Evaluating Faster R-CNN for cataract surgery tool detection using microscopy video

H.Y. Lee[1], R. Hisey[1], M. Holden[2], J. Liu[3], T. Ungi[1], G. Fichtinger[1], C. Law[3,4]

[1]Laboratory for Percutaneous Surgery, School of Computing, Queen's University, Kingston, Canada
[2]School of Computing, Carleton University, Ottawa, Canada
[3]Department of Ophthalmology, Queen's University, Kingston, Canada

Introduction: Traditional methods of cataract surgery skill assessment rely on human expert supervision. This exposes the trainee to interobserver variability and inconsistent feedback. Alternative measures such as sensor-based instrument motion analysis promise objective assessment [1]. However, sensor-based systems are logistically complicated and expensive to obtain. Previous studies have demonstrated a strong correlation between sensor-based metrics and two-dimensional motion metrics obtained from object detection [2]. Reliable object detection is the foundation for computing such performance metrics. Therefore, the objective of this study is to evaluate the performance of an object detection network, namely Faster Region-Based Convolutional Neural Network (FRCNN), in recognition of cataract surgery tools in microscopy video.

Methods: Microscope video was recorded for 25 trials of cataract surgery on an artificial eye. The trials were performed by a cohort consisting of one senior-surgeon and four junior-surgeons and manually annotated for bounding box locations of the cataract surgery tools (Figure 1) The surgical tools used included: forceps, diamond keratomes, viscoelastic cannulas, and cystotome needles. A FRCNN [3] was trained on a total of 130,614 frames for object detection. We used five-fold cross validation, using a leave-one-user-out method. In this manner, all videos from one surgeon were reserved for testing and the frames from the remaining 20 videos were divided among training and validation. Network performance was evaluated via mean average precision (mAP), which is defined as the area under the precision/recall curve. Samples were considered correctly identified when the intersection over union (IoU) between the ground truth and predicted bounding boxes was greater than 0.5.

Table 1. Mean average precision (mAP) by tool

Tool	*mAP*
Forceps	0.61
Diamond Keratome	0.49
Viscoelastic Cannula	0.59
Cystotome Needle	0.96

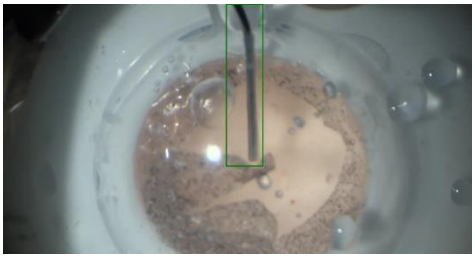

Figure 1. Manually annotated bounding box.

Results: The overall mAP of the network was 0.63. Tool-specific mAPs ranged between 0.49 and 0.96 (Table 1). The high accuracy in detection of the cystotome needle is likely due to the distinct size and shape of the tool tip. The diamond keratome had the lowest mAP of any of the tools recognized, however this may be attributed to variations in the appearance of the tool tip (Figure 2).

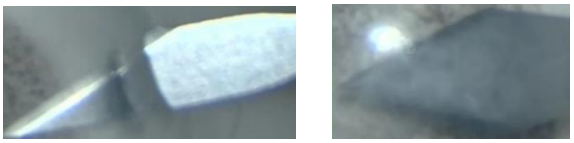

Figure 2. Right angle diamond keratome (left) and isosceles diamond keratome (right).

Conclusions: The FRCNN was able to recognize the surgical tools used in cataract surgery with reasonably high accuracy. Now that we know that the network can sufficiently recognize the surgical tools, our next goal is to use this network to compute motion-based performance metrics. Future work seek to validate these performance metrics against those obtained from sensor-based tracking and against expert evaluations. This serves as a first step towards providing consistent and accessible feedback for future trainees learning cataract surgery.

References:

[1] G. M. Saleh, et al, "Evaluating Surgical Dexterity During Corneal Suturing," *Archives of Ophthalmology*, vol. 124, no. 9, pp. 1263–1266, Sep. 2006.
[2] O. O'Driscoll *et al.*, "Object detection to compute performance metrics for skill assessment in central venous catheterization," in *SPIE Medical Imaging 2021*, Feb. 2021, vol. 11598, pp. 315–322.
[3] S. Ren, et al, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *arXiv:1506.01497 [cs]*, Jan. 2016.