Contents lists available at ScienceDirect

Surgery

journal homepage: www.elsevier.com/locate/surg

Computer vision in surgery

Thomas M. Ward, MD[a], Pietro Mascagni, MD[b,c], Yutong Ban, PhD[a,d], Guy Rosman, PhD[a,d], Nicolas Padoy, PhD[b], Ozanan Meireles, MD[a], Daniel A. Hashimoto, MD[a,*]

[a] Surgical Artificial Intelligence and Innovation Laboratory, Massachusetts General Hospital, Harvard Medical School, Boston, MA
[b] ICube, University of Strasbourg, CNRS, IHU Strasbourg, France
[c] Fondazione Policlinico Universitario A. Gemelli IRCCS, Rome, Italy
[d] Distributed Robotics Laboratory, Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA

ARTICLE INFO

Article history:
Accepted 10 October 2020
Available online 1 December 2020

ABSTRACT

The fields of computer vision (CV) and artificial intelligence (AI) have undergone rapid advancements in the past decade, many of which have been applied to the analysis of intraoperative video. These advances are driven by wide-spread application of deep learning, which leverages multiple layers of neural networks to teach computers complex tasks. Prior to these advances, applications of AI in the operating room were limited by our relative inability to train computers to accurately understand images with traditional machine learning (ML) techniques. The development and refining of deep neural networks that can now accurately identify objects in images and remember past surgical events has sparked a surge in the applications of CV to analyze intraoperative video and has allowed for the accurate identification of surgical phases (steps) and instruments across a variety of procedures. In some cases, CV can even identify operative phases with accuracy similar to surgeons. Future research will likely expand on this foundation of surgical knowledge using larger video datasets and improved algorithms with greater accuracy and interpretability to create clinically useful AI models that gain widespread adoption and augment the surgeon's ability to provide safer care for patients everywhere.

Highlights

Topic: Intraoperative applications of computer vision (CV) and artificial intelligence (AI)

Purpose: To provide today's surgeon with a summary of the history and recent advances in the field of CV and how CV is applied in the analysis of intraoperative video.

State of the art: CV and AI can accurately identify operative phases (steps) and tools in surgical video. In some cases, CV can identify operative phases with accuracy similar to surgeons.

Knowledge gaps: Research still needs to be done to develop algorithms that can better remember prior events in surgery to improve identifications and accurately identify deformable soft tissue structures.

Technology gaps: Training the algorithms requires a relatively large number of surgeon-annotated images and high-powered computing, which are not readily available in many operating rooms.

Future directions: Future research will build on the foundation of surgical phase and instrument recognition using larger video datasets and improved algorithms with greater accuracy and interpretability to create clinically useful and impactful algorithms that can augment surgeons and lead to safer surgery.

Introduction

Imagine operating under the watchful gaze of a surgeon who has done thousands of cases, seen every complication, never tires, and is ready to provide guidance at any moment. Artificial intelligence (AI) hopes to one day make this a reality. AI found little traction in the operating room before the last decade, because we could not teach computers the most important sense for a surgeon: sight. Research instead taught computers surgery indirectly with machine learning (ML) techniques that used data streams from the operating room, like electrosurgical energy device activation and human-annotated presence of tools. Without sight, AI was

[*] Reprint requests: Daniel A. Hashimoto, MD, MS, Surgical AI & Innovation Laboratory, 15 Parkman Street, WAC460, Boston, MA 02114.
E-mail address: dahashimoto@mgh.harvard.edu (D.A. Hashimoto).
Twitter: @laparoscopes

https://doi.org/10.1016/j.surg.2020.10.039
0039-6060/© 2020 Elsevier Inc. All rights reserved.

operating blindly, and its procedural understanding was inflexible and limited.

Recent advancements in computer vision (CV) have used deep learning to give computers human-like image recognition ability.[1] Deep learning processes information through networks of multiple neuron layers, which allows computers to learn complex tasks.[2] One common neuronal arrangement, the convolutional neural networks (CNNs), has a structure similar to a human's visual cortex, which facilitates learning the complex task of visual object recognition particularly well.[3] The robust nature of CNNs is important. For example, it is possible to train a traditional ML algorithm to recognize a particular instrument; however, this training is brittle and will likely fail once visual conditions change. Surgery is filled with visual disturbances, like smoke, blood, varying port site locations, and blurry cameras. CNNs, like humans, can more readily cope with these conditions and make appropriate identifications.

Using CNNs, researchers can now reliably teach computers to recognize objects in surgical images. Surgery is a process where past events influence future ones and is, therefore, best represented with video rather than a few photographs. However, researchers encountered performance limitations when they applied image recognition techniques to surgical video; computers could not fully understand a time point in the surgery because they had no context. They had a visual cortex but lacked the memory part of their "brains." This shortcoming was improved by adding long short-term memory (LSTM) neural networks, a "memory cortex" equivalent, to improve performance.[4]

The combination of CNN and LSTM gives computers the visual and temporal learning abilities to begin to understand surgery. Current research focuses on teaching computers 2 primary features of surgery: identification of the surgical phase (step) and identification/tracking of surgical tools. Even though automatically recognizing phases or instruments may not be valuable per se, algorithms have to reliably infer these fundamental surgical elements to understand the surgical context and provide information of higher surgical value.[5] The following sections will describe AI advances in these areas, which follows the previously described evolution from classical ML methods without video input, to visual understanding through CNNs, the addition of temporal memory with LSTMs, and more recent advances.

Automated identification of surgical phases

Teaching algorithms to understand operative events has been a focal point of surgical AI research. Similar to a medical student learning the operation's phase, researchers first trained algorithms to automatically identify the phase at a given time point. Prior to the CV revolution, computers had difficulty using video data and instead relied on ML techniques to learn from "computer-friendly" data streams, such as human-annotated binary tool presence (ie, a tool present in the surgical field at a certain time) and activation times of electrosurgical energy devices. ML techniques like hidden Markov models (HMM), random forests, and support vector machines were applied to laparoscopic surgery (whose operative phase is often predictable from the limited tools used) with successes in cholecystectomy and hysterectomy.[6–10] Some of these early successes also incorporated basic video processing to determine visually distinct events, like extracorporeal camera cleaning.[6]

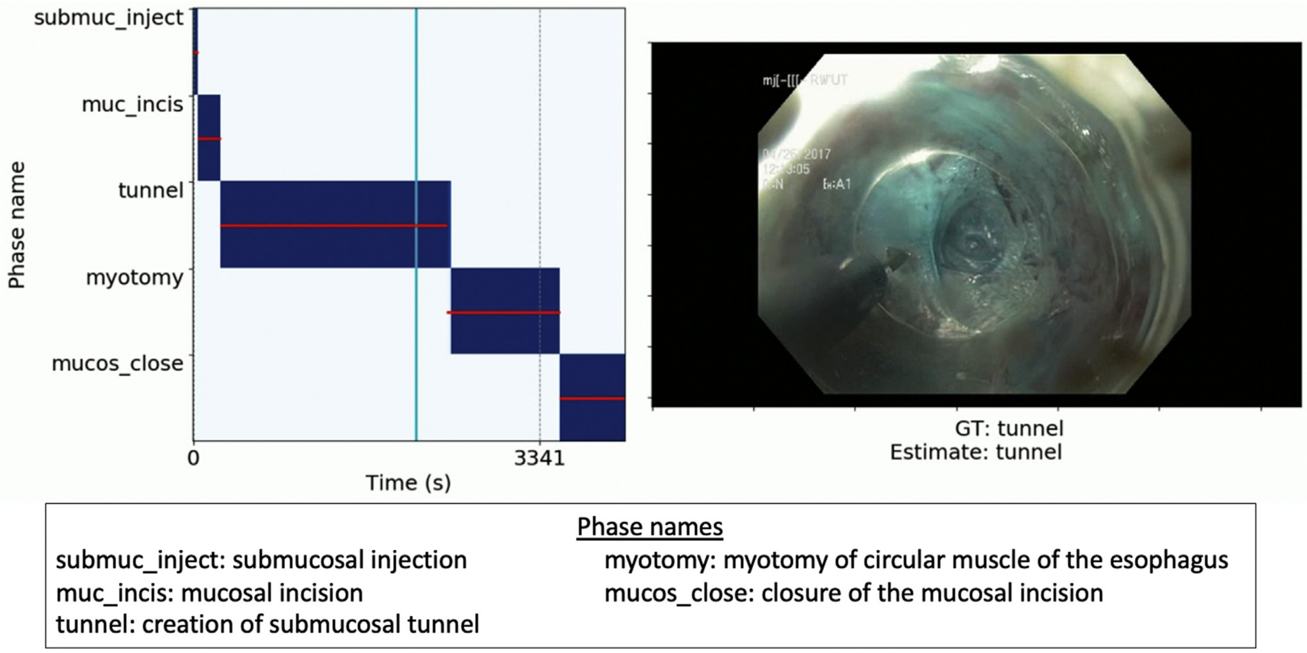
Prior to widespread use of CNNs, processing video data was a herculean task for traditional ML algorithms; 1 minute of high-resolution surgical video contains 25 times more data than a high-resolution computed tomography scan.[11] To circumvent this, initial efforts using surgical video relied on techniques to first reduce the amount of information in an image then learn from this reduced data. Efforts included phase identification in cholecystectomy, sleeve gastrectomy, and ophthalmologic procedures.[12–14] Like most CV methods of the time, they required a large amount of "hand-crafting" to achieve high accuracy on a particular collection of videos, which hampered their wide-spread utility.

CNNs allowed researchers to create accurate and generalizable AI algorithms to recognize surgical phases. Twinanda et al's landmark work in 2016 created EndoNet, a CNN capable of highly accurate phase recognition in laparoscopic cholecystectomy, along with the first publicly available reference dataset Cholec80.[15] They found that their identification performance was hampered by the CNN's lack of temporal context, which was improved by adding a traditional temporal ML modeling technique (HMM) to help with identification consistency. Similar works have augmented plain CNNs with HMMs or alternative techniques (like light gradient boosting machines).[16,17] A big milestone though was giving machines the temporal memory ability through LSTMs to improve their identifications through knowledge of previous surgical events. This technique found rapid uptake to create accurate phase recognition in laparoscopic (cholecystectomy, sleeve gastrectomy, and colectomy), ophthalmic, and endoscopic (peroral endoscopic myotomy) surgeries (Fig 1).[16–23] These algorithms can identify phases with 85% to 90% accuracy, which at first may seem low but actually equals inter-surgeon agreement when they annotate the same images.[16,20]

Now that reliable and accurate phase recognition has been achieved, the most recent advances have focused on incremental improvements to the CNN/LSTM base for small (1%–5%) accuracy boosts. Some efforts use additional input streams, like tool recognition to finetune phase recognition.[15,24] Others improve the memory component of the neural network, incorporating recent advances beyond LSTM to improve the computer's memory of previous events, which makes for more accurate identifications.[18,25] A last group of recent advances addressed a looming problem: data availability. Ideally, we would train these algorithms with thousands of videos. This many videos are simply not yet recorded, and even if they were, surgeons would also need to watch and label them to generate annotations from which the algorithms can learn. To ameliorate this, researchers have made training more efficient with "pre-training," where algorithms are given unlabeled surgical video data so they, independently, begin to recognize regular patterns occurring in surgery (eg, temporal ordering of different images or surgery duration).[26–28] Another tactic has been to have the machines teach themselves on machine-generated annotations created after they learn from only a few videos. Though seemingly paradoxical (machines improving by learning from what they self-generated), this method has led to improved performance.[29]

Automated identification of instruments

AI identification of instruments, just like surgical phases, underwent a stepwise evolution. Initial efforts were severely limited by the inherent difficulties in image recognition, but one success included tracking of surgical tooltips in retinal surgery.[30] The field blossomed after the incorporation of deep learning and CNNs. Reliable detection of an instrument's presence in a video frame was obtained for various laparoscopic and cataract surgeries.[15,31,32] In parallel, the field worked to discriminate precise instrument outlines in the images with newer systems, even tracking the tips of instruments in real time (Fig 2).[33–36] As seen before, incorporation of a "memory component" led to improved performance over visual (CNN) only models.[37] Even more than with phase training, training



Phase names

submuc_inject: submucosal injection
muc_incis: mucosal incision
tunnel: creation of submucosal tunnel
myotomy: myotomy of circular muscle of the esophagus
mucos_close: closure of the mucosal incision

Fig 1. Automated phase recognition in per oral endoscopic myotomy. Left half of image represents predicted operative phase in dark blue and surgeon-annotated GT in red.[16] (Color version of the figure is available online.) *GT,* ground truth.

a system to recognize instrument outlines is a time-intensive task, as the machine learns from a surgeon's manual outlines of an instrument's location. Efforts addressing this have included pretraining methods, like first having the algorithm self-teach the discovery of relevant visual features to standardize videos' colors (similar to "white balancing"), which has reduced the need for annotations by 75%.[38] Another technique trained a network from only binary tool presence annotations and obtained good results for tool localization and tracking.[39]

Future directions

The past decade's advances in deep learning have given computers the visual and memory capabilities needed to process surgical video and achieve human-like recognition of surgical phases and instruments. This ability is akin to intern year; it lays the foundation on which to build true surgical understanding. Our current technology can already perform rudimentary automated assessment, such as case difficulty assessment from phase recognition with "surgical fingerprints," operative performance assessment from real-time tool tracking, identification of critical events in surgical procedures (eg, the critical view of safety), and highlighting of key anatomy (eg, hepatocystic structures).[16,34,35,40–42]

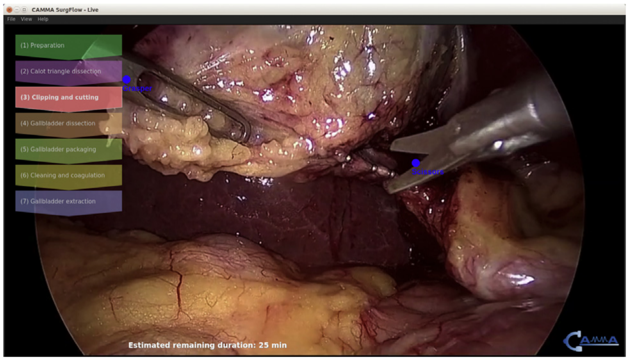

CAMMA SurgFlow - Live
File View Help
(1) Preparation
(2) Calot triangle dissection
(3) Clipping and cutting
(4) Gallbladder dissection
(5) Gallbladder packaging
(6) Cleaning and coagulation
(7) Gallbladder extraction
Grasper
Scissors
Estimated remaining duration: 25 min
CAMMA

Fig 2. Illustration of automated tool tracking in conjunction with automated prediction of remaining duration of a surgical phase.[28,29,39]

With advances in virtual reality simulation technology yielding early versions of automated surgical coaches,[43] interest has grown in investigating applications of the above-described CV functions for surgical coaching. The translation of CV to surgical education is an area of innovation with high potential for impact for both trainees and practicing surgeons. One can certainly imagine a virtual surgical coach providing formative feedback through quantitative metrics such as time spent in each operative phase,[20,25] use of instruments per operative phase, and time spent performing specific tasks such as retraction or dissection.[44] CV could also assist a virtual coach in highlighting key areas of anatomy or safe and unsafe areas of dissection, allowing surgeons to review their operations with AI-augmented feedback. While early work has begun on translating advances in CV to automated surgical coaching, additional research will be required before a virtual surgical coach of this nature is ready to leave the investigative arena.

To fully and effectively transition CV for surgery to the educational and clinical domain a few hurdles must be overcome. First, to be robust, AI models must learn from not just hundreds of 1 institution's videos but likely thousands of videos from multiple institutions. We must seek to collect a diverse dataset from different patients, practice patterns, and techniques to minimize bias and optimize generalizability. The more examples of surgical techniques, anatomic variants, and patient conditions we feed the models, the better they will be. Second, we need to develop improved metrics to measure the model's performance beyond simple accuracy. A surgeon who can identify 99% of the anatomy but fails to recognize the critical 1% of structures should receive a failing grade. Third, we need to design these models to maximize their explainability.[45] One way to improve explainability is to train the models to localize anatomy and use that in their decisions.[44] With the above directions and continued research, AI may even ultimately create that "ultimate surgical mentor" to help guide us through cases and create safer surgery for patients everywhere.

Support/Funding

Thomas Ward, Yutong Ban, Guy Rosman, Ozanan Meireles, and Daniel Hashimoto have received research support outside of this manuscript from Olympus Corporation. Guy Rosman has received research support outside of this manuscript from Toyota Research Institute.

Conflict of interest/Disclosures

Ozanan Meireles is a consultant for Medtronic and Olympus Corporation. Daniel Hashimoto is a consultant for Verily Life Sciences, Johnson & Johnson Institute, and Worrell. Pietro Mascagni and Nicolas Padoy have no conflicts of interest to declare.

References

1. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. *Adv Neural Inform Process Syst*. 2012:1097–1105.
2. Hashimoto DA, Rosman G, Rus D, Meireles OR. Artificial intelligence in surgery: promises and perils. *Ann Surg*. 2018;268:70–76.
3. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521:436–444.
4. Donahue J, Hendricks LA, Guadarrama S, et al. Long-term recurrent convolutional networks for visual recognition and description. *IEEE Trans Pattern Anal Mach Intell*. 2017;39:677–691.
5. Vercauteren T, Unberath M, Padoy N, Navab N. CAI4CAI: The rise of contextual artificial intelligence in computer-assisted interventions. *Proc IEEE*. 2020;108:198–214.
6. Padoy N, Blum T, Feussner H, Berger MO, Navab N. On-line recognition of surgical activity for monitoring in the operating room; 2008. https://www.aaai.org/Papers/IAAI/2008/IAAI08-015.pdf. Accessed September 20, 2020.
7. Padoy N, Blum T, Ahmadi SA, Feussner H, Berger MO, Navab N. Statistical modeling and recognition of surgical workflow. *Med Image Anal*. 2012;16:632–641.
8. Stauder R, Okur A, Peter L, et al. Random forests for phase detection in surgical workflow analysis. In: Stoyanov D, Collins DL, Sakuma I, Abolmaesumi P, Jannin P, eds. *Information Processing in Computer-Assisted Interventions*. New York: Springer International Publishing; 2014:148–157.
9. Meeuwsen FC, van Luyn F, Blikkendaal MD, Jansen FW, van den Dobbelsteen JJ. Surgical phase modelling in minimal invasive surgery. *Surg Endosc*. 2019;33:1426–1432.
10. Malpani A, Lea C, Chen CCG, Hager GD. System events: readily accessible features for surgical phase detection. *Int J Comput Assist Radiol Surg*. 2016;11:1201–1209.
11. Natarajan P, Frenzel JC, Smaltz DH. *Demystifying Big Data and Machine Learning for Healthcare*. Boca Raton (FL): CRC Press, Taylor & Francis Group; 2017.
12. Blum T, Feußner H, Navab N. Modeling and segmentation of surgical workflow from laparoscopic video. In: Jiang T, Navab N, Pluim JPW, Viergever MA, eds. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2010*. New York: Springer; 2010:400–407.
13. Lalys F, Bouget D, Riffaud L, Jannin P. Automatic knowledge-based recognition of low-level tasks in ophthalmological procedures. *Int J Comput Assist Radiol Surg*. 2013;8:39–49.
14. Volkov M, Hashimoto DA, Rosman G, Meireles OR, Rus D. Machine learning and coresets for automated real-time video segmentation of laparoscopic and robot-assisted surgery. *2017 IEEE International Conference on Robotics and Automation (ICRA)*. 2017:754–759.
15. 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;36:86–97.
16. Ward TM, Hashimoto DA, Ban Y, et al. Automated operative phase identification in peroral endoscopic myotomy [e-pub ahead of print]. *Surg Endosc*. 2020. https://doi.org/10.1007/s00464-020-07833-9. Accessed September 20, 2020.
17. 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;34:4924–4931.
18. Jin Y, Dou Q, Chen H, et al. SV-RCNet: Workflow recognition from surgical videos using recurrent convolutional network. *IEEE Trans Med Imaging*. 2018;37:1114–1126.
19. Zisimopoulos O, Flouty E, Luengo I, et al. DeepPhase: Surgical phase recognition in CATARACTS videos. In: Frangi AF, Schnabel JA, Davatzikos C, López Alberola C, Fichtinger G, eds. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*. New York: Springer International Publishing; 2018:265–272.
20. Hashimoto DA, Rosman G, Witkowski ER, et al. Computer vision analysis of intraoperative video: automated recognition of operative steps in laparoscopic sleeve gastrectomy. *Ann Surg*. 2019;270:414–421.
21. Kitaguchi D, Takeshita N, Matsuzaki H, et al. Automated laparoscopic colorectal surgery workflow recognition using artificial intelligence: Experimental research. *Int J Surg*. 2020;79:88–94.
22. Twinanda AP. Vision-based approaches for surgical activity recognition using laparoscopic and RBGD videos; 2017. https://www.theses.fr/2017STRAD005. Accessed September 10, 2020.
23. Twinanda AP, Mutter D, Marescaux J, de Mathelin M, Padoy N. Single- and multi-task architectures for surgical workflow challenge at M2CAI 2016; 2016. http://arxiv.org/abs/1610.08844. Accessed September 10, 2020.
24. Jin Y, Li H, Dou Q, et al. Multi-task recurrent convolutional network with correlation loss for surgical video analysis. *Med Image Anal*. 2020;59:101572.
25. Ban Y, Rosman G, Ward T, et al. Aggregating long-term context for learning surgical workflows; 2020. http://arxiv.org/abs/2009.00681. Accessed September 3, 2020.
26. Bodenstedt S, Rivoir D, Jenke A, et al. Active learning using deep Bayesian networks for surgical workflow analysis. *Int J Comput Assist Radiol Surg*. 2019;14:1079–1087.
27. 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; 2018. http://arxiv.org/abs/1805.08569. Accessed September 14, 2019.
28. Twinanda AP, Yengera G, Mutter D, Marescaux J, Padoy N. RSDNet: Learning to predict remaining surgery duration from laparoscopic videos without manual annotations. *IEEE Trans Med Imaging*. 2019;38:1069–1078.
29. Yu T, Mutter D, Marescaux J, Padoy N. Learning from a tiny dataset of manual annotations: a teacher/student approach for surgical phase recognition; 2019. http://icube-publis.unistra.fr/4-YMMP19. Accessed.
30. Richa R, Balicki M, Meisner E, Sznitman R, Taylor R, Hager G. Visual tracking of surgical tools for proximity detection in retinal surgery. In: Taylor RH, Yang GZ, eds. *International Conference on Information Processing in Computer-Assisted Interventions*. New York: Springer International Publishing; 2011:55–66.
31. Hu X, Yu L, Chen H, Qin J, Heng PA. AGNet: Attention-guided network for surgical tool presence detection. In: Cardoso MJ, Arbel T, Carneiro G, et al., eds. *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. New York: Springer International Publishing; 2017:186–194.
32. Al Hajj H, Lamard M, Conze PH, et al. CATARACTS: Challenge on automatic tool annotation for cataRACT surgery. *Med Image Anal*. 2019;52:24–41.
33. García-Peraza-Herrera LC, Li W, Fidon L, et al. ToolNet: Holistically-nested real-time segmentation of robotic surgical tools. *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 2017:5717-5722.
34. Jin A, Yeung S, Jopling J, 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)*. 2018:691-699.
35. 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;230:725–732.e1.
36. Laina I, Rieke N, Rupprecht C, et al. Concurrent segmentation and localization for tracking of surgical instruments. In: Descoteaux M, Maier-Hein L, Franz A, Jannin P, Collins DL, Duchesne S, eds. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2017*. New York: Springer International Publishing; 2017:664–672.
37. Attia M, Hossny M, Nahavandi S, Asadi H. Surgical tool segmentation using a hybrid deep CNN-RNN auto encoder-decoder. *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. 2017:3373-3378.
38. Ross T, Zimmerer D, Vemuri A, et al. Exploiting the potential of unlabeled endoscopic video data with self-supervised learning. *Int J Comput Assist Radiol Surg*. 2018;13:925–933.
39. Nwoye CI, Mutter D, Marescaux J, Padoy N. Weakly supervised convolutional LSTM approach for tool tracking in laparoscopic videos. *Int J Comput Assist Radiol Surg*. 2019;14:1059–1067.
40. Tokuyasu T, Iwashita Y, Matsunobu Y, et al. Development of an artificial intelligence system using deep learning to indicate anatomical landmarks during laparoscopic cholecystectomy [e-pub ahead of print]. *Surg Endosc*. 2020. https://doi.org/10.1007/s00464-020-07548-x. Accessed September 20, 2020.
41. Mascagni P, Vardazaryan A, Alapatt D, et al. Artificial intelligence for surgical safety: Automatic assessment of the critical view of safety in laparoscopic cholecystectomy using deep learning. Ann Surg. 2020. https://doi.org/10.1097/SLA.0000000000004351. Accessed September 20, 2020.
42. Korndorffer JR, Hawn MT, Spain DA, et al. Situating artificial intelligence in surgery: a focus on disease severity. *Ann Surg*. 2020;272:523–528.
43. Malpani A, Vedula SS, Lin HC, Hager GD, Taylor RH. Effect of real-time virtual reality-based teaching cues on learning needle passing for robot-assisted minimally invasive surgery: a randomized controlled trial. *Int J Comput Assist Radiol Surg*. 2020;15:1187–1194.
44. Nwoye CI, Gonzalez C, Yu T, et al. Recognition of instrument-tissue interactions in endoscopic videos via action triplets. In: Martel AL, Abolmaesumi P, Stoyanov D, et al., eds. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2020*. New York: Springer International Publishing; 2020:364–374.
45. Gordon L, Grantcharov T, Rudzicz F. Explainable artificial intelligence for safe intraoperative decision support. *JAMA Surg*. 2019;154:1064–1065.