Computer Vision in the Surgical Operating Room

François Chadebecq[a]   Francisco Vasconcelos[a]   Evangelos Mazomenos[a]
Danail Stoyanov[a]

[a]Department of Computer Science, Wellcome/EPSRC Centre for Interventional and Surgical Sciences (WEISS),
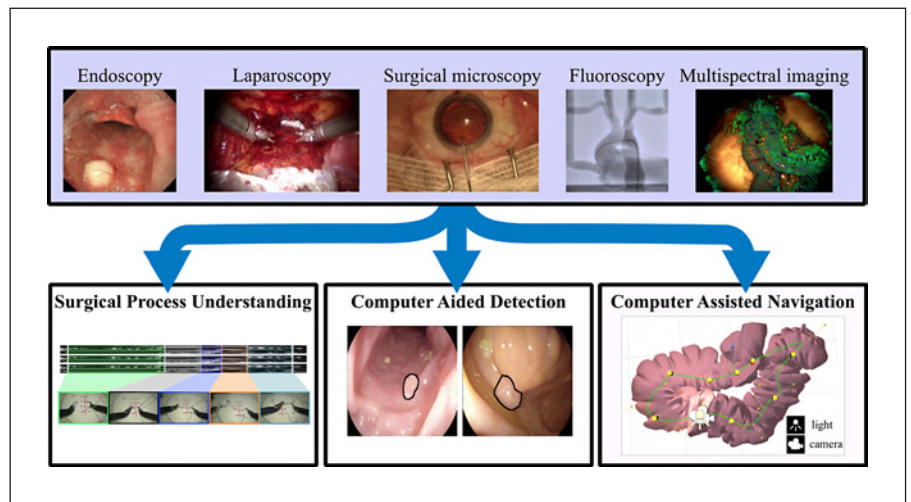University College London, London, UK

Abstract

Background: Multiple types of surgical cameras are used in modern surgical practice and provide a rich visual signal that is used by surgeons to visualize the clinical site and make clinical decisions. This signal can also be used by artificial intelligence (AI) methods to provide support in identifying instruments, structures, or activities both in real-time during procedures and postoperatively for analytics and understanding of surgical processes. *Summary:* In this paper, we provide a succinct perspective on the use of AI and especially computer vision to power solutions for the surgical operating room (OR). The synergy between data availability and technical advances in computational power and AI methodology has led to rapid developments in the field and promising advances. *Key Messages:* With the increasing availability of surgical video sources and the convergence of technologies around video storage, processing, and understanding, we believe clinical solutions and products leveraging vision are going to become an important component of modern surgical capabilities. However, both technical and clinical challenges remain to be overcome to efficiently make use of vision-based approaches into the clinic.

© 2020 S. Karger AG, Basel

Introduction

Surgery has progressively shifted towards the minimally invasive surgery (MIS) paradigm. This means that today most operating rooms (OR) are equipped with digital cameras that visualize the surgical site. The video generated by surgical cameras is a form of digital measurement and observation of the patient anatomy at the surgical site. It contains information about the appearance, shape, motion, and function of the anatomy and instrumentation within it. Once recorded over the duration of a procedure it also embeds information about the surgical process, actions performed, instruments used, possible hazards or complications, and even information about risk. While such information can be inferred by expert observers, this is not practical for providing assistance in routine clinical use and automated techniques are necessary to effectively utilize the data for driving improvements in practice [1, 2].

Currently, the majority of surgical video is either not recorded or it is stored for a limited period of time on the stack accompanying the surgical camera and then discarded at a later date. Perhaps some video is used in case presentations during clinical meeting discussions or society conferences or for educational purposes, and on an individual level surgeons may choose to record their case history. Storage has an associated cost and hence it is sensible to reduce data stores to only relevant and clinically useful information. This is largely due to the lack of tools that can synthesize the surgical video into meaningful information, either about

Danail Stoyanov
Wellcome/EPSRC Centre for Interventional and Surgical Sciences
University College London, 43–45 Foley St.
London, WC1 2BT (UK)
danail.stoyanov@ucl.ac.uk

Karger

Fig. 1. The use of computer vision to process data from a wide range of intraoperative imaging modalities or cameras can be grouped into the following 3 main applications: surgical process understanding, CAD, and computer-assisted navigation. We adopted this grouping for the purposes of this article, although additional applications are also discussed in the text.

the process or about physiological information contained in the video observations. For certain diagnostic procedures, e.g., endoscopic gastroenterology, storage of images from the procedure into the patient medical record to document observed lesions is becoming standard practice but this is largely not done for surgical video.

In addition to surgical cameras, it is also nowadays common for other OR cameras to be present. These can be used to monitor activity throughout the OR and not just at the surgical site [3]. As such, opportunities are present to capture this signal and provide an understanding of the entire room and activities or events that occur within it. This can potentially be used to optimize team performance or monitor room level events that can be used to improve the surgical process. To effectively make use of video data from the OR, it is necessary to build algorithms for video analysis and understanding. In this paper, we provide a short review of the state of the art in artificial intelligence (AI), and especially computer vision, for the analysis of surgical data and outline some of the concepts and directions for future development and practical translation into the clinic.

Computer Vision

The field of computer vision is a sub-branch of AI focused on building algorithms and methods for understanding information captured in images and video [4]. To make vision problems tractable, computational methods typically focus on sub-components of the human visual system, e.g., object detection or identification (classification), motion extraction, or spatial understanding (Fig. 1). Developing these building blocks in the context of surgery and surgeons' vision can lead to exciting possibilities for utilizing surgical video [4].

Computer vision has seen major improvements in the last 2 decades driven by breakthroughs in computing, digital cameras, mathematical modelling, and most recently deep learning techniques. While previous systems required human intervention in the design and modelling of image features that capture different objects in a certain domain, in deep learning the most discriminant features are learned autonomously from extensive amounts of annotated data. The increasing access to high volumes of digitally recorded image-guided surgeries is sparking a significant interest in translating deep learning to intraoperative imaging. Annotated surgical video datasets in a wide variety of domains are being made publicly available for training validating new algorithms in the form of competition challenges [5], resulting in a rapid progress towards reliable automatic interpretation of surgical data.

Surgical Process Understanding

Surgical procedures can be decomposed into a number of sequential tasks (e.g., dissection, suturing, and anastomosis) typically called procedural phases or steps [6–8]. Recognizing and temporally localizing these tasks allows for process surgical modelling and workflow analysis [9]. This further facilitates the current trend in MIS practice towards establishing standardized protocols for surgical workflow and guidelines for task execution, describing optimal tool positioning with respect to the anatomy, setting performance benchmarks, and ensuring operational safety and complication-free, cost-effective procedures [10, 11]. The ability to objectively quantify surgical performance could impact many aspects of the user and patient experience, like a reduced mental/physical load, increased safety, and more efficient training and planning [12, 13]. Intraoperative video is the main sensory cue for

surgical operators and provides a wealth of information about the workflow and quality of the procedure. Applying computer vision in the OR for workflow and skills analysis extends beyond the interventional video. Operational characteristics can be extracted from tracking and motion analysis of clinical staff using wall-mounted cameras and embedded sensors [14] and from tracking eye movements and estimating gaze patterns in MIS [15]. As in similar data science problems, learning-based AI has the potential to pioneer surgical workflow analysis and skills assessment and represents the focus of this section.

Surgical Phase Recognition

Surgical video has been used for segmenting surgical procedures into phases and the development of AI methods for workflow analysis (or phase recognition), facilitated by publicly annotated datasets [7, 8], has dramatically accelerated the stability and capability of recognizing and temporally localizing surgical tasks in different MIS procedures. The EndoNet architecture introduced a convolutional neural network for workflow analysis in laparoscopic MIS, specifically laparoscopic cholecystectomy, with the ability to recognize the 7 surgical phases of the procedure with over 80% accuracy [16]. More complex AI models (SV-RCNet, Endo3D) have increased the accuracy to almost 90% [17, 18]. One of the main requirements for learning techniques is data and annotations which are still limited in the surgical context. In robotic-assisted MIS procedures, instrument kinematics can be used in conjunction with the video to add explicit information on instrument motion. The JHU-ISI Gesture and Skill Assessment Working Set (JIGSAWS) is a dataset of synchronized video and robot kinematics data from benchtop simulations of 3 (suturing, knot tying, and needle passing) fundamental surgical tasks [6]. The JIGSAWS dataset has extended annotations at a sub-task level. AI techniques learn patterns and temporal interconnections of the sub-task sequences from combinations of robot kinematics and surgical video and detect and temporally localize each sub-task [19–24]. Recently, AI models for activity recognition have been developed and tested on annotated datasets from real cases of robotic-assisted radical prostatectomy and ocular microsurgery [18–20]. Future work in this area should focus on investigating the ability of AI methods for surgical workflow analysis to generalize with rigorous validation on multicenter annotated datasets of real procedures [20].

Surgical Technical Skill Assessment

Automated surgical skill assessment attempts to provide an objective estimation of the surgeons' performance and quality of execution [25]. AI models analyze the surgical video and learn high-level features to discriminate different performance and experience levels during the execution of surgical tasks. In studies on robotic surgical skills, using the JIGSAWS dataset, such systems can estimate manually assigned OSATS-based scores with more than 95% accuracy [26–28]. In interventional ultrasound (US) imaging, AI methods can automatically measure the operator's skills by evaluating the image quality of the captured US images with respect to their medical content [29, 30].

Computer-Aided Detection

Automatically detecting structures of interest in digital images is a well-established field in computer vision. Its real-time application to surgical video provides assistance in visualizing clinical targets and sensitive areas to optimize and increase the safety of a procedure.

Lesion Detection in Endoscopy

AI vision systems for computer-aided detection (CAD) can provide assistance during diagnostic interventions by automatically highlighting lesions and abnormalities that could otherwise be missed. CAD systems were firstly introduced in radiology, with existing US Food and Drug Administration (FDA)- and European Economic Area (EEA)-approved systems for mammography and chest computed tomography (CT) [31]. In the interventional context there has been particular interest in developing CADe systems for gastrointestinal endoscopy. Colonoscopy has received the most attention to date, and prototype CAD systems for polyp detection report accuracies as high as 97.8% using magnified narrow band imaging [32]. Similar systems have also been developed for endocytoscopy, capsule endoscopy, and conventional white light colonoscopes [33]. Research on CADe systems is also targeting esophageal cancer and early neoplasia in Barret esophagus [34].

Anatomy Detection

Detection and highlighting of anatomical regions during surgery may provide assisted guidance and avoid accidental damage to critical structures, such as vessels and nerves. While significant research on critical anatomy representation focuses on registration and display of data from preoperative scans (see Surgical Process Understanding), more recent approaches directly detect these structures from intraoperative images and video. In robotic prostatectomy, subtle pulsation of vessels can be detected and magnified to an extent that is perceivable by a surgeon [35]. In laparoscopic cholecystectomy, automated video retrieval can help in assessing whether a critical view of safety was achieved [36], with potential risk reduction and a safer removal of the gallbladder. Additionally, the detection and classification of anatomy enables

the automatic generation of standardized procedure reports for quality control assessment and clinical training [37].

Surgical Instrument Detection

Automatic detection and localization of surgical instruments, when integrated with anatomy detection, can also contribute to accurate positioning and ensure critical structures are not damaged. In robotic MIS (RMIS), this information has the added benefit of making progress toward active guidance, e.g., during needle suturing [38]. For this reason, research on surgical instrument detection has seen its largest share of research targeting the articulated tools in RMIS [39]. With RMIS there are interesting possibilities in using robotic systems to generate data for training AI models and bypassing the need for expensive manual labelling [40]. However, instrument detection has also received some attention in nonrobotic procedures including colorectal, pelvis, spine, and retinal surgery [41]. In such cases, vision for instrument analysis may assist in building systems that can report analytics about instrument usage for reporting, or instrument motion and activity for surgical technical skill analysis or verification.

Computer-Assisted Navigation

Vision-based methods for localization and mapping of the environment using the surgical camera have advanced rapidly in recent years. This is crucial in both diagnostic and surgical procedures because it may enable more complete diagnosis or fusion of pre- and intraoperative information to enhance clinical decision making.

Enhancing Navigation in Endoscopy

For providing physicians with navigation assistance in MIS, these systems must be able to locate the position of the endoscope within explored organs while simultaneously inferring their shape. Simultaneous localization and mapping in endoscopy is, however, a challenging problem [42]. The ability of deep learning approaches to learn characteristic data features has proven to outperform hand-crafted features detectors and descriptors in laparoscopy, colonoscopy, and sinus endoscopy [43]. These approaches have also demonstrated promising results for registration and mosaicking in fetoscopy with the aim of augmenting the fetoscope field of view [44]. Nonetheless, endoscopy image registration remains an open problem due to the complex topological and photometrical properties of organs producing significant appearance variations and complex specular reflections. Deep learning-based simultaneous localization and mapping approaches rely on the ability of neural networks to

learn a depth map from a single image, overcoming the need for image registration. It has recently been shown that these approaches are able to infer dense and detailed depth maps in colonoscopy [45]. By fusing consecutive depth maps and simultaneously estimating the endoscope motion using geometric constraints, it has been demonstrated that long range colon sections could be reconstructed [46]. A similar approach has also been successfully applied to 3-D reconstruction of the sinus anatomy from endoscopic video so as to propose an alternative to CT scans – expensive procedures using ionizing radiation – for longitudinal monitoring of patients after nasal obstruction surgery [47]. However, critical limitations, such as navigation within deformable environments, need to be overcome.

Navigation in Robotic Surgery

Surgical robots such as the da Vinci Surgical System generally use stereo endoscopes which have significant advantages over monocular endoscopes in their ability to capture 3-D measurements. Estimating a dense depth map from a pair of stereo images generally consists of estimating dense disparity maps defining the apparent pixel motion between 2 images. Most of the stereo registration approaches rely on geometric methods [48]. It has, however, been shown that DL-based approaches could be successfully applied to partial nephrectomy outperforming state of the art stereo reconstruction methods [49]. Surgical tool segmentation and localization contribute to safe tool-tissue interaction and are essential to visually guided manipulation tasks. Recent DL approaches demonstrate significant improvements over hand-crafted tool tracking methods offering a high degree of flexibility, accuracy, and reliability [50].

Image Fusion and Image-Guided Surgery

A key concept in enhancing surgical navigation has been the idea of fusing multiple preoperative and intraoperative imaging modalities in an augmented reality (AR) view of the surgical site [48]. Vision-based AR systems generally involve mapping and localization of the environment in addition to blocks that align any preoperative 3-D data models to that reconstruction and then display the fused information to the surgeon [51]. The majority of surgical AR systems have been founded on geometric vision algorithms but deep learning methods are emerging, e.g., for US to CT in spine surgery [52] or to design efficient deformable registration in laparoscopic liver surgery [53]. Despite methodological advances, significant open problems persist in surgical AR, such as adding contextual information to the visualization (e.g., identifying anatomical structures and critical surgical areas and detecting surgical phases and complications) [54], ensuring robust localization despite occlusions and

displaying relevant information to different stakeholders in the OR. Work is advancing to address these challenges and evaluation of the state-of-the-art learning-based method for visual human pose estimation in the OR has recently been reported [55] alongside a review dedicated to face detection into the OR [56] and methods to estimate both surgical phases and remaining surgery durations [57] which can be used to alter information displayed at different times.

Discussion

In this paper, we have provided a succinct review of the broad possibilities for using computer vision in the surgical OR. With the increasing availability of surgical video sources and the convergence of technologies around video storage, processing, and understanding, we believe clinical solutions and products leveraging vision are going to become an important component of modern surgical capabilities. However, both technical and clinical challenges remain and we try to outline them below.

Priorities for technical research and development are:

- Availability of datasets with labels and ground truth: despite efforts from challenges, the quality and availability of large scale surgical datasets remains a bottleneck. Efforts are needed to address this and cause a similar catalyst effect as was observed in wider vision and AI communities.
- Technical development in unsupervised methods: developing approaches that do not require any labelled sensor data (ground truth) is needed to bypass the need for large scale dataset or adapt to new domains (i.e., adapt method dedicated to nonmedical data to medical imaging). Furthermore, even if the data gap is bridged, the domain of surgical problems and axes of variation (patient, disease, etc.) is huge and solutions need to be adaptive to be able to scale.

Challenges for clinical deployment are:

- Technical challenges in infrastructure: computing facilities in the OR, access to cloud computing using limited bandwidth, and latency of delivering solutions are all practical problems that require engineering resources beyond the core AI development.
- Regulatory requirements around solutions: various levels of regulation are needed for integrating medical devices and software within the OR. Because of their complexity, assessing the limitation and capabilities of AI-based solutions is difficult, particularly for problems in which human supervision cannot be used to validate their precision (e.g., simultaneous localization and mapping).
- User interfaces design: it is critical to ensure that only relevant information is provided to the surgical teams

and, for advanced AI-based solutions, a direct practitioner-surgical platform communication can be established. Integrating contextual information (e.g., surgical phase recognition and practitioner identification) is a major challenge for developing efficient user interfaces.

Finally, this short and succinct review has focused on research directions that are in active development. Due to limitations of space, we have not discussed opportunities around using computer vision with different imaging systems or spectral imaging despite the opportunities in AI systems to resolve ill posed inverse problems in that domain [58]. Additionally, we have not covered in detail work in vision for the entire OR but this is a very active area of development with exciting potential for a wider team workflow understanding [3].

Conflict of Interest Statement

D. Stoyanov is part of Digital Surgery Ltd. and a shareholder in Odin Vision Ltd.

The authors have no conflict of interests to declare.

Funding Sources

Author Contributions

E. Mazomenos wrote the section dedicated to surgical video analysis. F. Vasconcelos wrote the section dedicated to CAD and F. Chadebecq the section dedicated to computer-assisted interventions. D. Stoyanov wrote the Introduction and Discussion and was responsible for the organization of this paper.

References

1 Maier-Hein L, Vedula SS, Speidel S, Navab N, Kikinis R, Park A, et al. Surgical data science for next-generation interventions. Nat Biomed Eng. 2017 Sep;1(9):691–6.

2 Stoyanov D. Surgical vision. Ann Biomed Eng. 2012 Feb;40(2):332–45.

3 Jung JJ, Jüni P, Lebovic G, Grantcharov T. First-year Analysis of the Operating Room Black Box Study. Ann Surg. 2020 Jan;271(1):122–7.

4 Prince SJ. Computer vision: Models learning and inference. Cambridge: Cambridge University Press; 2012.

5 Bernal J, Tajkbaksh N, Sánchez FJ, Matuszewski BJ, Hao Chen, Lequan Yu, et al. Comparative validation of polyp detection methods in video colonoscopy: results from the MICCAI 2015 endoscopic vision challenge. IEEE Trans Med Imaging. 2017 Jun;36(6): 1231–49.

6 Ahmidi N, Tao L, Sefati S, Gao Y, Lea C, Haro BB, et al. A Dataset and Benchmarks for Segmentation and Recognition of Gestures in Robotic Surgery. IEEE Trans Biomed Eng. 2017 Sep;64(9):2025–41.

7 Stauder R, Ostler D, Kranzfelder M, Koller S, Feußner H, Navab N. The TUM LapChole dataset for the M2CAI 2016 workflow challenge [Internet]. 2016. Available from: https://arxiv.org/abs/1610.09278.

8 Flouty E, Kadkhodamohammadi A, Luengo I, Fuentes-Hurtado F, Taleb H, Barbarisi S, et al. CaDIS: Cataract dataset for image segmentation [Internet]. 2019. Available from: https://arxiv.org/abs/1906.11586.

9 Gholinejad M, J Loeve A, Dankelman J. Surgical process modelling strategies: which method to choose for determining workflow? Minim Invasive Ther Allied Technol. 2019 Apr; 28(2):91–104.

10 Gill S, Stetler JL, Patel A, Shaffer VO, Srinivasan J, Staley C, et al. Transanal Minimally Invasive Surgery (TAMIS): Standardizing a Reproducible Procedure. J Gastrointest Surg. 2015 Aug;19(8):1528–36.

11 Jajja MR, Maxwell D, Hashmi SS, Meltzer RS, Lin E, Sweeney JF, et al. Standardization of operative technique in minimally invasive right hepatectomy: improving cost-value relationship through value stream mapping in hepatobiliary surgery. HPB (Oxford). 2019 May;21(5):566–73.

12 Reiley CE, Lin HC, Yuh DD, Hager GD. Review of methods for objective surgical skill evaluation. Surg Endosc. 2011 Feb;25(2):356–66.

13 Mazomenos EB, Chang PL, Rolls A, Hawkes DJ, Bicknell CD, Vander Poorten E, et al. A survey on the current status and future challenges towards objective skills assessment in endovascular surgery. J Med Robot Res. 2016; 1(3):1640010.

14 Azevedo-Coste C, Pissard-Gibollet R, Toupet G, Fleury É, Lucet JC, Birgand G. Tracking Clinical Staff Behaviors in an Operating Room. Sensors (Basel). 2019 May;19(10): 2287.

15 Khan RS, Tien G, Atkins MS, Zheng B, Panton ON, Meneghetti AT. Analysis of eye gaze: do novice surgeons look at the same location as expert surgeons during a laparoscopic operation? Surg Endosc. 2012 Dec;26(12):3536–40.

16 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 Jan;36(1):86–97.

17 Chen W, Feng J, Lu J, Zhou J. Endo3D: Online workflow analysis for endoscopic surgeries based on 3D CNN and LSTM. In: Stoyanov D, Taylor Z, Sarikaya D, et al., editors. OR 2.0 context-aware operating theaters, computer assisted robotic endoscopy, clinical image-based procedures, and skin image analysis. Cham: Springer International; 2018.

18 Jin Y, Dou Q, Chen H, Yu L, Qin J, Fu CW, et al. SV-RCNet: Workflow Recognition From Surgical Videos Using Recurrent Convolutional Network. IEEE Trans Med Imaging. 2018 May;37(5):1114–26.

19 Funke I, Bodenstedt S, Oehme F, von Bechtolsheim F, Weitz J, Speidel S. Using 3D Convolutional Neural Networks to Learn Spatiotemporal Features for Automatic Surgical Gesture Recognition in Video. Conference on Medical Image Computing and Computer-Assisted Intervention. 2019. https://doi.org/10.1007/978-3-030-32254-0_52.

20 Bodenstedt S, Rivoir D, Jenke A, Wagner M, Breucha M, Müller-Stich B, et al. Active learning using deep Bayesian networks for surgical workflow analysis. Int J CARS. 2019 Jun; 14(6):1079–87.

21 Liu D, Jiang T. Deep Reinforcement Learning for Surgical Gesture Segmentation and Classification. Conference on Medical Image Computing and Computer-Assisted Intervention. 2018; vol 11073, pp 247–55.

22 Despinoy F, Bouget D, Forestier G, Penet C, Zemiti N, Poignet P, et al. Unsupervised Trajectory Segmentation for Surgical Gesture Recognition in Robotic Training. IEEE Trans Biomed Eng. 2016 Jun;63(6):1280–91.

23 Van Amsterdam B, Nakawala H, De Momi E, Stoyanov D. Weakly Supervised Recognition of Surgical Gestures. IEEE International Conference on Robotics and Automation. 2019; pp 9565–71.

24 Van Amsterdam B, Clarkson MJ, Stoyanov D. Multi-Task Recurrent Neural Network for Surgical Gesture Recognition and Progress Prediction. IEEE International Conference on Robotics and Automation. 2020.

25 Vedual SS, Ishii M, Hager GD. Objective assessment of surgical technical skill and competency in the operating room. Annu Rev Biomed Eng. 2017 Jun 21;19:301–25.

26 Nguyen XA, Ljuhar D, Pacilli M, Nataraja RM, Chauhan S. Surgical skill levels: classification and analysis using deep neural network model and motion signals. Comput Methods Programs Biomed. 2019 Aug;177:1–8.

27 Ismail Fawaz H, Forestier G, Weber J, Idoumghar L, Muller PA. Accurate and interpretable evaluation of surgical skills from kinematic data using fully convolutional neural networks. Int J CARS. 2019 Sep;14(9):1611–7.

28 Wang Z, Majewicz Fey A. Deep learning with convolutional neural network for objective skill evaluation in robot-assisted surgery. Int J CARS. 2018 Dec;13(12):1959–70.

29 Fawaz HI, Forestier G, Weber J, Idoumghar L, Muller PA. Automated Performance Assessment in Transoesophageal Echocardiography with Convolutional Neural Networks. Conference on Medical Image Computing and Computer Assisted Intervention. 2018.

30 Abdi AH, Luong C, Tsang T, Allan G, Nouranian S, Jue J et al. Automatic quality assessment of echocardiograms using convolutional neural networks: Feasibility on the apical four-chamber view. IEEE Trans Med Imaging. 2017 Jun;36(6):1221–30.

31 Castellino RA. Computer aided detection (CAD): an overview. Cancer Imaging. 2005 Aug;5(1):17–9.

32 Ahmad OF, Soares AS, Mazomenos E, Brandao P, Vega R, Seward E, et al. Artificial intelligence and computer-aided diagnosis in colonoscopy: current evidence and future directions. Lancet Gastroenterol Hepatol. 2019 Jan;4(1):71–80.

33 Brandao P, Zisimopoulos O, Mazomenos E, Ciuti G, Bernal J, Visentini-Scarzanella M, et al. Towards a computed-aided diagnosis system in colonoscopy: automatic polyp segmentation using convolution neural networks. J Med Robot Res. 2018;3(2):1840002.

34 Hussein M, Puyal J, Brandao P, Toth D, Sehgal V, Everson M, et al. Deep neural network for the detection of early neoplasia in Barret's oesophagus. Gastrointest Endosc. 2020; 91(6):AB250.

35 Janatka M, Sridhar A, Kelly J, Stoyanov D. Higher order of motion magnification for vessel localisation in surgical video. Conference on Medical Image Computing and Computer-Assisted Intervention. 2018. https://doi.org/10.1007/978-3-030-00937-3_36.

36 Mascagni P, Fiorillo C, Urade T, Emre T, Yu T, Wakabayashi T, et al. Formalizing video documentation of the critical view of safety in laparoscopic cholecystectomy: a step towards artificial intelligence assistance to improve surgical safety. Surg Endosc. 2020 Jun;34(6): 2709–14.

37 He Q, Bano S, Ahmad OF, Yang B, Chen X, Valdastri P, et al. Deep learning-based anatomical site classification for upper gastrointestinal endoscopy. Int J CARS. 2020 Jul; 15(7):1085–94.

38 D'Ettorre C, Dwyer G, Du X, Chadebecq F, Vasconcelos F, De Momi E, et al. Automated pick-up of suturing needles for robotic surgical assistance. IEEE International Conference on Robotics and Automation. 2018; pp 1370-1377.

39 Du X, Kurmann T, Chang PL, Allan M, Ourselin S, Sznitman R, et al. Articulated multi-instrument 2-D pose estimation using fully convolutional networks. IEEE Trans Med Imaging. 2018 May;37(5):1276–87.

40 Colleoni E, Edwards P, Stoyanov D. Synthetic and Real Inputs for Tool Segmentation in Robotic Surgery. Conference on Medical Image Computing and Computer Assisted Intervention. 2020. https://doi.org/10.1007/978-3-030-59716-0_67.

41 Bouget D, Allan M, Stoyanov D, Jannin P. Vision-based and marker-less surgical tool detection and tracking: a review of the literature. Med Image Anal. 2017 Jan;35:633–54.

42 Maier-Hein L, Mountney P, Bartoli A, Elhawary H, Elson D, Groch A, et al. Optical techniques for 3D surface reconstruction in computer-assisted laparoscopic surgery. Med Image Anal. 2013 Dec;17(8):974–96.

43 Liu X, Zheng Y, Killeen B, Ishii M, Hager GD, Taylor RH, et al. Extremely dense point correspondences using a learned feature descriptor [Internet]. 2020. Available from: https://arxiv.org/abs/2003.00619.

44 Bano S, Vasconcelos F, Tella Amo M, Dwyer G, Gruijthuijsen C, Deprest J, et al. Deep Sequential Mosaicking of Fetoscopic Videos. Conference on Medical Image Computing and Computer Assisted Intervention. 2019.

45 Rau A, Edwards PJ, Ahmad OF, Riordan P, Janatka M, Lovat LB, et al. Implicit domain adaptation with conditional generative adversarial networks for depth prediction in endoscopy. Int J CARS. 2019 Jul;14(7):1167–76.

46 Ma R, Wang R, Pizer S, Rosenman J, McGill SK, Frahm JM. Real-time 3d reconstruction of colonoscopic surfaces for determining missing regions. Conference on Medical Image Computing and Computer-Assisted Intervention. 2019. https://doi.org/10.1007/978-3-030-32254-0_64.

47 Liu X, Stiber M, Huang J, Ishii M, Hager GD, Taylor RH, et al. Reconstructing sinus anatomy from endoscopic video: towards a radiation-free approach for quantitative longitudinal assessment [Internet]. 2020. Available from: https://arxiv.org/abs/2003.08502.

48 Peters TM, Linte CA, Yaniv Z, Williams J. Mixed and augmented reality in medicine. Boca Raton: CRC Press; 2018.

49 Luo H, Hu Q, Jia F. Details preserved unsupervised depth estimation by fusing traditional stereo knowledge from laparoscopic images. Healthc Technol Lett. 2019 Nov;6(6):154–8.

50 Colleoni E, Moccia S, Du X, De Momi E, Stoyanov D. Deep learning based robotic tool detection and articulation estimation with spatio-temporal layers. IEEE Robot Autom Lett. 2019;4(3):2714–21.

51 Park BJ, Hunt SJ, Martin C 3rd, Nadolski GJ, Wood BJ, Gade TP. Augmented and Mixed Reality: Technologies for Enhancing the Future of IR. J Vasc Interv Radiol. 2020 Jul;31(7):1074–82.

52 Chen F, Wu D, Liao H. Registration of CT and ultrasound images of the spine with neural network and orientation code mutual information. Med Imaging Augment Real. 2016;9805:292–301.

53 Brunet JN, Mendizabal A, Petit A, Golse N, Vibert E, Cotin S. Physics-based deep neural network for augmented reality during liver surgery. Conference on Medical Image Computing and Computer Assisted Intervention. 2019. https://doi.org/10.1007/978-3-030-32254-0_16.

54 Vercauteren T, Unberath M, Padoy N, Navab N. CAI4CAI: The Rise of Contextual Artificial Intelligence in Computer Assisted Interventions. Proc IEEE Inst Electr Electron Eng. 2020 Jan;108(1):198–214.

55 Srivastav V, Issenhuth T, Abdolrahim K, de Mathelin M, Gangi A, Padoy N. MVOR: A multi-view RGB-D operating room dataset for 2D and 3D human pose estimation. Conference on Medical Image Computing and Computer Assisted Intervention. 2018.

56 Issenhuth T, Srivastav V, Gangi A, Padoy N. Face detection in the operating room: comparison of state-of-the-art methods and a self-supervised approach. Int J CARS. 2019 Jun;14(6):1049–58.

57 Yengera G, Mutter D, Marescaux J, Padoy N. Less is more: surgical phase recognition with less annotations through self-supervised pretraining of CNN–LSTM networks [Internet]. 2018. Available from: https://arxiv.org/abs/1805.08569.

58 Clancy NT, Jones G, Maier-Hein L, Elson DS, Stoyanov D. Surgical spectral imaging. Med Image Anal. 2020 Jul;63:101699.