Video labelling robot-assisted radical prostatectomy and the role of artificial intelligence (AI): training a novice

Samy Cheikh Youssef1 · Nadine Hachach-Haram2 · Abdullatif Aydin1 · Taimur T. Shah3 · Nikhil Sapre3 · Rajesh Nair3 · Sonpreet Rai3 · Prokar Dasgupta1,3

Received: 29 July 2022 / Accepted: 10 October 2022 / Published online: 30 October 2022 © The Author(s) 2022

Abstract
Video labelling is the assigning of meaningful information to raw videos. With the evolution of artificial intelligence and its intended incorporation into the operating room, video datasets can be invaluable tools for education and the training of intelligent surgical workflow systems through computer vision. However, the process of manual labelling of video datasets can prove costly and time-consuming for already busy practising surgeons. Twenty-five robot-assisted radical prostatectomy (RARP) procedures were recorded on Proximie, an augmented reality platform, anonymised and access given to a novice, who was trained to develop the knowledge and skills needed to accurately segment a full-length RARP procedure on a video labelling platform. A labelled video was subsequently randomly selected for assessment of accuracy by four practising urologists. Of the 25 videos allocated, 17 were deemed suitable for labelling, and 8 were excluded on the basis of procedure length and video quality. The labelled video selected for assessment was graded for accuracy of temporal labelling, with an average score of 93.1%, and a range of 85.6–100%. The self-training of a novice in the accurate segmentation of a surgical video to the standard of a practising urologist is feasible and practical for the RARP procedure. The assigning of temporal labels on a video labelling platform was also studied and proved feasible throughout the study period.

Keywords Computer vision · Robot-assisted radical prostatectomy · Artificial intelligence · Video labelling · Surgical video

Introduction

Video labelling is the process through which different aspects of a video are assigned specific informative labels which a machine can utilise in machine learning and in the development of computer vision. Computer vision is the field of artificial intelligence (AI) which enables machines to develop a human-like understanding of different aspects of images and videos and enables the human-like learning of information from visual data. Machines to date can be taught to infer, analyse, and detect subtle data patterns and compute on their own through unsupervised machine learning, necessitating no explicit instruction [1].

One application of video labelling in surgery is the automatic segmentation of surgical videos. This application could be employed in the education of surgical trainees with research supporting the utilisation of video-based educational interventions [2, 3]. The automatic segmentation of surgical videos may also enable easy navigation through video indexing, improve interpretability of surgical video, and aid assessment of surgical skill [4, 5].

Although automation in surgical procedures is likely to remain out of reach in the near future, there is potential in the synergy between computer vision in AI and automation of surgical video labelling. A difficulty in the training of machine learning algorithms for classifying features of visual data is the prerequisite for large quantities of pre-labelled data [6]. Due to the nature of surgical videos and the heterogeneity in surgeon approaches, patient anatomy, and intraoperative events—to achieve accurate annotation, surgical expertise is necessary. However, the feasibility of incorporating experienced surgeons in such studies may not
be feasible at every institution and is costly, both in their employment to perform such research and in time which could be spent treating patients [7].

Research for crowd annotation with surgical video is scarce, and current evidence suggesting a strong correlation between crowd annotators and clinical experts remains inconclusive for tasks of higher complexity or the labelling of full-length surgical procedures [7]. The established difficulties with labelling libraries of surgical video for computer vision research warrant further research into alternative methods in which accurately labelled video datasets can be produced. This study investigates the feasibility of training a novice student, whereby a foundational knowledge of anatomy and research is existent, in labelling a full-length robot-assisted radical prostatectomy (RARP) video.

Methods

Over a period of 2 months, 25 RARP procedures performed on the Da Vinic Si HD dual console system were recorded on Proximie, a novel commercially available GDPR and HIPAA compliant augmented reality platform, approved by Guys and St. Thomas’ NHS trust (GSTT) IT department.

The footage was captured in the endoscopic view, producing a library of two-dimensional (2D) full-length RARP videos which were deidentified and anonymised. This video set was accessible via a secure, online, cloud-based storage which only the study participants had access to.

The student was trained on video labelling through self-directed learning, a review of the literature and reference to online video materials. Subsequently, a random video was selected from the dataset to be labelled on an online video labelling platform. This same video was then assessed according to an agreed pre-set checklist, on a 5-point Likert scale by four practising urologists experienced in performing RARP. The accuracy of video labelling was then calculated and documented.

Patient and public involvement

Patients were not involved in the design and conduction of this study. Prostatectomy patients at GSTT undergoing RARP completed a standardised consent form for the storage and usage of surgical video. Surgical videos were then obtained and accessed via Proximie, a GDPR and HIPAA compliant platform.

Data collection

The student was able to access the video through a secure online cloud-based storage with files protected by a 256-bit advanced encryption standard (AES). Patients had full knowledge of recording prior to the procedure and completed a consent form enabling the subsequent usage of the operative video. Videos were anonymised using an arbitrary numerical system deidentifying patients. A random number generator was utilised to select a video for assessment of video labelling accuracy [8].

Video labelling

A comprehensive library of operative steps was outlined according to those previously defined in the literature, considering the steps which would be visible throughout operative video (robot setup and positioning, pneumoperitoneum, and port placement omitted as these steps are not visible in surgical video). The nine final steps assessed are outlined in Table 1. These steps were defined based on the review of relevant literature [9, 10], and were then validated for video labelling by a robotics clinical fellow and expert urologist.

Video labelling was performed on the VGG image annotator (VIA) platform [11] by the student (Fig. 1). For the purposes of review, operative video was disseminated via a cloud-based storage network using a private link, and time stamps were listed and sent separately as a word document.

Assessment of accuracy

Following the learning period, a video amongst the data set was randomly selected [8] to test the video labelling accuracy of the student. The accuracy was graded on a five-point Likert scale [12] and completed forms were collected from the participants, the mean scores were calculated for the 9 procedural steps (Tables 1 and 2).

Procedural steps for video labelling

See Table 1.

Results

Quality of surgical video

Of the 25 analysed videos, 8/25 (32%) videos were deemed incomplete/low quality, and these videos were not used for review or labelling by the student due to the operative video being less than 1 h in length (incomplete), missing significant steps or the video being pixelated and unclear to the viewer. Of the 17 videos which were deemed to be of sufficient quality for analysis, videos were time stamped by the student as part of the learning process (Fig. 2).
The annotated video then underwent review by the four participating urologists, who then quantified the accuracy of labelling of each procedural step, over a 5-point Likert scale. The mean accuracy of the labelled video across all steps was 93.06%, with a range from 85.6 to 100%. Refer to Table 2 for complete data.
Accuracy scores

See Table 2.

Discussion

This purpose of this pilot study is to investigate the feasibility of video labelling by a novice student, as an initial step towards the development of an AI algorithm, automating the process of video labelling in RARP as previously done for laparoscopic sleeve gastrectomy [13], laparoscopic sigmoidectomy [4], and laparoscopic cholecystectomy [5].

The results of this study demonstrated the feasibility of a novice student to train in the labelling and accurate segmentation of operative video over a short-term period on a part-time basis; in this case, the robotic prostatectomy procedure performed primarily via a transperitoneal approach. A procedural video amongst the data set was then selected for review by four practising urologists. The results demonstrated an average of over 90% accuracy in the time stamping and video labelling of the procedural steps. The step which had the lowest accuracy score was step 5 (Bladder neck transection, see Table 2) at 4.13/5.00, which post-assessment feedback suggested was resultant to a misunderstanding of the relevant surgical anatomy. This finding suggests the potential utility in a standardised pre-labelling training programme, delivered by the practising surgeons, to the novices intending to perform video labelling. During such training, complex anatomy, potential anatomical variants, and operative steps which may incite confusion can be identified and novices can develop a more comprehensive understanding prior to the commencement of video labelling.

For the purposes of this study, the VIA annotation software [14] was not used by the assessors in grading the accuracy of the labelling; rather, the assessors were sent the surgical video via a collaborative cloud link which allowed access to the full-length case, accompanied

---

**Table 2.** Accuracy scores—accuracy scores less than 60% (on average less than neutral overall) are deemed inaccurate on the merit of the proposed scoring system

| Reviewer + stage of training | Score of operative steps (out of 5) | Overall accuracy |
|-----------------------------|-------------------------------------|-----------------|
| NS (post-CCT clinical fellow) | 5.00 5.00 5.00 5.00 5.00 5.00 4.00 5.00 4.00 43.00/45.00 | 95.60% |
| SR (post-CCT clinical fellow) | 5.00 5.00 5.00 5.00 5.00 5.00 5.00 5.00 5.00 45.00/45.00 | 100.00% |
| TS (post-CCT clinical fellow) | 5.00 3.00 2.50 5.00 4.00 4.00 5.00 5.00 3.00 38.50/45.00 | 85.60% |
| RN (consultant urological surgeon) | 5.00 5.00 4.00 4.00 4.00 4.00 5.00 5.00 4.00 41.00/45.00 | 91.11% |

Average scores 5.00 4.50 4.75 4.75 4.13 4.75 4.25 4.75 4.75 4.75 41.88/45.00 Mean = 93.06%
by typed time stamps. The justification for this was the inconvenience which would be incurred if VIA were to be accessed to review a 3-h-long procedure. All the participants intending to review the surgical video with the temporal labels would be required to download the video which occupied 4.32 gigabytes (GB) and the alternative of a shared link was deemed sufficient in the context of this study and for review of the accuracy of time stamping.

The fields of surgical education, robotic surgery, and innovation in surgical technology with artificial intelligence are constantly evolving in the light of new technological developments. The video labelling performed in this project has several applications. In the immediate context, the organisation of a video dataset into a cloud storage network and its segmentation into the constituting steps of the procedure can be used for: research purposes, increasing efficiency of post-operative review of surgeon performance, and educational purposes whereby clarity of video materials is possibly improved through the addition of time stamps [15]. The educational value of video-based education in surgery has been established in recent years [3, 16]. As for the educational impact of adding labels to surgical video, this will require the conduction of future randomised-controlled trials. With the reduced working hours, the increasing financial strains faced by the NHS and reduced exposure to core surgical procedures by core surgical trainees—a high-quality surgical video library across specialties may augment current surgical training and practise [17, 18]. The COVID-19 pandemic has also caused significant disruption to surgical training, with reports of substantially reduced operative experience [19, 20], and the employment of alternative teaching methods could prove advantageous given the ability to remotely learn from video-based resources.

Labelled video has multiple further applications, for example, the development of context aware operating rooms with surgical workflow analysis [4, 21–23], labelled video is required as an initial reference point for machine learning and the establishment of algorithms which automate labelling and segmentation of surgical video. The future prospect of video labelling algorithms lies beyond the simple segmentation of surgical video, and is directed towards higher reasoning functions, such as surgical skill feedback, analysis of operative skill metrics, and intraoperative clinical decision support [13, 15].

However, an inherent difficulty encountered with this process is the time and expertise investment due as part of the manual labelling process, often demanding the participation of expert surgeons or motivated trainees [13, 15, 24]. This is where potential lies for student participation, and the current study has proven the possibility of student participation in a video labelling project. With numerous studies suggesting high interest to participate in research projects amongst students [25, 26], it may be an affordable, feasible alternative of benefit to both students and researchers.

Though this study has limitations. The results of this project may not be generalisable to all other endoscopic procedures. The results suggest promise in the ability of a student to learn to accurately identify the procedural steps within a surgical video, and subsequently label these. However, replicating this study for other procedures may demonstrate a difference in learning curves. Another significant feature of this study is that the data set was obtained from a single hospital, and the approach taken in performing RARP procedures in other hospitals may not be representative. It is known that there are several dissection techniques which may be done for RARP, with each differing in the anatomical structures encountered, the ability for labelling through different anatomical approaches was not tested in this study [27].

The context in which this study was conducted was during the peak of the COVID-19 pandemic, where clinical responsibilities of the participant surgeons and contact restrictions limited the recruitment of a student cohort. Despite the limitations, this study serves as a pilot study for more comprehensive future video labelling research. A formalised educational process can be applied amongst a cohort of students, to determine interrater reliability for students who have undergone the same educational process.

**Implications for future research**

Advancements in AI and surgical technology have prompted further research studies. A barrier to the conduction of such studies can often be due to the requirement of high expertise and associated funding. This study has shown that a B.Sc. student, over the course of 2 months, was able to self-train in the understanding and accurate segmentation of the RARP procedure with no prior exposure to robotic surgery, having analysed and manually labelled an entire dataset. This could form the basis for an educational surgical video library, a reference point for assessors, however most applicable to this project, the use of video datasets in computer vision research and AI applications. Studies assessing the ability of students to perform such roles are scarce in the literature and may be of benefit to the scientific community due to the existent interest for research participation in students.

**Author contributions** SC, AA, PD, and NHH contributed to the conception and design of the study. NHH provided the data set, which SC then analysed as per the methods and then wrote the paper with input from all the other authors. NS, TS, RN, and SR assessed the labelling performance of SC and provided performance data for analysis. All authors read and approved the final manuscript.

**Funding** This research did not receive any specific grant on behalf of a public/commercial funding agency and was conducted as part of a B.Sc.
Data availability  No additional data available.

Declarations

Conflict of interest  NHH is the CEO of Proximie, and SC, AA, PD, TS, RN, NS, and SR declare no conflict of interest.

Ethical approval  Proximie was approved as a quality improvement project by the information technology (IT) team and Caldicott Guardian at Guy's and St. Thomas' NHS Foundation Trust. All patients provided informed consent and there were no patient identifiable data.

Consent to participate  Patients completed written consent forms enabling the recording and subsequent usage of endoscopic video for research purposes, composing the video dataset used in this study.

Consent to publish  Article does not contain any personal medical information.

Open Access  This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Chen J, Remulla D, Nguyen JH et al (2019) Current status of artificial intelligence applications in urology and their potential to influence clinical practice. BJU Int 124:567–577. https://doi.org/10.1111/bju.14852
2. Akl MN, Giles DL, Long JB et al (2008) The efficacy of viewing an educational video as a method for the acquisition of basic laparoscopic suturing skills. J Minim Invasive Gynecol 15:410–413. https://doi.org/10.1016/j.mijgyn.2008.03.012
3. Ahmet A, Gamze K, Rustem M, Karaborklu Argut S (2018) Is video-based education an effective method in surgical education? A systematic review. J Surg Educ 75:1150–1158. https://doi.org/10.1016/j.jsurg.2018.01.014
4. Kitaguchi D, Takeshita N, Matsuzaki H et al (2020) Real-time automatic surgical phase recognition in laparoscopic sigmoidectomy using the convolutional neural network-based deep learning approach. Surg Endosc 34:4924–4931. https://doi.org/10.1007/s00464-019-07281-0
5. Twinanda AP, Shehata S, Mutter D et al (2017) EndoNet: a deep architecture for recognition tasks on laparoscopic videos. IEEE Trans Med Imaging 36:86–97. https://doi.org/10.1109/TMI.2016.2593957
6. Deo RC (2015) Machine learning in medicine. Circulation. https://doi.org/10.1161/CIRCULATIONAHA.115.01593
7. Ward TM, Fer DM, Ban Y et al (2021) Challenges in surgical video annotation. Comput Assist Surg. https://doi.org/10.1080/24099322.2021.1937320
8. Urbaniaik GC, Scott Plous (1997) https://www.randomizer.org/. In: Res. Randomizer
9. Huynh LM, Ahlering TE (2018) Robot-assisted radical prostatectomy: a step-by-step guide. J Endourol 32:S28–S32. https://doi.org/10.1089/end.2017.0723
10. Lovegrove C, Novara G, Mottrie A et al (2016) Structured and modular training pathway for robot-assisted radical prostatectomy (RARP): validation of the RARP assessment score and learning curve assessment. Eur Urol 69:526–535. https://doi.org/10.1016/j.eururo.2015.10.048
11. Andrew Zisserman, Abhishek Dutta, Ankush Gupta VGG Image Annotator (VIA). https://www.robots.ox.ac.uk/~vgg/software/via/. Accessed 24 Aug 2021
12. Sullivan GM, Artino AR (2013) Analyzing and interpreting data from likert-type scales. J Grad Med Educ 5:541–542. https://doi.org/10.4300/jgme-5-4-18
13. Hashimoto DA, Rosman G, Witkowski ER et al (2019) Computer vision analysis of intraoperative video: automated recognition of operative steps in laparoscopic sleeve gastrectomy. Ann Surg 270:414–421. https://doi.org/10.1097/SLA.0000000000003460
14. Dutta A, Zisserman A (2019) The VIA annotation software for images, audio and video. In: MM 2019 - Proceedings of the 27th ACM International Conference on Multimedia. https://doi.org/10.1145/3343031.3350535. Accessed 24 Aug 2021
15. Al Abbas AI, Jung JP, Rice MJK et al (2019) Methodology for developing an educational and research video library in minimally invasive surgery. J Surg Educ 76:745–755. https://doi.org/10.1016/j.jsurg.2018.10.011
16. Green JL, Suresh V, Bittar P et al (2019) The utilization of video technology in surgical education: a systematic review. J Surg Res 235:171–180. https://doi.org/10.1016/j.jsrs.2018.09.015
17. Toll E, Davis C (2010) More trainees and less operative exposure: a quantitative analysis of training opportunities for junior surgical trainees. Bull R Coll Surg Engl. https://doi.org/10.1308/147363510x12689756996360
18. Nagendran M, Kiew G, Raine R et al (2019) Financial performance of English NHS trusts and variation in clinical outcomes: a longitudinal observational study. BMI Open. https://doi.org/10.1136/bmjopen-2018-021854
19. Khan KS, Keay R, McLellan M, Mahmud S (2020) Impact of the COVID-19 pandemic on core surgical training. Scott Med J. https://doi.org/10.1177/0036933020949217
20. Hope C, Reilly JJ, Griffiths G et al (2021) The impact of COVID-19 on surgical training: a systematic review. Tech Coloproctol. https://doi.org/10.1007/s10151-020-02404-5
21. Stauder R, Okur A, Peter L, et al (2014) Random forests for phase identification in peroral endoscopic myotomy. Surg Endosc. https://doi.org/10.1007/s00464-019-07833-9
22. Ward TM, Hashimoto DA, Ban Y et al (2020) Automated operative phase identification in peroral endoscopic myotomy. Surg Endosc. https://doi.org/10.1007/s00464-020-07833-9
23. Ward TM, Mascagni P, Ban Y et al (2020) Computer vision in surgery. Surg (United States). https://doi.org/10.1016/j.surg.2020.10.039
24. Padov N (2019) Machine and deep learning for workflow recognition during surgery. Minim Invasive Ther Allied Technol 28:82–90. https://doi.org/10.1080/13645706.2019.1584116
25. Moraes DW, Jotz M, Menegazzo WR et al (2016) Interest In Artificial Intelligence Applications in Urology and their Potential to Influence Clinical Practice. BJU Int 124:567–577. https://doi.org/10.1111/bju.14852
26. Sayedalamin Z, Halawa TF, Baig M, et al (2018) Undergraduate medical research in the Gulf Cooperation Council (GCC) countries: a descriptive study of the students’ perspective. In: BMC Research Notes. https://doi.org/10.1186/s13104-018-3381-y
27. Martini A, Falagario UG, Villers A et al (2020) Contemporary techniques of prostate dissection for robot-assisted prostatectomy. Eur Urol. https://doi.org/10.1016/j.eururo.2020.07.017

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.