



Original Investigation | Surgery

Utility of the Simulated Outcomes Following Carotid Artery Laceration Video Data Set for Machine Learning Applications

Guillaume Kugener, MEng; Dhiraj J. Pangal, BS; Tyler Cardinal, BS; Casey Collet, BS; Elizabeth Lechtholz-Zey, BS; Sasha Lasky, BS; Shivani Sundaram, BS; Nicholas Markarian, BS; Yichao Zhu, MS; Arman Roshannai; Aditya Sinha, BS; X. Y. Han, MS; Vardan Papyan, PhD; Andrew Hung, MD; Animashree Anandkumar, PhD; Bozena Wrobel, MD; Gabriel Zada, MD, MS; Daniel A. Donoho, MD

Abstract

IMPORTANCE Surgical data scientists lack video data sets that depict adverse events, which may affect model generalizability and introduce bias. Hemorrhage may be particularly challenging for computer vision-based models because blood obscures the scene.

OBJECTIVE To assess the utility of the Simulated Outcomes Following Carotid Artery Laceration (SOCAL)—a publicly available surgical video data set of hemorrhage complication management with instrument annotations and task outcomes—to provide benchmarks for surgical data science techniques, including computer vision instrument detection, instrument use metrics and outcome associations, and validation of a SOCAL-trained neural network using real operative video.

DESIGN, SETTING, AND PARTICIPANTS For this quality improvement study, a total of 75 surgeons with 1 to 30 years' experience (mean, 7 years) were filmed from January 1, 2017, to December 31, 2020, managing catastrophic surgical hemorrhage in a high-fidelity cadaveric training exercise at nationwide training courses. Videos were annotated from January 1 to June 30, 2021.

INTERVENTIONS Surgeons received expert coaching between 2 trials.

MAIN OUTCOMES AND MEASURES Hemostasis within 5 minutes (task success, dichotomous), time to hemostasis (in seconds), and blood loss (in milliliters) were recorded. Deep neural networks (DNNs) were trained to detect surgical instruments in view. Model performance was measured using mean average precision (mAP), sensitivity, and positive predictive value.

RESULTS SOCAL contains 31 443 frames with 65 071 surgical instrument annotations from 147 trials with associated surgeon demographic characteristics, time to hemostasis, and recorded blood loss for each trial. Computer vision-based instrument detection methods using DNNs trained on SOCAL achieved a mAP of 0.67 overall and 0.91 for the most common surgical instrument (suction). Hemorrhage control challenges standard object detectors: detection of some surgical instruments remained poor (mAP, 0.25). On real intraoperative video, the model achieved a sensitivity of 0.77 and a positive predictive value of 0.96. Instrument use metrics derived from the SOCAL video were significantly associated with performance (blood loss).

CONCLUSIONS AND RELEVANCE Hemorrhage control is a high-stakes adverse event that poses unique challenges for video analysis, but no data sets of hemorrhage control exist. The use of SOCAL, the first data set to depict hemorrhage control, allows the benchmarking of data science applications, including object detection, performance metric development, and identification of metrics associated with outcomes. In the future, SOCAL may be used to build and validate surgical data science models.

JAMA Network Open. 2022;5(3):e223177. doi:10.1001/jamanetworkopen.2022.3177

Open Access. This is an open access article distributed under the terms of the CC-BY License.

Key Points

Question What is the utility of a data set that contains videos of surgeons managing hemorrhage?

Findings This quality improvement study of the Simulated Outcomes Following Carotid Artery Laceration (SOCAL), a public data set of surgeons managing catastrophic surgical hemorrhage in a cadaveric training exercise included 65 071 instrument annotations with recorded outcomes. Computer vision-based instrument detection achieved a mean average precision of 0.67 on SOCAL and a sensitivity of 0.77 and a positive predictive value of 0.96 at detecting surgical instruments from real intraoperative video.

Meaning A corpus of videos of surgeons managing catastrophic hemorrhage is a novel, valuable resource for surgical data science.

+ Supplemental content

Author affiliations and article information are listed at the end of this article.

Introduction

Advances in machine learning and computer vision (CV) facilitate the discovery of unknown patterns in large bodies of visual data.¹⁻⁵ Applying these methods to analyze surgical video from the operating room may improve patient care by detecting complications, assisting trainee development, and assessing performance.⁶⁻¹⁰ Algorithmic methods thrive when presented with comprehensive data sets. Surgical video data sets, however, are traditionally collected to answer a single question by a single group of investigators, depicting a particular surgical scenario (eg, one data set for object detection in neurosurgery and a different data set for phase detection in general surgery).¹¹⁻²³ This patchwork approach risks fostering data set bias, a condition that occurs when model training data are not representative of the true data landscape (ie, all surgical scenarios).²⁴

We cataloged surgical video data sets into a single corpus (eTable 1 in the [Supplement](#)) and detected a clinically relevant data set bias: the absence of surgical adverse events. Because of institutional and medicolegal sensitivities, no video data sets of surgeons encountering complications, particularly bleeding events, are available, which creates a dangerous lacuna for CV algorithms. Investigators cannot develop tools to assess complication management and cannot evaluate the performance of existing tools during scenarios that compromise visualization, such as hemorrhage.

To address this gap, we recorded videos from a multiyear, nationwide educational curriculum in which surgeons of all experience levels encountered catastrophic hemorrhage in a high-fidelity, validated cadaveric model of internal carotid artery (ICA) laceration. Laceration of the ICA is a life-threatening risk inherent in anterior cranial base surgery (eg, endoscopic sinus surgery and pituitary tumor surgery) that can result in stroke, neurologic injury, or death.^{25,26} Because ICA laceration occurs in less than 0.2% of cases,^{25,27} most surgical trainees never experience this event during formal surgical training. Video of these scenarios often contains bleeding that can obscure the entire surgical scene. These cadaveric simulations allowed surgeons to hone their skills and receive feedback in a highly realistic, structured environment without risking patient harm.²⁸⁻³⁰

We used these videos to develop the Simulated Outcomes Following Carotid Artery Laceration (SOCAL) data set. SOCAL's size, high level of realism, varied anatomy and operative views, and validated performance outcomes³⁰⁻³³ make it a unique resource. These otherwise unobtainable videos of hemorrhage complication management can support surgical data science tasks and be used to develop machine learning models. This article publicly releases the SOCAL data set³⁴ and provides several illustrative cases central to surgical data science using this data set: deep learning performance benchmarking, metric development, and intraoperative video instrument detection.

Methods

Surgeon verbal consent for video recording was obtained, and the cadaveric experiment was approved by the institutional review board of the University of Southern California. For the patient case discussed, consent was waived by the institutional review board given the study's retrospective nature and the nonidentifiable nature of the intraoperative video. This quality improvement study followed the Standards for Quality Improvement Reporting Excellence ([SQUIRE](#)) reporting guideline.

Data Set Development

We collected videos of surgeons participating in nationwide training courses for ICA laceration management from January 1, 2017, to December 31, 2020. These courses were held at the North American Skull Base Society, Keck School of Medicine of the University of Southern California, the Emory University School of Medicine, and the Combined ENT & Neurosurgery Advanced Resident Sinus and Skull Base Dissection course in Park City, Utah (sponsored by Stryker Corp). Surgeon characteristics and task outcomes were recorded. Details of the methods,³⁵ validation,³² costs,³¹ and results³⁰ of the training course have previously been published. In brief, surgeons attempted to

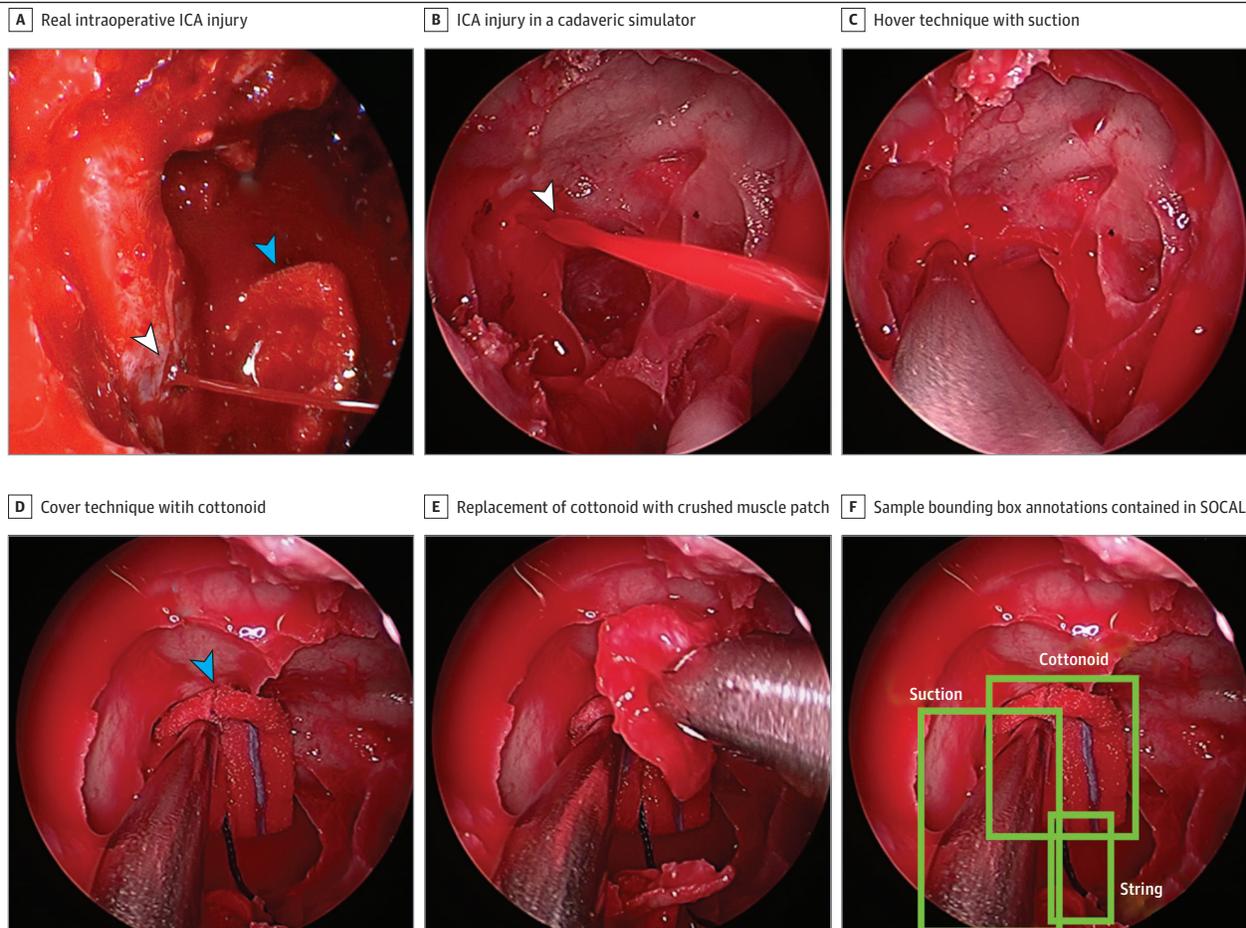
manage an ICA laceration in a perfused (bleeding) human cadaveric head during a standardized exercise: the first attempt was performed before coaching and the second attempt after personalized coaching instruction was delivered. A real ICA example is shown in **Figure 1A**, and the steps to achieve hemostasis in the cadaver model are depicted in Figure 1B-E and described in the eMethods in the Supplement. These experiments have been validated as having exceptionally high realism and transferability to the operating room.^{30-33,35}

The Karl Storz Video Neuro-Endoscope records intraoperative video during each of these trials. The SOCAL data set consists of 147 annotated videos from 7 different cohorts of surgeons; trial duration varies from 46 seconds to 5 minutes. We developed the data set following previously published methods.³⁶ Surgical instruments were hand annotated in each video frame with bounding boxes using the annotation tool Vott (Figure 1F).³⁷

Object Detection Algorithm and Evaluation

Our data set was used to train 2 object detection models: RetinaNet and YOLOv3.^{38,39} Both RetinaNet and YOLOv3 are convolutional neural network (CNN)-based deep neural networks (DNNs) developed to detect spatial patterns within images. RetinaNet is composed of a backbone network that computes a feature map over an input image using CNNs and 2 task-specific subnetworks that use CNNs to classify objects detected in the backbone network and perform bounding box regression. YOLOv3 is a single-stage object detection method that uses CNNs to predict bounding

Figure 1. Endoscopic Images of Internal Carotid Artery (ICA) Injury Management



An actual ICA and steps to achieve hemostasis in the cadaver model. Surgical instruments were hand annotated in each video frame with bounding boxes using the annotation tool Vott. White arrows indicate injury; blue arrows, instrument.

boxes and assign probabilities to class types from the entire image. These architectures achieved competitive results on ImageNet,^{38,39} which is a standardized data set used to benchmark and compare object detection methods.⁴⁰ An intersection over union greater than 0.5 (ie, where >50% of the area of the detection lies within the area of the ground truth annotation) was considered a successful detection. Further details on model implementation are in the eMethods in the [Supplement](#).

Instrument Use Pattern Features

We generated rudimentary instrument use variables from ground truth annotations of surgical instruments. Features of interest were selected based on known ICA injury surgical progression and clinical expertise. Instrument use patterns for each key instrument in the procedure (suction, grasper, cottonoid, string, and muscle) were generated by summing the number of frames with the instrument in view and dividing that value by the total number of frames for that trial.

Intraoperative Video Analysis

Three and a half minutes (210 seconds) of video was taken from the case of an uncomplicated endoscopic, endonasal tumor resection of a sellar craniopharyngioma (not published within SOCAL) to validate automatic tool detection in a real intraoperative setting. We parsed the video into individual frames at 0.5 frames per second (total of 105 frames). We used the RetinaNet model trained on SOCAL and validated it by predicting the location of instruments in each frame of the operative case video. The model output bounding boxes around detected instruments without specifying tool type (ie, instruments were labeled as *tool* instead of *grasper* or *suction*). A member of the research team performed ground truth annotations. Videos were annotated from January 1 to June 30, 2021.

Statistical Analysis

Object detection model performance on SOCAL was evaluated using mean average precision (mAP) (average area under the precision-recall curve), a standard metric used to evaluate object detection networks. Object detection model performance on real intraoperative video was evaluated using sensitivity and positive predictive value for instrument detections. For performance metric associations with outcomes, we fit individual linear regression models to predict blood loss from each instrument use pattern while accounting for the time to hemostasis for that trial as a covariate and reported the effect size, variance explained, and associated *P* value of each fitted model. When analyzing instrument usage patterns, we consider a 2-sided *P* < .05 to be statistically significant.

Results

Data Set Description

SOCAL contains 147 videos of 75 individual surgeons (neurosurgeons and otorhinolaryngologists with 1 to 30 years' experience [mean, 7 years]) managing a simulated life-threatening hemorrhage in a cadaveric model before and after expert coaching during a 4-year period (7 cohorts) (eTable 2 and eMethods in the [Supplement](#)).³⁰ Surgeon experience ranged from junior trainees to experts (48 trainees, 25 attending physicians, and 2 not reported). Almost half of the surgeons failed to control hemorrhage on the first attempt, resulting in simulated patient mortality. Paired videos with outcome data are available for 71 surgeons (4 unpaired trials; 1 surgeon with a third trial). Thirty-one participants failed in trial 1 and succeeded in trial 2 after expert instruction. Blood loss decreased by 29% (154 mL; 95% CI, 65-243 mL) after expert instruction. Surgeon participant outcomes are given in eTable 2 in the [Supplement](#).

A total of 31 443 frames of video were annotated with instrument-level bounding boxes across 5 surgical instruments (**Table 1**) for a total of 65 071 unique instrument instances. Each frame had 0 to 5 surgical instruments. An image from the SOCAL data set compared with an actual intraoperative

ICA injury is shown in Figure 1, illustrating the high-fidelity nature of the simulator. A description of surgical instrument findings within the data set is given in Table 1.

We followed the NeurIPS 2021 Code and Data Submission Guidelines for data set publishing to ensure accessibility.⁴¹ In following these guidelines, we placed the Digital Object Identifier (DOI) of the public SOCAL data set in a repository that ensures long-term preservation, makes the data set publicly available for download,³⁴ and licenses the data set under a Creative Commons Attribution 4.0 International Public License. To illustrate the utility of SOCAL as a test bed for surgical data science techniques, we outline 3 use cases: benchmarking deep learning object detection, developing performance metrics from intraoperative video, and validating instrument detection using real intraoperative video.

Benchmarks for Object Detection Using SOCAL

We used 2 off-the-shelf object detection DNNs to demonstrate the feasibility of instrument detection using SOCAL. The overall mAPs were 0.67 for RetinaNet and 0.53 for YOLOv3. We then evaluated the mAP for each object in our data set; the results are visualized in Figure 2. The models predict suction (mAPs, 0.91 for RetinaNet and 0.82 for YOLOv3) and grasper (mAPs, 0.77 for RetinaNet and 0.68 for YOLOv3) with high accuracy. In contrast, the mAPs for cottonoid (0.52 for RetinaNet and 0.33 for YOLOv3) and string (0.5 for RetinaNet and 0.33 for YOLOv3) are lower, and the mAPs for muscle are poor (0.25 for RetinaNet and 0.1 for YOLOv3). Training, validation, and test splits and outcomes are given in Table 2.

Benchmarks for Instrument Use Patterns Using SOCAL

As a feasibility demonstration of objectively describing instrument movement, instrument use patterns were calculated from on-screen bounding box location data for the 5 critical instruments used in this simulation (suction, grasper, cottonoid, string, and muscle). We found that the following features accounted for a significant amount of the variance in blood lost: the proportion of frames with grasper in use, the proportion of frames with cottonoid in use, the proportion of frames with muscle in use, and the proportion of frames with suction in use (Table 2). Time to hemostasis alone accounts for 49.84% (effect size, 2.9; $P < .001$) of the variance in blood lost.

Instrument Detection in a Real Intraoperative Video

As a proof-of-concept for the SOCAL database to serve as training data for video from real patient cases, a 3-minute intraoperative video of endoscopic resection of a tumor of the anterior cranial base was analyzed. One hundred seventy-nine unique instances of surgical instruments were annotated as ground truth data. The RetinaNet model that was trained using the SOCAL data set accurately detected 138 instruments (sensitivity, 0.77). There were 41 instances of missed instruments. There

Table 1. Simulated Outcomes Following Carotid Artery Laceration Video Data Set Instrument Instances, Training, Validation, and Test Sets

Variable	All videos	Training	Validation	Test
No. of trials	147	124	9	14
No. of frames	31 443	27 223	2292	1928
No. of instruments				
Suction	22 356	19 106	1862	1388
Grasper	15 943	13 576	1084	1283
String	11 917	9944	1214	759
Cottonoid	10 005	8491	610	904
Muscle	4560	3681	223	656
Tool	76	76	0	0
Drill	210	159	0	51
Scalpel	4	4	0	0

were 6 instances of double annotations or mislabeling of noninstruments as instruments (positive predictive value, 0.96). Examples of true-positive, false-positive, and false-negative results are shown in **Figure 3**. Fascia (Figure 3A) and absorbable hemostatic agent (Figure 3B) were not counted as surgical instruments.

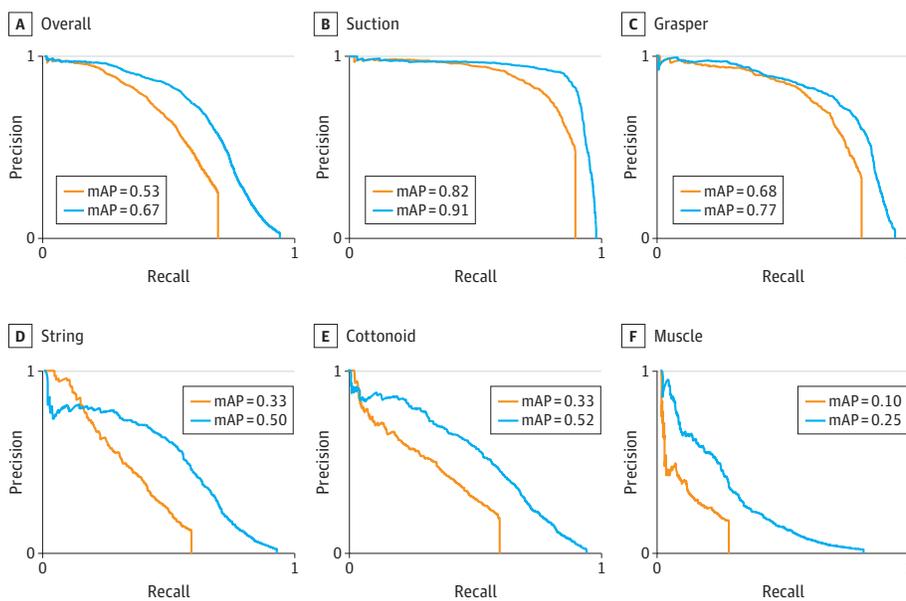
Discussion

In this quality improvement study, we introduce SOCAL, a novel video data set of surgeons managing catastrophic hemorrhage. No public video data sets that depict surgical adverse events are available, and it is unlikely that a similar data set of intraoperative hemorrhage on actual patients (instead of cadavers) could ever be created. We illustrate the utility of SOCAL as a valuable resource for surgical data science by describing 3 use cases that showcase the importance of novel test bed data sets, the importance of studying surgical adverse events, and the need for additional public data sets of surgical video.

Use of Video Data Sets as a Test Bed for Surgical Data Science Methods

SOCAL offers a test bed for methods that analyze intraoperative video and can stress test CV models because it depicts relevant and challenging conditions not present in other data sets (eg, object obscuration by bleeding). After training and testing 2 detectors (RetinaNet and YOLOv3) on the SOCAL data set, we found clinically relevant decrements in the performance of object detection in

Figure 2. Deep Learning Instrument Detection Model Performance



Orange lines represent the YOLOv3 deep neural network; blue lines represent the RetinaNet deep neural network. mAP indicates mean average precision.

Table 2. Preliminary Instrument Use Patterns and Corresponding Variance of Blood Loss

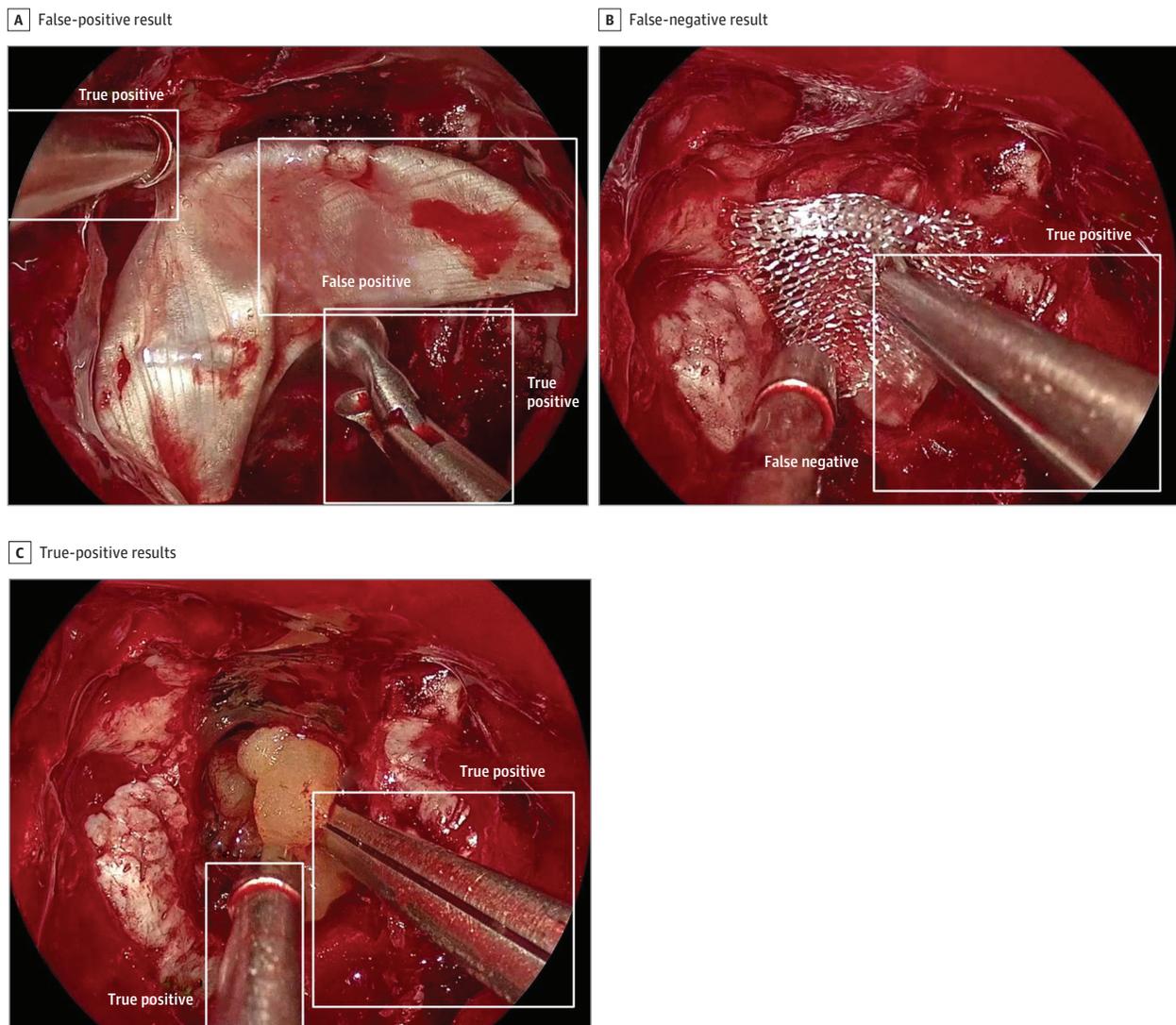
Instrument use pattern feature	Effect size	Variance of blood loss	P value
Time to hemostasis	2.9	49.84	.001
Frames with			
Grasper	-679.5	6.24	.001
Cottonoid	-684.6	6.18	.001
Muscle	-466.2	1.99	.02
Suction	-403.5	1.71	.03
String	-259.3	1.37	.05

the setting of massive hemorrhage. Rigid metal objects (eg, suction and grasper) were more easily detected than malleable and discolored objects (eg, cottonoid and string). These decrements were not seen in frames without bleeding. We obtained comparative evaluations of the 2 networks and were able to identify the best-performing DNN within this context. The overall performance of these models using SOCAL can now be compared with previous results of surgical detection in lower-fidelity or more constrained surgical tasks.^{15,21,42,43} Validation using the SOCAL video may allow for future stepwise improvements in the validity of surgical object detection.

SOCAL can be used to perform core machine learning tasks, such as instrument detection and instrument movement description.⁷ We show that even rudimentary instrument movement characteristics (eg, seconds that the suction instrument is visible) are associated with operative performance. These rudimentary benchmarks will be improved by more sophisticated models that quantify surgical movement to describe surgical signatures.^{7,13}

We validated a SOCAL-trained DNN using real case video, demonstrating a new pathway for augmenting small surgical data sets. This transfer learning approach uses larger surgical data sets to train and test a model and then validates findings on a smaller set of intraoperative video use cases.

Figure 3. Instrument Detection Results



A total of 138 instruments were identified (true-positive results), 6 noninstruments were identified (false-positive results), and 41 instruments were missed (false-negative results).

Laboratory or simulation data sets may be promising for model training and testing because we can replicate adverse events in high-fidelity simulation without waiting for patients to be harmed, and these videos may be easier to produce at scale. Then, once trained on simulator-based data, models can be validated on a markedly smaller number of operative video frames, solving for institutional and medicolegal sensitivities.

Instrument Annotations and Outcome Labels in Surgical Video Data Sets

SOCAL includes data on surgeon characteristics and performance. Throughout surgical data science, researchers are interested in exploring the associations among surgeon characteristics, tool movement, task outcome, and, eventually, patient outcome, complication occurrence, and technical competency.⁴⁴ Surgical video is the primary data source for this information for most operations aside from a few robotic-assisted procedures.⁴⁵⁻⁴⁷ Despite its importance, no publicly available data sets exist of surgical video labeled with clinically meaningful outcomes.

Using SOCAL, we can quantify the association between instrument use patterns and task outcomes and provide this feedback to surgeons. The criterion standard for feedback to surgeons is expert review, but the scarcity of experts limits scalability.^{1,48-50} The generation of summaries of surgical actions using video-based instrument use patterns combined with other objectively validated rating scales⁵¹ may provide surgeons with meaningful feedback. SOCAL depicts a wide heterogeneity of surgeon skill levels from junior trainees to world experts and contains paired records of surgeons before and after expert coaching. Investigators can use SOCAL to explore the association between surgeon experience and outcomes and between surgeon coaching and changes in instrument movement. SOCAL illustrates the need for future data sets to contain these labels.

Studying Surgical Adverse Events

SOCAL is the only instrument-annotated surgical video data set that depicts catastrophic hemorrhage, a surgical adverse event. Hemorrhage control tasks are valuable to study because they are life threatening and occur in all surgical fields.⁵²⁻⁵⁴ These tasks are well suited for CV methods because their outcomes can be labeled as success or failure and blood loss can be quantified.^{55,56} Hemorrhage also introduces a degree of realism and camera obscuration that is not captured in benchtop simulators or routine procedures.^{11,22} Finally, a direct association is seen between surgical skill and ability to achieve hemostasis, which is often exacerbated during high-stakes complication management when surgeons may potentially deviate from standard techniques because of stress, inexperience, or lack of specific skills.^{30,35} Using SOCAL, we saw that neither surgeon confidence nor experience was strongly associated with task success or blood loss in this hemorrhage control task,³⁰ whereas surgeon instrument use characteristics were strongly associated with task success and blood loss.

Although the specific hemorrhage conditions observed in SOCAL are rare, hemorrhage control is crucial for all surgical procedures. Accordingly, SOCAL-validated detectors may have improved performance in the setting of hemorrhage or extensive blood product deposition as can be seen in cardiac, vascular, and some neurosurgical scenarios.

Need for Publicly Available Data Sets

The current landscape of surgical video analysis consists of individual surgeon-research teams that developed databases for intraoperative video analysis with limited overlap in size, aims, types of videos recorded, annotation styles, and overall functionality (eTable 1 in the [Supplement](#)). The development of intraoperative video databases faces many challenges. There is often a gap between surgeons who hold intraoperative video and engineers who can use these data. Converting raw video into an annotated, usable data set is tedious and expensive. Even if stripped of protected health information, intraoperative video may be prone to institutional or individual resistance to sharing. We hope to overcome the limitations of other data sets by making the video, annotations, outcomes, and code for development publicly available as scaffolds for other groups.

We followed the NeurIPS 2021 Code and Data Submission Guidelines such that our data set adheres to best practices and ensures accessibility and reproducibility of results.⁴¹ We encourage further exploration by other groups and hope the release of this data set spurs the development of other intraoperative video data sets.

Limitations

This study has several limitations. The analysis presented serves to illustrate the utility of the SOCAL data set and leaves much room for further development of object detectors, performance metric evaluation, and outcomes associations. Many novel DNN architectures have been described in the literature but are not validated in the setting of bleeding. SOCAL could serve as a test bed for these models in future work (eg, more advanced object detection systems that use temporal sequences [long short-term memory and 3-dimensional CNN^{5,57}]). Hyperparameters, such as intersection over union threshold and frame rate for detection, were not thoroughly explored in this pilot analysis and require further experimentation. SOCAL only contains data from endoscopic endonasal neurosurgery; other surgical fields may have unique instruments and camera characteristics that could challenge SOCAL-trained networks. Further description and clinical characterization of the detected features have yet to be performed.

Conclusions

The lack of videos of adverse events creates a data set bias that hampers surgical data science. SOCAL, the first publicly accessible video data set of surgeons managing catastrophic hemorrhage, was developed to meet this need. SOCAL includes instrument annotations, surgeon experience data, and task outcomes. SOCAL serves as a unique test bed for surgical data science techniques, such as object detection, performance metric derivation, and outcome associations using deep learning models.

ARTICLE INFORMATION

Accepted for Publication: January 31, 2022.

Published: March 21, 2022. doi:[10.1001/jamanetworkopen.2022.3177](https://doi.org/10.1001/jamanetworkopen.2022.3177)

Open Access: This is an open access article distributed under the terms of the [CC-BY License](https://creativecommons.org/licenses/by/4.0/). © 2022 Kugener G et al. *JAMA Network Open*.

Corresponding Author: Guillaume Kugener, MEng, Department of Neurosurgery, Keck School of Medicine of the University of Southern California, 1200 N State St, Ste 3300, Los Angeles, CA 90033 (kugener@usc.edu).

Author Affiliations: Department of Neurosurgery, Keck School of Medicine of the University of Southern California, Los Angeles (Kugener, Pangal, Cardinal, Collet, Lechtholz-Zey, Lasky, Sundaram, Markarian, Roshannai, Sinha, Zada); Department of Computer Science, Viterbi School of Engineering, University of Southern California, Los Angeles (Zhu); Department of Operations Research and Information Engineering, Cornell University, Ithaca, New York (Han); Department of Mathematics, University of Toronto, Toronto, Ontario, Canada (Papayan); Center for Robotic Simulation and Education, USC Institute of Urology, Keck School of Medicine of the University of Southern California, Los Angeles (Hung); Department of Computer Science and Mathematics, California Institute of Technology, Pasadena (Anandkumar); Department of Otolaryngology, Keck School of Medicine of the University of Southern California, Los Angeles (Wrobel); Division of Neurosurgery, Center for Neuroscience, Children's National Hospital, Washington, DC (Donoho).

Author Contributions: Dr Donoho had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Concept and design: Kugener, Pangal, Roshannai, Papayan, Hung, Anandkumar, Wrobel, Zada, Donoho.

Acquisition, analysis, or interpretation of data: Kugener, Pangal, Cardinal, Collet, Lechtholz-Zey, Lasky, Sundaram, Markarian, Zhu, Roshannai, Sinha, Han, Wrobel, Donoho.

Drafting of the manuscript: Kugener, Pangal, Lechtholz-Zey, Sundaram, Roshannai, Donoho.

Critical revision of the manuscript for important intellectual content: Kugener, Pangal, Cardinal, Collet, Lasky, Markarian, Zhu, Sinha, Han, Papyan, Hung, Anandkumar, Wrobel, Zada, Donoho.

Statistical analysis: Kugener, Pangal, Collet, Lechtholz-Zey, Zhu, Roshannai, Anandkumar, Donoho.

Obtained funding: Zada.

Administrative, technical, or material support: Kugener, Pangal, Lasky, Sundaram, Markarian, Zhu, Sinha, Han, Wrobel, Zada, Donoho.

Supervision: Kugener, Pangal, Papyan, Hung, Wrobel, Zada, Donoho.

Conflict of Interest Disclosures: None reported.

REFERENCES

1. Greenberg CC, Dombrowski J, Dimick JB. Video-based surgical coaching: an emerging approach to performance improvement. *JAMA Surg.* 2016;151(3):282-283. doi:10.1001/jamasurg.2015.4442
2. Hashimoto DAM, Rosman G, Rus D, Meireles OR. Surgical video in the age of big data. *Ann Surg.* 2018;268(6):e47-e48. doi:10.1097/SLA.0000000000002493
3. Hung AJ, Chen J, Gill IS. Automated performance metrics and machine learning algorithms to measure surgeon performance and anticipate clinical outcomes in robotic surgery. *JAMA Surg.* 2018;153(8):770-771. doi:10.1001/jamasurg.2018.1512
4. Kitaguchi D, Takeshita N, Matsuzaki H, et al. Real-time automatic surgical phase recognition in laparoscopic sigmoidectomy using the convolutional neural network-based deep learning approach. *Surg Endosc.* 2020;34(11):4924-4931. doi:10.1007/s00464-019-07281-0
5. Kitaguchi D, Takeshita N, Matsuzaki H, Igaki T, Hasegawa H, Ito M. Development and validation of a 3-dimensional convolutional neural network for automatic surgical skill assessment based on spatiotemporal video analysis. *JAMA Netw Open.* 2021;4(8):e2120786. doi:10.1001/jamanetworkopen.2021.20786
6. Ward TM, Mascagni P, Madani A, Padoy N, Perretta S, Hashimoto DA. Surgical data science and artificial intelligence for surgical education. *J Surg Oncol.* 2021;124(2):221-230. doi:10.1002/jso.26496
7. Ward TM, Mascagni P, Ban Y, et al. Computer vision in surgery. *Surgery.* 2021;169(5):1253-1256. doi:10.1016/j.surg.2020.10.039
8. Chadebecq F, Vasconcelos F, Mazomenos E, Stoyanov D. Computer vision in the surgical operating room. *Visc Med.* 2020;36(6):456-462. doi:10.1159/000511934
9. Staartjes VE, Volokitin A, Regli L, Konukoglu E, Serra C. Machine vision for real-time intraoperative anatomic guidance: a proof-of-concept study in endoscopic pituitary surgery. *Oper Neurosurg (Hagerstown).* 2021;21(4):242-247. doi:10.1093/ons/opab187
10. Pangal DJ, Kugener G, Cardinal T, et al. Use of surgical video-based automated performance metrics to predict blood loss and success of simulated vascular injury control in neurosurgery: a pilot study. *J Neurosurg.* 2021. Published online December 31, 2021. doi:10.3171/2021.10.JNS211064
11. Twinanda AP, Shehata S, Mutter D, Marescaux J, de Mathelin M, Padoy N. EndoNet: a deep architecture for recognition tasks on laparoscopic videos. *IEEE Trans Med Imaging.* 2017;36(1):86-97. doi:10.1109/TMI.2016.2593957
12. Hashimoto DAM, Rosman G, Witkowski ERM, et al. Computer vision analysis of intraoperative video: automated recognition of operative steps in laparoscopic sleeve gastrectomy. *Ann Surg.* 2019;270(3):414-421. doi:10.1097/SLA.0000000000003460
13. Ward TM, Hashimoto DA, Ban Y, et al. Automated operative phase identification in peroral endoscopic myotomy. *Surg Endosc.* 2021;35(7):4008-4015. doi:10.1007/s00464-020-07833-9
14. Grammatikopoulou M, Flouty E, Kadkhodamohammadi A, et al. CaDIS: cataract dataset for surgical RGB-image segmentation. *Med Image Anal.* 2021;71:102053. doi:10.1016/j.media.2021.102053
15. Jin A, Yeung S, Jopling J, et al. Tool detection and operative skill assessment in surgical videos using region-based convolutional neural networks. *arXiv.* Preprint posted online July 22, 2018. Accessed February 13, 2022. <https://arxiv.org/abs/1802.08774>
16. Bawa VS, Singh G, Kaping A F, et al. ESAD: endoscopic surgeon action detection dataset. *arXiv.* Preprint posted online June 12, 2020. Accessed February 13, 2022. <https://arxiv.org/abs/2006.07164>
17. Speidel S, Maier-Hein L, Stoyanov D, et al. Endoscopic vision challenge 2021. Paper presented at: 24th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI 2021); March 2, 2021; Strasbourg, France.

18. Roß T, Reinke A, Full PM, et al. Comparative validation of multi-instance instrument segmentation in endoscopy: results of the ROBUST-MIS 2019 challenge. *Med Image Anal*. 2021;70:101920. doi:10.1016/j.media.2020.101920
19. Maier-Hein L, Wagner M, Ross T, et al. Heidelberg colorectal data set for surgical data science in the sensor operating room. *Sci Data*. 2021;8(1):101. doi:10.1038/s41597-021-00882-2
20. Allan M, Shvets A, Kurmann T, et al. 2017 Robotic instrument segmentation challenge. *arXiv*. Preprint posted online February 21, 2019. Accessed September 30, 2021. <https://arxiv.org/abs/1902.06426>
21. Sarikaya D, Corso JJ, Guru KA. Detection and localization of robotic tools in robot-assisted surgery videos using deep neural networks for region proposal and detection. *IEEE Trans Med Imaging*. 2017;36(7):1542-1549. doi:10.1109/TMI.2017.2665671
22. Gao Y, Vedula SS, Reiley CE, et al. JHU-ISI Gesture and Skill Assessment Working Set (JIGSAWS): a surgical activity dataset for human motion modeling. Accessed February 8, 2022. https://cirl.lcsr.jhu.edu/research/hmm/datasets/jigsaws_release/
23. Bouget D, Benenson R, Omran M, Riffaud L, Schiele B, Jannin P. Detecting surgical tools by modelling local appearance and global shape. *IEEE Trans Med Imaging*. 2015;34(12):2603-2617. doi:10.1109/TMI.2015.2450831
24. Ashraf A, Khan S, Bhagwat N, Chakravarty M, Taati B. Learning to unlearn: building immunity to dataset bias in medical imaging studies. *arXiv*. Preprint posted online December 2, 2018. Accessed February 8, 2022. <https://arxiv.org/abs/1812.01716>
25. Laws ER Jr. Vascular complications of transsphenoidal surgery. *Pituitary*. 1999;2(2):163-170. doi:10.1023/A:1009951917649
26. Chin OY, Ghosh R, Fang CH, Baredes S, Liu JK, Eloy JA. Internal carotid artery injury in endoscopic endonasal surgery: a systematic review. *Laryngoscope*. 2016;126(3):582-590. doi:10.1002/lary.25748
27. Ciric I, Ragin A, Baumgartner C, Pierce D. Complications of transsphenoidal surgery: results of a national survey, review of the literature, and personal experience. *Neurosurgery*. 1997;40(2):225-236. doi:10.1097/00006123-199702000-00001
28. James HK, Pattison GTR, Fisher JD, Griffin D. Cadaveric simulation versus standard training for postgraduate trauma and orthopaedic surgical trainees: protocol for the CAD:TRAUMA study multicentre randomised controlled educational trial. *BMJ Open*. 2020;10(9):e037319. doi:10.1136/bmjopen-2020-037319
29. Strickland BA, Ravina K, Kammen A, et al. The use of a novel perfusion-based human cadaveric model for simulation of dural venous sinus injury and repair. *Oper Neurosurg (Hagerstown)*. 2020;19(3):E269-E274. doi:10.1093/ons/opz424
30. Donoho DA, Pangal DJ, Kugener G, et al. Improved surgeon performance following cadaveric simulation of internal carotid artery injury during endoscopic endonasal surgery: training outcomes of a nationwide prospective educational intervention. *J Neurosurg*. 2021;1-9. doi:10.3171/2020.9.JNS202672
31. Donoho DA, Johnson CE, Hur KT, et al. Costs and training results of an objectively validated cadaveric perfusion-based internal carotid artery injury simulation during endoscopic skull base surgery. *Int Forum Allergy Rhinol*. 2019;9(7):787-794. doi:10.1002/alr.22319
32. Shen J, Hur K, Zhang Z, et al. Objective validation of perfusion-based human cadaveric simulation training model for management of internal carotid artery injury in endoscopic endonasal sinus and skull base surgery. *Oper Neurosurg (Hagerstown)*. 2018;15(2):231-238. doi:10.1093/ons/opx262
33. Zada G, Bakhsheshian J, Pham M, et al. Development of a perfusion-based cadaveric simulation model integrated into neurosurgical training: feasibility based on reconstitution of vascular and cerebrospinal fluid systems. *Oper Neurosurg (Hagerstown)*. 2018;14(1):72-80. doi:10.1093/ons/opx074
34. Kugener G, Pangal DJ, Zada G. Simulated outcomes following carotid artery laceration. *figshare*. Published online August 10, 2021. Accessed March 2, 2022. https://figshare.com/articles/dataset/Simulated_Outcomes_following_Carotid_Artery_Laceration/15132468/1
35. Pham M, Kale A, Marquez Y, et al. A perfusion-based human cadaveric model for management of carotid artery injury during endoscopic endonasal skull base surgery. *J Neurol Surg B Skull Base*. 2014;75(5):309-313. doi:10.1055/s-0034-1372470
36. Pangal DJ, Kugener G, Shahrestani S, Attenello F, Zada G, Donoho DA. A guide to annotation of neurosurgical intraoperative video for machine learning analysis and computer vision. *World Neurosurg*. 2021;150:26-30. doi:10.1016/j.wneu.2021.03.022
37. Microsoft/VoTT. Microsoft; 2020. Accessed July 7, 2020. <https://github.com/microsoft/VoTT>
38. Lin TY, Goyal P, Girshick R, He K, Dollar P. Focal Loss for Dense Object Detection. Accessed October 23, 2020. https://openaccess.thecvf.com/content_iccv_2017/html/Lin_Focal_Loss_for_ICCV_2017_paper.html

39. Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: unified, real-time object detection. *arXiv*. Preprint posted online May 9, 2016. Accessed February 13, 2022. <https://arxiv.org/abs/1506.02640>
40. Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L. ImageNet: a large-scale hierarchical image database. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*. 2009:248-255.
41. Paper Information/Code Submission Policy. Accessed October 14, 2021. <https://nips.cc/Conferences/2021/PaperInformation/CodeSubmissionPolicy>
42. Liu Y, Zhao Z, Chang F, Hu S. An anchor-free convolutional neural network for real-time surgical tool detection in robot-assisted surgery. *IEEE Access*. 2020;8:78193-78201. doi:10.1109/ACCESS.2020.2989807
43. Yamazaki Y, Kanaji S, Matsuda T, et al. Automated surgical instrument detection from laparoscopic gastrectomy video images using an open source convolutional neural network platform. *J Am Coll Surg*. 2020;230(5):725-732.e1. doi:10.1016/j.jamcollsurg.2020.01.037
44. Hung AJ, Liu Y, Anandkumar A. Deep learning to automate technical skills assessment in robotic surgery. *JAMA Surg*. 2021;156(11):1059-1060. doi:10.1001/jamasurg.2021.3651
45. Hung AJ, Chen J, Ghodoussipour S, et al. A deep-learning model using automated performance metrics and clinical features to predict urinary continence recovery after robot-assisted radical prostatectomy. *BJU Int*. 2019;124(3):487-495. doi:10.1111/bju.14735
46. Chen J, Oh PJ, Cheng N, et al. Use of automated performance metrics to measure surgeon performance during robotic vesicourethral anastomosis and methodical development of a training tutorial. *J Urol*. 2018;200(4):895-902. doi:10.1016/j.juro.2018.05.080
47. Ghodoussipour S, Reddy SS, Ma R, Huang D, Nguyen J, Hung AJ. An objective assessment of performance during robotic partial nephrectomy: validation and correlation of automated performance metrics with intraoperative outcomes. *J Urol*. 2021;205(5):1294-1302. doi:10.1097/JU.0000000000001557
48. Birkmeyer JD, Finks JF, O'Reilly A, et al; Michigan Bariatric Surgery Collaborative. Surgical skill and complication rates after bariatric surgery. *N Engl J Med*. 2013;369(15):1434-1442. doi:10.1056/NEJMsa1300625
49. Augestad KM, Butt K, Ignjatovic D, Keller DS, Kiran R. Video-based coaching in surgical education: a systematic review and meta-analysis. *Surg Endosc*. 2020;34(2):521-535. doi:10.1007/s00464-019-07265-0
50. Brajcich BC, Stulberg JJ, Palis BE, et al. Association between surgical technical skill and long-term survival for colon cancer. *JAMA Oncol*. 2021;7(1):127-129. doi:10.1001/jamaoncol.2020.5462
51. Martin JA, Regehr G, Reznick R, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. *Br J Surg*. 1997;84(2):273-278. doi:10.1046/j.1365-2168.1997.02502.x
52. Quasarano RT, Kashef M, Sherman SJ, Hagglund KH. Complications of gynecologic laparoscopy. *J Am Assoc Gynecol Laparosc*. 1999;6(3):317-321. doi:10.1016/S1074-3804(99)80068-1
53. Sandadi S, Johannigman JA, Wong VL, Blebea J, Altose MD, Hurd WW. Recognition and management of major vessel injury during laparoscopy. *J Minim Invasive Gynecol*. 2010;17(6):692-702. doi:10.1016/j.jmig.2010.06.005
54. England EC, Spear CR, Huang DD, et al. REBOA as a rescue strategy for catastrophic vascular injury during robotic surgery. *J Robot Surg*. 2020;14(3):473-477. doi:10.1007/s11701-019-01011-3
55. Jukes AK, Mascarenhas A, Murphy J, et al. Stress response and communication in surgeons undergoing training in endoscopic management of major vessel hemorrhage: a mixed methods study. *Int Forum Allergy Rhinol*. 2017;7(6):576-583. doi:10.1002/alr.21941
56. Fruhwirth J, Koch G, Mischinger HJ, Werkgartner G, Tesch NP. Vascular complications in minimally invasive surgery. *Surg Laparosc Endosc*. 1997;7(3):251-254. doi:10.1097/OO019509-199706000-00016
57. Yengera G, Mutter D, Marescaux J, Padoy N. Less is more: surgical phase recognition with less annotations through self-supervised pre-training of CNN-LSTM networks. *arXiv*. Preprint posted online May 22, 2018. <https://arxiv.org/abs/1805.08569>

SUPPLEMENT.

eMethods. Supplemental Methods

eTable 1. Publicly Available Tool-Annotated Surgical Video Data Sets Across Surgical Specialties

eTable 2. Baseline Demographic and Outcome Characteristics of Surgeon-Trials Contained in the SOCAL Video Database