endolaser area applied | 0.687±0.033 | 0.667±0.035 |
| Macular Hole | 0.92±0.018 | 0.916±0.021 |

**TABLE 1**
Mean area under the precision-recall curve (AUPR) for the detection and segmentation of each feature.

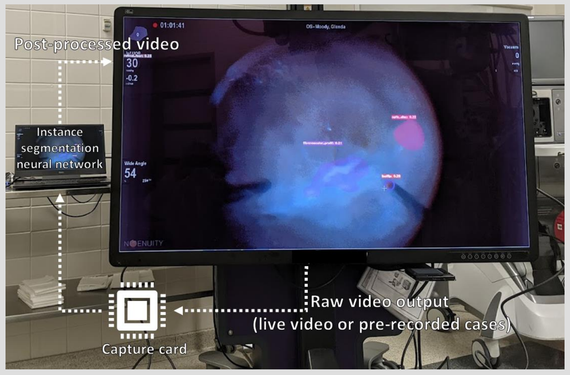


**FIGURE 2**
Integration of the platform in the operating room.

The model detected and classified the vitrector tooltip with a mean area under the precision-recall curve (AUPR) of 0.972±0.009. Segmentation of target tissues such as optic disc, fovea, and macular hole reached mean AUPR values of 0.928±0.013, 0.844±0.039, and 0.916±0.021, respectively (Table 1 and Figure 3). Our platform rendered the post-processed image at a resolution of 1920x1080 pixels at 38.77±1.52 frames per second when attached to an intraoperative visualization system (Figure 2).

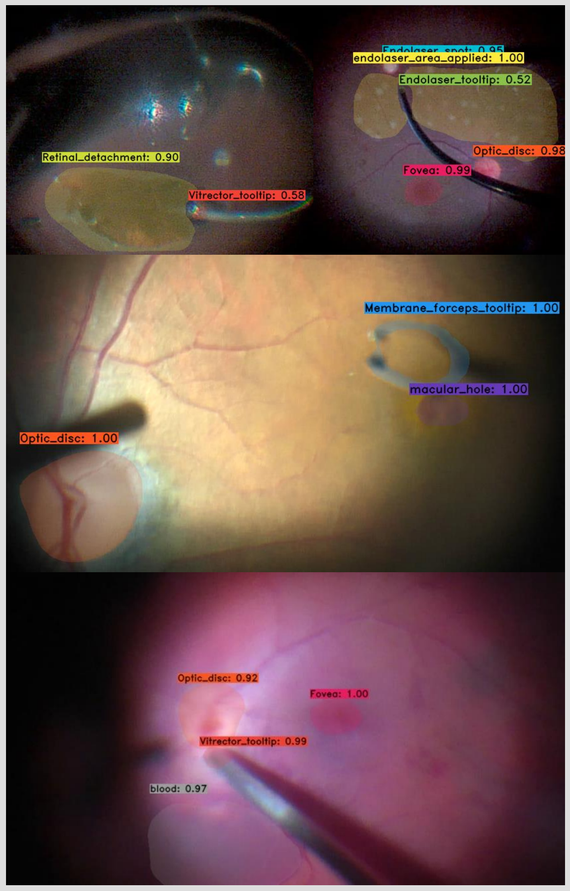


**FIGURE 3 - POST-PROCESSED IMAGES FROM OUR MODEL.**
Top left: Retinal detachment segmented and vitrector tooltip location detected. Top right: Detection and segmentation of retinal elements and instruments during endolaser photocoagulation. Bottom: Detection and segmentation of elements during membranectomy and with the presence of blood.

## CONCLUSION

The application of artificial intelligence-based image processing to ophthalmic microsurgery has the potential to improve performance, impact safety, and deepen our understanding of surgical training. The number of artificial intelligence-based solutions in ophthalmology has vastly increased during the last decade. Although, few solutions have been created with the potential to enhance vitreoretinal procedures.

We proved that deep learning neural networks can detect, classify, and segment tissues and instruments during different phases of a vitreoretinal surgery in real time. We proposed a model that creates a framework to develop surgical guidance tools that may guide surgical decision-making in real time. Potential applications include warning unintended instrument-tissue interactions, assistance for the surgeon with positioning surgical instruments, detection of potential intraoperative complications, data extraction to control equipment parameters such as the vitrector cutting rate, and retroactive data acquisition of instrument maneuvers for surgeon's skills analysis.

## REFERENCES

1. Bolya D, Zhou C, Xiao F, Lee YJ. ***YOLACT++: Better Real-time Instance Segmentation.*** IEEE Transactions on Pattern Analysis and Machine Intelligence 2020.
2. He K, Zhang X, Ren S, Sun J. ***Deep Residual Learning for Image Recognition.*** *2015.* Available at: http://arxiv.org/abs/1512.03385.
3. Lin T-Y, Dollár P, Girshick R, et al. ***Feature Pyramid Networks for Object Detection.*** 2016. Available at: http://arxiv.org/abs/1612.03144.
4. Krizhevsky A, Sutskever I, E. Hinton G. ***ImageNet Classification with Deep Convolutional Neural Networks*** *2012.* Advances in Neural Information Processing Systems 25 (NIPS 2012)

## FINANCIAL DISCLOSURES

Yannek Leiderman: Alcon (C, R, S, E); Microsurgical Guidance Solutions LLC (O, P); Allergan (C,S); Genentech (C); Regeneron (C,S); RegenXBio (C).

Rogerio Nespolo: Microsurgical Guidance Solutions LLC (O, P).

## ACKNOWLEDGEMENTS & SUPPORT

Special thanks to Dr. Emily Cole, Dr. Alexis Warren, and Dr. Daniel Wang for helping annotating the dataset.

Work supported by the Chancellor's Translational Research Initiative (CTRI)

Supported By
Research to Prevent Blindness