2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Automated detection of electrocautery instrument in videos of open neck procedures using YOLOv3 *

Tingyan Deng[1], Shubham Gulati[2], William Rodriguez[3], Benoit M. Dawant[4], Alexander Langerman[5]

***Abstract*— With the rapid development of deep learning approaches, tremendous progress has been made in computer-assisted analysis of minimally-invasive, videoscopic surgery. However, surgery through open incisions ("open surgery"), which constitutes a much larger portion of surgical procedures performed, is rarely investigated because of the difficulty in obtaining high-quality open surgical video footage. Automated detection of surgical instruments shows promise for evaluating surgical activities, and provides a foundation for quality/safety review, education, and identification of surgical performance. In this paper, we present results using YOLOv3 to successfully identify an electrocautery surgical instrument in a library of images derived from 22 open neck procedures (an 887-image training/validation set, and a 1149-image testing set) captured using a wearable surgical camera. We show that our method effectively detects the spatial bounds of the electrocautery pencil in still images and we further demonstrate the ability of our method to detect the location of this instrument in video footage. Our work serves as the first demonstration of open surgical instrument detection using first-person video footage from a wearable camera and sets the stage for further work in this field.**

***Clinical Relevance*— Detection of instrumentation in surgical video is the necessary first step towards automating surgical task identification and skills assessment, which will be useful for surgical quality improvement and training.**

I. INTRODUCTION

A. Videography in Surgery

The routine use of video in the operating room has introduced a paradigm shift in surgical quality improvement. Recordings from minimally-invasive, videoscopic procedures (e.g., laparoscopy, robotic surgery) have been used to identify surgical errors [1] and identified strong correlations between surgeon skill performance and patient outcomes [2]. Increasingly, deep learning is being used to automate these assessments [3]. Despite the promise of surgical video, this work has been almost exclusively limited to videoscopic procedures, leaving a gap in applications for procedures performed through open incisions ("open surgery"). The lag in progress for open surgery video analysis has been primarily due to limitations with recording these procedures, including poor video quality due to obstructions, excessive motion, light overexposure, and other limitations related to battery life and surgeon comfort [4][5]. These difficulties have until now limited the translation of existing deep learning research on footage from videoscopic and microscope-mounted cameras [6][7][8] to video footage captured by wearable cameras.

B. Cleopatra Surgical Camera

To address the challenges in recording open surgery, we developed a novel video platform designed for open procedures: Clearer Operative Analysis and Tracking ("Cleopatra"). Cleopatra is a neck-worn camera that provides an inherently stable, first-person view of the operative field; this both avoids the instability and movement of head mounted cameras and the potential for obstruction suffered by overhead, boom- or light-mounted cameras.

Our group has previously used Mask R-CNN to achieve the automated recognition and segmentation of the surgical wounds in videos of open procedures captured using Cleopatra [9].

C. Surgical Instrument

Beyond the wound, the two other major elements of the surgical field are the surgeons' hands and the surgical instruments. The detection of surgical instruments during a procedure is currently an area of interest for automating analysis of surgical activities and skills assessment [10]. As the next step towards tracking surgical activities during open procedures, we have focused on detection of the widely-used electrocautery pencil device (colloquially, the "bovie", Figure 1) in this work. Because in many open procedures, much of the cutting action is accomplished using this instrument, this is an ideal target for assessing surgical skills.

D. YOLOv3

In order to construct the object detection network and accomplish instrument detection at real-time speed, we utilized the real-time object detection method: "you only look once" (YOLOv3) described in [11]. Compared to other algorithms like faster R-CNN and RetinaNet, YOLO is faster and uses a single convolutional network to detect objects; these features allow it to be employed by small processors and thus are more applicable to wearable recording devices for real-time analysis. In this manuscript, we present our success using YOLOv3 to identify the electrocautery pencil instrument in video stills and video sequences from a variety of open surgical scenarios captured by a wearable camera.

*This work was supported by the Vanderbilt Institute for Surgery and Engineering (VISE)

[1]T. Deng is with the Department of Electrical Engineering and Computer Science, Nashville, Tennessee, USA (e-mail: tingyan.deng@vanderbilt.edu)

[2]S. Gulati is with the Department of Biomedical Engineering, Nashville, Tennessee, USA (e-mail: shubham.gulati@vanderbilt.edu)

[3]W. Rodriguez is with the Vanderbilt Institute for Surgery and Engineering, Nashville, Tennessee, USA. (bill.rodriguez01@gmail.com)

[4]B.M. Dawant is with the Department of Electrical Engineering and Computer Science and Vanderbilt Institute for Surgery and Engineering, Nashville, Tennessee, USA. (benoit.dawant@vanderbilt.edu)

[5]A. Langerman is with the Department of Otolaryngology-Head and Neck Surgery, the Department of Radiology and Radiological Sciences, the Center for Biomedical Ethics and Society, and the Vanderbilt institute for Surgery and Engineering, Vanderbilt University Medical Center, Nashville, Tennessee, USA (AlexanderLangerman.com)

978-1-7281-1178-0/21/$31.00 ©2021 IEEE

II. Methods

A. Data Source

Our data source was a library of 22 videos of open neck procedures captured using Cleopatra, which records point-of-view video footage of the surgical field at 1080p/30fps in h264 format; these are subsequently converted to mp4 format for editing and exportation of images. These videos were collected under an IRB-approved quality improvement protocol and contained no identifiable images or metadata. A total of 430 images were initially extracted from 12 of those videos to generate a still-image library for our training set (Data Set A); 521 images were extracted from 4 separate videos for testing (videos A-D, in Figure 4). After initial testing we identified that most false positive identification errors were due to the presence of other linear, blue apparatuses in the surgical field, (e.g., elastic retraction hooks, vessel loops; Figure 2). Thus, to further improve our model and lower false positive and negative identification, we subsequently generated an additional 457 training images from our original 12 training videos; this set included 53 images that contained both the electrocautery pencil and other linear, blue items that were our most common "distractors" (Data Set B). We also added another 6 videos for testing (videos E-J in Figure 4) and extracted an additional 628 images from these 6 videos for a total final testing set of 1149 still images from 10 videos.

B. Video Markup

The focus of this experiment is the automatic recognition of the blue, insulated tip of the electrocautery pencil. A practicing surgeon identified this instrument in every image and Labellmg (discussed in [12]) was used to draw a bounding box and export the annotated images in YOLO format containing the four (x,y) coordinates that defined the bounding box; these data were used to train the network.

C. Network Training

Each training involved forty-thousand iterations with 100 training steps per epoch (batch size = 64, learning rate = 0.001). Training was repeated after doubling the training/validation set and adding distractor images.

D. Network Testing

Both networks, Data Set A and Data Set B, were tested on a set of 1149 still images derived from 10 videos from separate procedures, none of which were used for training or validation. We then also tested our network on ten video clips (nine 24-minute videos and one 12-minute video) to examine performance on video as opposed to still images, and quantitatively analyzed these results.

E. Outcome Measures

Our primary outcome measures for model performance on images and video were 1) true positive – recognition of the instrument when in frame; 2) false positive – recognition of non-instrument as instrument; 3) false negatives – failure to recognize the instrument when in frame; and 4) true negative – no instrument detected when no instrument was visible in the frame. F-1 score of both rounds of testing was then calculated from these data.

III. Results

A. Instrument Recognition

Our model recognized the electrocautery pencil across multiple procedures in our testing images and videos. Figure 1 shows the bounding box found by our model in an example image. Note that the number near the bounding box represents the detection probability (i.e. the confidence level of the bounding box).

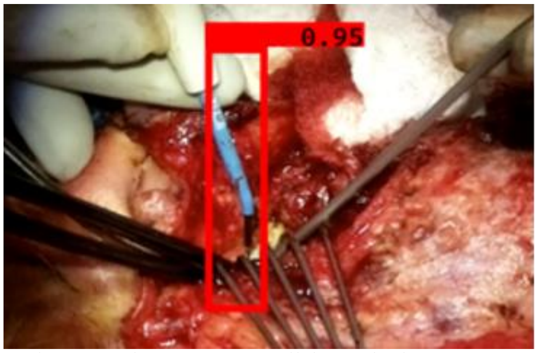

Figure 1 – Electrocautery pencil detected using YOLOv3

B. Model Performance

The model performed well on the bounding box detection task, with a final F-1 score of 0.906 for Data Set B (Table 1) after the addition of more training images, including those of blue distractors. Despite this, and likely because of the lower detection probability threshold, false positives remained high at 10%; this was often realized as recognition of a blue vessel loop or elastic retractor as a secondary electrocautery pencil rather than complete non-recognition of the actual electrocautery pencil when it was in frame (Figure 2).

TABLE 1. Comparative performance of two YOLOv3 networks on the same testing set of 1149 surgical images

	Data Set A Network (430 images)	Data Set B Network (887 images)
True Positive	801	882
False Positive	162	113
False Negative	103	71
True Negative	83	83
F-1 Score	0.8581	0.9055

C. Errors in Instrument Detection

On visual inspection of images with misidentification in our initial model, we identified that eccentric positioning of the electrocautery pencil was the major source of false negatives. False positives were due to the presence of other linear, blue "distractors" which look similar to the insulated tip of the electrocautery pencil. We tuned our model by enlarging our training set and adding training images that specifically include "distractors", reducing the rate of false negatives upon retraining. Examples of a persistent false positive is shown in Figure 2. An example of our model being able to decrease the false positives after re-training with more "distractors" (Data Set B) is shown in Figure 3.

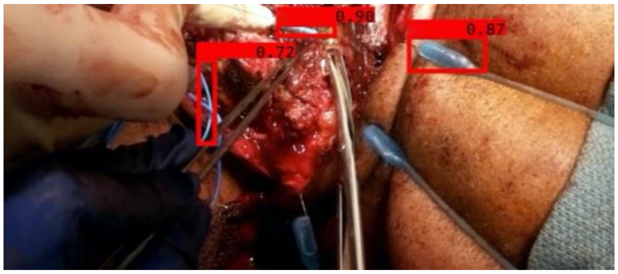

Figure 2 – False-positive detection of a blue vessel loop (left) and a blue elastic retaining hook (right) as the electrocautery pencil, in addition to the actual electrocautery pencil (top center).

D. *Application of Model to Video Sequences*

As proof-of-concept, we also tested our Data Set B model on video clips and discovered that it performed well in detecting the instrument as it appeared in the surgical field and tracking it through the frame. We analyzed 228 total minutes of video footage across the 10 videos in the testing set. Of these, the electrocautery pencil was present in-frame for 141 minutes (64.1%). Of these 141 minutes, our model successfully detected the instrument for 132 minutes (98.6% true positive rate) and failed detection for 9 minutes (1.36% false negative rate). For 87 minutes, the instrument was not in frame, and of this time, our model accurately detected no instrument for over 85 minutes (92.4% true negative rate) and incorrectly detected a non-instrument as an instrument for 1.5 minutes (7.6% false positive rate). The electrocautery pencil frequently moved in and out of the frame, so detections were scattered across the length of the video as shown in Figure 4 and there were more true negatives than in the image sets. The model performed well and similarly on the initial four videos for our first testing (A-D; F-1=0.933) and on the additional 6 videos we added after retraining (E-J; F-1=0.917).

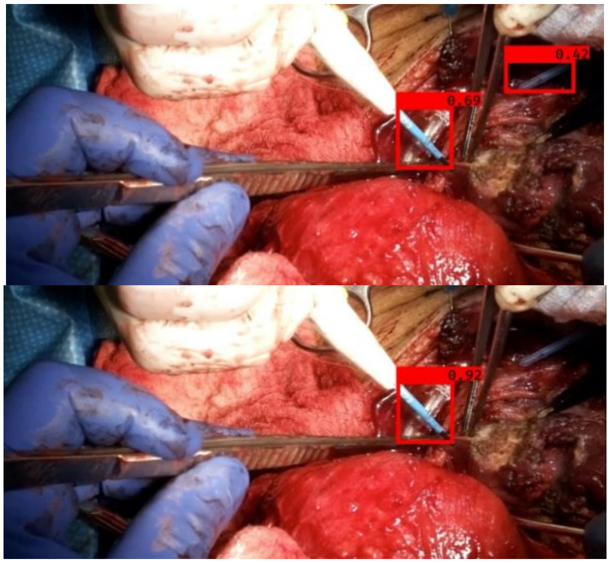

Figure 3 – Re-training the model with additional images that included "distractor" items like blue elastic resulted in decreased false-detection of these objects. Upper panel is still image from testing of Data Set A, and lower panel is testing results of same still image using the re-trained, Data Set B.

IV. Discussion

A. *Key Findings*

Our work is the first demonstration of automated recognition of a surgical instrument in point-of-view open surgical video footage captured by a wearable device. Our identification of the electrocautery pencil instrument using YOLOv3 was successful and our algorithm showed promise with a relatively small training set. This work paves the way for further research on open surgical video capture optimization.

B. *Future Directions*

The preliminary results presented in this manuscript serve as proof-of-concept for deep learning network applications to wearable surgical video footage. YOLOv3 provided reasonable performance, but better results might be obtained with more recent algorithms such as YOLOv4 and v5, and real-time performance on a small device may require using YOLO-lite. To augment our model, we utilized visual review of false detection frames, identifying "distractors" that, when specifically introduced into a re-training set, reduced the false positive rate. However, we did not simply fit the model to a specific test set; rather we identified and corrected distractors that are common to all videos, as evidenced by the similar Data Set B model performance on the four videos used for initial testing (A-D) and the additional six testing videos (E-J) added after retraining.

This work paves the way towards two future goals: motion analysis and multi-item analysis. Motion analysis will allow us to track the movement of the instrument through the frame, which may correlate with surgical performance. For example, one of the features of a widely-used surgical skills metric is "time and motion" [13]; plots of instrument motion could be assessed for smoothness as a quantitative measure of this quality. The correlation between instrument motion and skill level has already been demonstrated in previous work using an electromagnetic sensor directly attached to a surgical instrument [14], and this same principle may be applied in a "no-touch" manner using video analysis. Our preliminary video results presented in Figure 4 demonstrate such instrument tracking to be feasible across long video segments.

Our second goal is to achieve a simultaneous identification and tracking of the multiple instruments as well as anatomical features of the surgical field. While single-object detection is relatively straightforward, a more comprehensive analysis of multi-item activity (e.g., interaction of instruments and tissue, switching of instruments, and case phase identification) will require correspondingly more complex network training methods. We anticipate that we can apply the same training insights gleaned from this study to these more complex networks to improve our results.

C. *Impact on Current Medicine/Biology*

Identification of surgical instruments is necessary to develop computational methods for assessing surgical technical skill. This is the first demonstration of surgical instrument recognition in open surgical video footage from a wearable device. While this method is limited to the detection

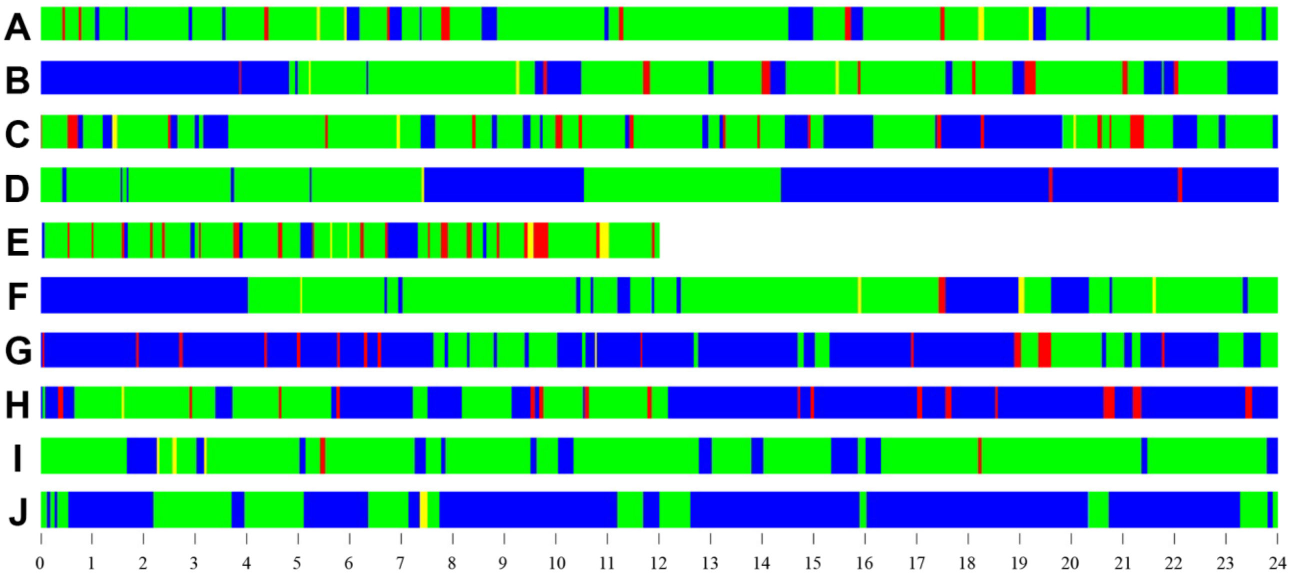

Figure 4 – Full sequences of the ten testing videos (initial 4 were A-D, second 6 were E-J). Colors indicate results of model testing, and duration of results are represented by width of colored segment on a time axis of 24 minutes. Video E was only 12 minutes long. Green = True Positive, Blue = True Negative, Red = False Positive, Yellow = False Negative

of one instrument, further development will allow us to expand to the detection of multiple instruments at a time. With the advancement of instrument tracking in open surgery, methods of assessing laparoscopic surgical skill via instrument movement can be adapted to more procedure types. This wearable videography method coupled with surgical instrument tracking expands the current scope of automated surgical skill assessment, leading to enhancements in training, credentialing, and remediation.

V. Conclusion

Our automatic recognition of an electrocautery instrument in images obtained from the "Cleopatra" wearable surgical camera during open neck procedures using YOLOv3 was successful. We identified images with an eccentric placement of the instrument and with other similar blue apparatuses as the primary sources of model errors on visual review, and refined our model to address this, ultimately achieving a low rate of false positive and negative errors. Our model was also qualitatively and quantitatively successful when applied to video clips. This work forms a foundation for automated assessment of surgical skills in open surgical procedures.

References

[1] A. Langerman, "Using Surgical Video to Classify Intraoperative Events," *Ann Surg*. 272(2):227-228, Aug 2020, doi: 10.1097/SLA.0000000000003934. PMID: 32675486.

[2] KR Chhabra , JR Thumma, OA Varban, JB Dimick, "Associations Between Video Evaluations of Surgical Technique and Outcomes of Laparoscopic Sleeve Gastrectomy," *JAMA Surg*. Feb 2021.

[3] C. Loukas, "Video content analysis of surgical procedures," *Surg Endosc*.32(2):553-568. Feb 2018 doi: 10.1007/s00464-017-5878-1. Epub 2017 Oct 26. PMID: 29075965.

[4] T.J. Saun, K.J. Zuo, T.P. Grantcharov, "Video Technologies for Recording Open Surgery: A Systematic Review" in *Surgical Innovation*, 26(5):599-612, 2019.

[5] E.Kapi, "Surgeon-Manipulated Live Surgery Video Recording Apparatuses: Personal Experience and Review of Literature". *Aesth Plast Surg* 41, 738–746, 2017.

[6] H.Sugimori, T.Sugiyama, N.Nakayama, A.Yamashita, K.Ogasawara, "Development of a Deep Learning-Based Algorithm to Detect the Distal End of a Surgical Instrument," *Applied Sciences*; 10(12):4245, 2020.

[7] I. Funke, S.T. Mees, J. Weitz, S. Speidel, "Video-based surgical skill assessment using 3D convolutional neural networks," *Int J CARS* **14,** 1217–1225, 2019. https://doi.org/10.1007/s11548-019-01995-1

[8] A. Jin, S. Yeung, J. Jopling, J. Krause, D. Azagury, A. Milstein *et al.*, "Tool Detection and Operative Skill Assessment in Surgical Videos Using Region-Based Convolutional Neural Networks," *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, Lake Tahoe, NV, USA, pp. 691-699, 2018, doi: 10.1109/WACV.2018.00081.

[9] T. Y. Deng, S. Gulati, A. Kumar, W. Rodriguez, B. Dawant, and A. Langerman, "Automated detection of surgical wounds in videos of open neck procedures using a mask R-CNN," *Medical Imaging 2021: Image-Guided Procedures, Robotic Interventions, and Modeling*, 2021.

[10] J.L., Lavanchy, J. Zindel, K. Kirtac, I. Twick, E. Hosgor *et al.,* "Automation of surgical skill assessment using a three-stage machine learning algorithm," *Sci Rep* 11, 5197, 2021. https://doi.org/10.1038/s41598-021-84295-6

[11] J. Redmon, A. Farhadi, "YOLOv3: An Incremental Improvement," in *ArXiv* abs/1804.02767, 2018.

[12] Tzutalin, "Labellimg," *Git code*[online], June 20 2021. Available : https://github.com/tzutalin/labelImg

[13] Martin JA, Regehr G, Reznick R, MacRae H, Murnaghan J, Hutchison C, Brown M. "Objective structured assessment of technical skill (OSATS) for surgical residents," *Br J Surg*. 84(2):273-8. doi: 10.1046/j.1365-2168.02502.x, 1997 Feb, PMID: 9052454.

[14] Ahmidi N, Poddar P, Jones JD, Vedula SS, Ishii L, Hager GD, Ishii M. "Automated objective surgical skill assessment in the operating room from unstructured tool motion in septoplasty," *Int J Comput Assist Radiol Surg*.10(6):981-91, 2015 Jun, doi: 10.1007/s11548-015-1194-1. PMID: 25895080.