Deep learning for surgical phase recognition using endoscopic videos

Guédon, Annetje C.P.; Meij, Senna E.P.; Osman, Karim N.M.M.H.; Kloosterman, Helena A.; van Stralen, Karlijn J.; Grimbergen, Matthijs C.M.; Eijsbouts, Quirijn A.J.; van den Dobbelsteen, John J.; Twinanda, Andru P.

DOI
10.1007/s00464-020-08110-5

Publication date
2020

Document Version
Accepted author manuscript

Published in
Surgical Endoscopy

Citation (APA)
Guédon, A. C. P., Meij, S. E. P., Osman, K. N. M. M. H., Kloosterman, H. A., van Stralen, K. J., Grimbergen, M. C. M., Eijsbouts, Q. A. J., van den Dobbelsteen, J. J., & Twinanda, A. P. (2020). Deep learning for surgical phase recognition using endoscopic videos. *Surgical Endoscopy*, *35 (2021)*(11), 6150-6157. https://doi.org/10.1007/s00464-020-08110-5

Deep learning for surgical phase recognition using endoscopic videos

Authors:

Annetje C.P. Guédon, PhD
Department of Clinical Physics, Spaarne Gasthuis, the Netherlands.

Senna E.P. Meij, MSc
Department of Biomechanical Engineering, Delft University of technology, the Netherlands.

Karim N.M.M.H. Osman, BSc
Department of Biomechanical Engineering, Delft University of technology, the Netherlands.

Helena A. Kloosterman, MSc
Cosmonio, the Netherlands.

Karlijn J. van Stralen, PhD
Spaarne Gasthuis Academie, Spaarne Gasthuis, the Netherlands.

Matthijs C.M. Grimbergen, PhD
Department of Radiology, Amsterdam UMC, the Netherlands

Quirijn A.J. Eijsbouts, PhD
Department of Surgery, Spaarne Gasthuis, the Netherlands.

John J. van den Dobbelsteen, PhD
Department of Biomechanical Engineering, Delft University of technology, the Netherlands.

Andru P. Twinanda, PhD
Cosmonio, the Netherlands.

Corresponding author:

Annetje Guédon

Spaarnepoort 1, 2134TM, Hoofddorp, the Netherlands

+31 6 24 22 59 30

aguedon@spaarnegasthuis.nl

No funding

Abstract

Operating room planning is a complex task as pre-operative estimations of procedure duration have a limited accuracy. This is due to large variations in the course of procedures. Therefore, information about the progress of procedures is essential to adapt the daily operating room schedule accordingly. This information should ideally be objective, automatically retrievable and in real-time. Recordings made during endoscopic surgeries are a potential source of progress information. A trained observer is able to recognize the ongoing surgical phase from watching these videos. The introduction of deep learning techniques brought up opportunities to automatically retrieve information from surgical videos. The aim of this study was to apply state-of-the art deep learning techniques on a new set of endoscopic videos to automatically recognize the progress of a procedure, and to assess the feasibility of the approach in terms of performance, scalability and practical considerations.

A dataset of 33 laparoscopic cholecystectomies (LC) and 35 total laparoscopic hysterectomies (TLH) was used. The surgical tools that were used and the ongoing surgical phases were annotated in the recordings. Neural networks were trained on a subset of annotated videos. The automatic recognition of surgical tools and phases was then assessed on another subset. The scalability of the networks was tested and practical considerations were kept up.

The performance of the surgical tools and phase recognition reached an average precision and recall between 0,77 and 0,89. The scalability tests showed diverging results. Legal considerations had to be taken into account and a considerable amount of time was needed to annotate the datasets.

This study shows the potential of deep learning to automatically recognize information contained in surgical videos. This study also provides insights in the applicability of such a technique to support operating room planning.

Keywords

Surgical phase, deep learning, endoscopic videos, automatic recognition.

Intro

Planning surgical procedures is known to be a complex task. Many factors have to be taken into account, such as the availability of surgeons and qualified operating room (OR) personnel, emergency procedures, constraints regarding limited OR facilities, and the large diversity in patients and procedures [1, 2]. A key element in OR planning is surgery duration [1–5]. In current practice, the prediction of surgery duration is based on average durations, while a large variability is observed for many surgical procedures [2–4]. This causes suboptimal OR planning. Surgical procedures that finish later than expected induce a delay (or cancellation) of the subsequent procedures, longer waiting times for patients and overtime work of the OR personnel. Procedures that finish ahead of planning can cause ORs to remain unnecessarily unused. The variability in surgery duration is monitored by the OR schedulers, either by observation or verbal communication with the OR teams. They estimate case-by-case the remaining procedure duration to adapt the daily schedule accordingly. This requires a vast experience with many procedures and the accuracy of the OR teams' estimation.

Real-time updates on procedure progress and a reliable end time prediction would be valuable to support the OR schedulers in their tasks. Additionally, it is desirable to retrieve this information automatically, instead of adding compulsory registration for the OR team or interrupting the surgical process to communicate with them [6, 7]. The different technologies used in the OR may in fact offer a source of information. For instance, the use of surgical tools and devices can provide information about the progress of the surgery [8–13]. Specifically for endoscopic procedures, the video camera is a valuable source of information as it shows anatomical structures as well as the use of surgical tools. An experienced observer is able to give an indication of the progress of the procedure by watching the videos. In theory, the same information should be automatically retrievable. The challenge is to train systems to do so.

The introduction of machine learning techniques has brought up opportunities to recognize information contained in surgical images. Various studies have worked on retrieving information about the progress of a surgical procedure by recognizing the different surgical phases of the procedure from videos. These studies first extracted visual features from videos and recognized the surgical phases using techniques such as Support Vector Machines, Hidden Markov Models and Dynamic Time Warping [12, 14–16]. The visual features are pre-defined to capture certain information in the images. However, the choice of features might not cover all significant features present in the images. Deep learning techniques, which work without any pre-defined features, overcome this limitation. These have recently been applied by Aksamentov et al. [17] and Twinanda et al. [18, 19] and showed promising results on laparoscopic cholecystectomies and bypass surgeries.

However, these recent deep learning applications used mainly the same large database of procedure, such as laparoscopic cholecystectomies or cataract surgeries [17-20]. In order to investigate the applicability of such a technique for operating room planning, different databases with various types of surgical procedures need to be tested. Moreover, information about the practical considerations that go with deep learning applications is required. The aim of this study is to apply deep learning techniques on a new set of endoscopic videos, and to assess the feasibility of the technique in medical practice, in terms of performance, scalability and practical considerations.

Material & methods

Dataset

The dataset consists of endoscopic videos of 33 laparoscopic cholecystectomies (LC) and 35 total laparoscopic hysterectomies (TLH), which were performed between 2016 and 2019 in the Spaarne Gasthuis, a teaching hospital in the Netherlands. The videos were obtained using a recording and archiving system (Dax archiving solutions, the Netherlands). The laparoscopic cholecystectomies were performed by five general surgeons and the total laparoscopic hysterectomies by three gynecologists. For both procedures, the definitions of surgical phases and corresponding usage of surgical tools were based on literature [12, 18, 21] and adapted during meetings with performing surgeons and gynecologists. Extensive bleeding was defined as a possible additional phase that could occur during any of the phases. The surgical tools used during the procedures are shown in table 1 and the surgical phases are shown in table 2.

Deep learning

For this deep learning application, a neural network was trained to recognize information contained in surgical images. The first step was to annotate one frame per second of all the videos. The tools that were present in a frame, as well as the ongoing surgical phase, were annotated using the NOUS application (COSMONiO, the Netherlands, www.cosmonio.com). NOUS is an interactive AI platform which allows to annotate data and train a deep neural network. The annotation of the tools was performed by instructed students, the annotation of the phases was performed by two of the authors. None of them were performing surgeons, however they received instructions to enable correct annotations of tools and phases. The tools were annotated per frame and phases were annotated by selecting the frames corresponding to the same phase.

The second step was to perform five random splits in the dataset for both procedures. For the LC procedures, each split contained a training subset of 24 videos and a validation subset of nine videos. For the TLH procedures, each split contained a training subset of 26 videos and a validation subset of nine videos. This step was done to average the performance results so that they are less dependent of a single split.

The third step was to train the neural network, for both procedures, with the annotated frames of the training subsets, in order to be able to recognize the same type of information in the validation subsets. The network was not previously trained on any other surgery data. In this study, an InceptionV3 network with ResNet50 backbone was implemented and adapted to fit the needs of this application, this is a specific architecture of convolutional neural network [22]. The network ran on a dedicated server (NOUS Learner, COSMONiO, the Netherlands), which was located in the hospital. The training of the network ran for 50 epochs with batches of 32 frames. Scaling, shifting, rotation and color transformation were used as data augmentation. No additional information about characteristics of the patients and OR team was provided to the network. No temporal information about the progress of the procedure was provided to the network. It only used visual information contained in the frames of the videos.
The fourth step was to let the trained network automatically recognize tools and phases in the frames of the validation subsets. The annotations of the validation subsets were not used to train the network, they were only used to evaluate the performance of the automatic phase and tool recognition.
The last step was to perform temporal smoothing on the automatic recognition of the phases in order to avoid shifting between phases for only a few frames. For instance, as the network only used visual information, a few frames without instruments present in the frames may cause a wrong phase recognition. Therefore, a window of 15 to 30 frames was used for the smoothing on the automatic recognition of the phases. When the network was unsure between different phases (which was defined with a probability vector smaller than 0.8), the recognized phase was adapted to the one recognized the most often in the previous 15 to 30 frames.
When the network was sure of the phase that corresponded to a certain frame (which was defined with a probability vector larger than 0,8), the recognized phase was kept as is.

Feasibility assessment

Performance

The feasibility was assessed in terms of performance of the automatic recognition of surgical tools and phases for both procedures. The recognized tools and phases were compared with the manually annotated tools and phases in the validation subsets. From this comparison, a confusion matrix was calculated which showed the true positives (TP), true negative (TN), false positives (FP) and false negatives (FN) recognitions. In deep learning, the accuracy, precision and recall are commonly used performance metrics. The accuracy is the percentage of frames that are recognized correctly by the network (see equation 1). This metric is appropriate if the data is balanced (when the number of negatives and positives is similar). In this case, the data was unbalanced. There were many frames where a certain tool was not present, resulting in a very large proportion of TN. Therefore, the accuracy was not representative for the actual performance of the network. The precision is the percentage of the frames in which the network recognized a tool or phase, where the network recognized it correctly (see equation 2). The recall, which is the same as the metric sensitivity, is the percentage of the frames in which a tool or phase is actually present, where the network recognized it correctly (see equation 3).

$$Accuracy = (TP+TN) / (TP+TN+FP+FN) \quad \text{equation 1}$$

$$Precision = TP / (TP+FP) \quad \text{equation 2}$$

Recall = TP / (TP+FN) *equation 3*

The precision and recall were evaluated per tool and phase individually for each recording of the validation subsets. The average precision and recall were then calculated per tool and phase over all the recordings of the validation subsets. Finally, the weighted average precision and recall were calculated in order to provide an average performance of the network for both procedures. The weighting was done on the numbers of frames in which a specific tool or phase was actually present.

Scalability

The scalability is assessed by evaluating the performance of a network, that is trained on a certain type of surgical procedure, to automatically recognize information in other procedures. A scalable network would ask less effort to be trained on other types of procedures, as the information that is common to both the procedure used to train the network and the other procedures would have already be added to the network. In this study, the scalability was assessed for the network trained on LC procedures and the network trained on TLH procedures. The assessment was performed on the recognition of tools that are used in both the LC and TLH procedures. These were the grasper, scissors, monopolar hook and irrigation & suction device. The tool recognition was again evaluated by calculating the average precision and recall per tool over all recordings of the validation subsets.

Practical considerations

In order to assess the feasibility of such a technique in healthcare, practical considerations have to be taken into account. The efforts needed to annotate the datasets in order to train the network were assessed in terms of time. Moreover, the amount of data needed for the network to perform adequately was assessed. The numbers of frames in the training subsets that were annotated with a surgical tool or phase were compared to the performance in the recognition of this tool or phase in the validation subsets.

Results

Performance

The performance of the automatic recognition of surgical tools and surgical phases for both procedures types is shown in table 3 and 4. For the LC procedures, the weighted average precision and recall were respectively 0.89 and 0.81 for the surgical tool recognition, and respectively 0.79 and 0.77 for the surgical phase recognition. For the TLH procedures, the weighted average precision and recall were respectively 0.88 and 0.79 for the surgical tool recognition, and respectively 0.78 and 0.79 for the surgical phase recognition.

The performance of the phase recognition after temporal smoothing is shown in figure 1a and 1b, and figure 2a and 2b. The annotated surgical phases are shown in red, the automatically recognized phases in blue. A perfectly performing automatic phase recognition would show a complete overlap of the red and blue lines. Figures 1a and 2a show the results of the phase recognition before the temporal smoothing. Figures 1b and 2b show the ones after the temporal smoothing.

Scalability

The scalability of the trained networks of both procedures is shown in table 5. The performance of the automatic recognition of surgical tools with the network trained on LC procedures was assessed on TLH procedures and vice versa.

Practical considerations

The use of endoscopic videos was approved by the hospital. In this study, all videos were kept within the hospital. The annotations and the networks trained on the videos remained the property of the hospital. After getting access to the videos, the annotations of the surgical tools and phases was performed. This process took on average 3.5 times the length of the video. For instance, a video of 100 minutes took about 350 minutes to annotate.

The amount of annotated data needed to reach a certain performance in the recognition of surgical tools and phases was assessed. A precision and recall higher than 0.8 was reached by annotating at least 20000 frames with a surgical tool or phase in the training subsets. However, annotating at least 10000 frames showed already stable results and almost all precision and recall results were higher than 0.8. This applied to the following surgical tools: the grasper and hook for the LC procedures, and the grasper, ligasure, needle feeder, needle & thread, and uterus mobiliser for the TLH procedures. Moreover, this applied to the following surgical phases: 'preparation and dissection' and 'gallbladder dissection' for the LC procedures, and 'uterus dissection', 'uterus separation from the vagina' and 'vaginal cuff closure'. Surgical tools and phases that were present in less than respectively 1000 and 2000 frames in the training subsets resulted in a very low precision and recall. This applied to the drain for the LC procedures, and the scissors and bag for the TLH procedures. This also applied to the closing phase in both procedures.

Discussion

This study applied deep learning techniques on a new set of laparoscopic cholecystectomy and total laparoscopic hysterectomy videos, and aimed to recognize automatically surgical tools and phases in these videos. The performance of the neural network in terms of automatic recognition was very similar for both procedures. The weighted average recall and precision were between 0.77 and 0.89, which shows that such a technique is suited to retrieve information from these types of endoscopic procedures automatically. This study strengthens the idea that this method is suitable for various types of surgical procedures, which show weighted average recall and precision were between 0.75 and 0.91 [17–19]. It also shows the possibility to provide automated information from the OR to the planning team, without interruptions of the surgical process or involvement of any member of the surgical team.

The neural network was based on visual features only. A temporal smoothing was applied instead of using a more complicated temporal methodology (such as Long Short Term Memory networks [23, 24]) in order to improve the phase recognition. Variation in performance results was observed in the recognition of surgical tools and phases. Stable and high recall and precision results were obtained by training the network on at least 10000 annotated frames. This does not necessarily mean that more frames need to be annotated to get usable results for medical practice. Other studies show that there is not a great improvement of the recall and precision with a large increase of recordings (and consequently a large increase of frames) [17-19]. Besides, this is definitely dependant of the intended use of the network. For the goal of optimizing the OR planning, the recognition of some phases is more important than others. The accurate recognition of e.g. the closing phase is less important than the recognition of earlier phases. The reason is that the actions needed to be taken for an optimal OR planning (such as preparing the next patient) are already initiated at that moment. Another example is the clipping & cutting phase for the LC procedures. This phase was not well recognized but the recognition of the previous and following phases presented good results. The clipping & cutting phase is short, therefore not many frames could be used to train the network. The influence of the recognition of this phase on OR planning optimization would be minimal as the recognition of the previous and following phase would most certainly provide the needed information. The most significant information would be the recognition of the phases that vary the most in time and thereby influence the most the daily OR planning. A phase that is expected to influence largely the progress of a procedure is the possible additional phase 'bleeding'. The recognition of bleeding was low for both procedures, despite the clear visual features that goes with a bleeding. The network most certainly did not have enough frames annotated with bleeding to be trained on, as bleeding did not occur often in the videos of the database.

Regarding the scalability of the used method, the automatic recognition of surgical tools with the network trained on another procedure showed a decrease in performance. The grasper and irrigator, which both presented high scores in recall and precision with the network trained on the same procedure, showed decreased but still comparable results. This shows that networks might be used in the future for different types of procedures. The hook, which was expected to present comparable scores as well, did not. Two different brands, with shafts of different colors, were unfortunately used for the LC and TLH procedures. This shows the limitations of the scalability of such a network. At last, the scissors, which presented unsatisfactory results with the network trained on the same procedure, showed a dramatic decrease in performance with the network trained on the other procedure. The network should be trained with more frames containing scissors to see if the results can be improved.

Getting access to endoscopic videos for research purposes involves practical concerns that differ between countries and between hospitals. Some considerations have to be taken into account, such as the consent of patients to use the videos for research purposes, the prerequisites to be able to use the (pseudonymised) videos retrospectively without consent, or the privacy-by-design principle for future implementation of a dedicated system [25]. The property of the videos, the annotations and the networks trained on the videos can as well be an area of concern. Moreover, the 'black box' which represents deep learning algorithms can be an obstacle for acceptance in medical practice [26]. Another practical consideration is the large amount of time needed to annotate the dataset. However, other approaches can be used in the future to make the use of deep learning more feasible in medical practice. The tools annotations in this study were used indirectly for the phase recognition, but they may be dispensable in applications for which phase recognition only is required. Annotating merely the phases would considerably decrease the annotation time, as the annotation can be done by selecting at once a large amount of frames corresponding to one surgical phase.

There are some limitations to this study. To further test the applicability in medical practice, datasets of various surgery types are needed to train the networks and to test the phase recognition and the scalability of the networks. Furthermore, the amount of frames per phase to train the networks should be less variable. For all phases that are important for the intended use, a minimum of 10000 frames should be used to train the networks. Different ways of annotating should also be investigated to establish a faster method. Adding additional information to the networks, such as patient information or data of other equipment in the OR, should as well be investigated. Last but not least, several steps still need to be taken in order to use this method to improve the OR planning. The remaining surgery time needs to be predicted in real-time based on the automatic phase recognition, and the interaction of such a prediction system with the OR staff needs be designed.

To conclude, this study offers insights in the applicability of deep learning techniques and the feasibility of phase and tool recognition. Future work will focus on estimating the remaining surgery duration based on this automatic phase recognition. Moreover, the required accuracy of the remaining surgery duration estimation needs to be assessed in medical practice in order to successfully implement such a system in ORs.

Disclosure

A.C.P. Guédon, S.E.P. Meij, K.N.M.M.H. Osman, H.A. Kloosterman, K.J. van Stralen, M.C.M. Grimbergen, Q.A.J. Eijsbouts, J.J. van den Dobbelsteen and A.P. Twinanda have no conflict of interest or financial ties to disclose.

References

1. Eijkemans MJC, Van Houdenhoven M, Nguyen T, et al (2010) Predicting the unpredictable: A new prediction model for operating room times using individual characteristics and the surgeon's estimate. Anesthesiology. https://doi.org/10.1097/ALN.0b013e3181c294c2
2. Dexter F, Ph D, Epstein RH, et al (2017) Making Management Decisions on the Day of Surgery Based on Operating Room Efficiency and Patient Waiting. 1444–1453
3. Edelman ER, Kuijk SMJ Van, Hamaekers AEW, et al (2017) Improving the Prediction of Total Surgical Procedure Time Using Linear Regression Modeling. 4:1–5. https://doi.org/10.3389/fmed.2017.00085
4. Eijk RPA van, Veen-berkx E Van, Kazemier G, Eijkemans MJC (2016) Effect of Individual Surgeons and Anesthesiologists on Operating Room Time. https://doi.org/10.1213/ANE.0000000000001430
5. Gupta N, Ranjan G, Arora MP, et al (2013) Validation of a scoring system to predict difficult laparoscopic cholecystectomy. Int J Surg 11:1002–1006. https://doi.org/10.1016/j.ijsu.2013.05.037
6. Wiegmann DA, ElBardissi AW, Dearani JA, et al (2007) Disruptions in surgical flow and their relationship to surgical errors: An exploratory investigation. Surgery 142:658–665. https://doi.org/10.1016/j.surg.2007.07.034
7. Arora S, Hull L, Sevdalis N, et al (2010) Factors compromising safety in surgery: stressful events in the operating room. Am J Surg 199:60–65. https://doi.org/10.1016/j.amjsurg.2009.07.036
8. Blum T, Padoy N, Feußner H, Navab N (2008) Modeling and online recognition of surgical phases using hidden Markov models. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) 5242 LNCS:627–635. https://doi.org/10.1007/978-3-540-85990-1-75
9. Guédon ACP, Paalvast M, Meeuwsen FC, et al (2016) 'It is Time to Prepare the Next patient' Real-Time Prediction of Procedure Duration in Laparoscopic Cholecystectomies. J Med Syst. https://doi.org/10.1007/s10916-016-0631-1
10. Meeuwsen FC, van Luyn F, Blikkendaal MD, et al (2019) Surgical phase modelling in minimal invasive surgery. Surg Endosc. https://doi.org/10.1007/s00464-018-6417-4
11. Padoy N, Blum T, Ahmadi SA, et al (2012) Statistical modeling and recognition of surgical workflow. Med Image Anal 16:632–641. https://doi.org/10.1016/j.media.2010.10.001
12. Blum T, Feußner H, Navab N (2010) Modeling and segmentation of surgical workflow from laparoscopic video. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) 6363 LNCS:400–407. https://doi.org/10.1007/978-3-642-15711-0_50
13. Bouarfa L, Jonker PP, Dankelman J (2011) Discovery of high-level tasks in the operating room. J Biomed Inform 44:455–462. https://doi.org/10.1016/j.jbi.2010.01.004
14. Lalys F, Riffaud L, Morandi X, Jannin P (2011) Surgical phases detection from microscope videos by combining SVM and HMM. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)
15. Volkov M, Hashimoto DA, Rosman G, et al (2017) Machine learning and coresets for automated real-time video segmentation of laparoscopic and robot-assisted surgery. Proc - IEEE Int Conf Robot Autom 754–759. https://doi.org/10.1109/ICRA.2017.7989093
16. Lalys F, Riffaud L, Bouget D, Jannin P (2012) A framework for the recognition of high-level surgical tasks from video images for cataract surgeries. IEEE Trans Biomed Eng 59:966–976. https://doi.org/10.1109/TBME.2011.2181168
17. Aksamentov I, Twinanda AP, Mutter D, et al (2017) Deep neural networks predict remaining surgery duration from cholecystectomy videos. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)
18. 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

19. Twinanda AP, Yengera G, Mutter D, et al (2018) RSDNet: Learning to Predict Remaining Surgery Duration from Laparoscopic Videos Without Manual Annotations. IEEE Trans Med Imaging 1–10. https://doi.org/10.1109/TMI.2018.2878055

20. Yu F, Croso SG, et al (2019) Assessment of Automated Identification of Phases in Videos of Cataract Surgery Using Machine Learning and Deep Learning Techniques. JAMA Netw. open. https://doi.org/10.1001/jamanetworkopen.2019.1860

21. Blikkendaal MD, Driessen SRC, Rodrigues SP, et al (2017) Surgical flow disturbances in dedicated minimally invasive surgery suites: an observational study to assess its supposed superiority over conventional suites. Surg Endosc. https://doi.org/10.1007/s00464-016-4971-1

22. Szegedy C, Vanhoucke V, Ioffe S, et al (2016) Rethinking the Inception Architecture for Computer Vision. IEEE Conf. Comput. Vis. Pattern Recognit. https://doi.org/10.1109/CVPR.2016.308

23. Nwoye CI, Mutter D, Marescaux J, Padoy N (2019) Weakly supervised convolutional LSTM approach for tool tracking in laparoscopic videos. Int J Comput Assist Radiol Surg 14:. https://doi.org/10.1007/s11548-019-01958-6

24. Chen W, Feng J, Lu J, Zhou J (2018) Endo3D: Online workflow analysis for endoscopic surgeries based on 3D CNN and LSTM. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)

25. Van Dalen ASHM, Legemaate J, Schlack WS, et al (2019) Legal perspectives on black box recording devices in the operating environment. Br. J. Surg. 106(11):1433–144. http://doi.org/10.1002/bjs.11198

26. Gordon L, Grantcharov T, Rudzicz F (2019) Explainable Artificial Intelligence for Safe Intraoperative Decision Support. JAMA Surg. 154(11):1064–1065. http://doi.org/10.1001/jamasurg.2019.2821

Tables and figures

Table

Laparoscopic cholecystectomy	Total laparoscopic hysterectomy
Grasper	Grasper
Scissors	Scissors
Monopolar hook	Monopolar hook
Irrigator & suction device	Irrigation & suction device
Clipper	Ligasure
Bag	Uterus mobiliser
Drain	Morcellator
	Bag
	Needle feeder
	Needle & thread

Table

Phase	Laparoscopic cholecystectomy	Total laparoscopic hysterectomy
1	Trocars & tools insertion	Trocars & tools insertion
2	Preparation & dissection	Uterus dissection
3	Clipping & cutting	Uterus separation from the vagina
4	Gallbladder dissection	Uterus retrieval: transvaginal
5	Gallbladder packaging & retrieval	Uterus retrieval: morcellation
6	Final check & irrigation	Vaginal cuff closing
7	Closing & desufflation	Final check & irrigation
8		Closing & desufflation
Additional	Bleeding	Bleeding

Table

Laparoscopic cholecystectomy		
Tools	**Precision**	**Recall**
Grasper	0.90	0.86
Clipper	0.84	0.65
Scissors	0.72	0.44
Hook	0.92	0.80
Bag	0.78	0.65
Irrigator	0.84	0.68
Drain	0.23	0.02
Weighted average	**0.89**	**0.81**
Phases		
Trocars & tools insertion	0.69	0.95
Preparation & dissection	0.86	0.88
Clipping & cutting	0.56	0.29
Gallbladder dissection	0.79	0.74
Gallbladder packaging & retrieval	0.57	0.52
Final check & irrigation	0.50	0.30
Closing & desufflation	0.45	0.11
Weighted average	**0.79**	**0.77**
Additional phase: Bleeding	0.63	0.30

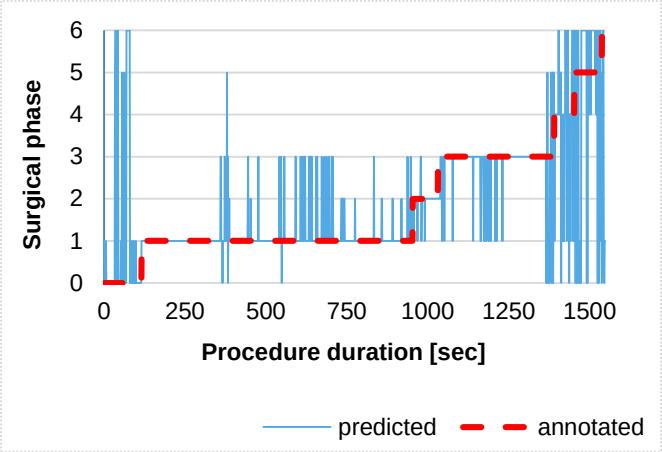
Table

Total laparoscopic hysterectomy		
Tools	**Precision**	**Recall**
Grasper	0.89	0.81
Ligasure	0.87	0.89
Hook	0.95	0.87
Uterus mobiliser	0.87	0.70
Morcellator	0.72	0.46
Bag	0.00	0.00
Needle feeder	0.94	0.82
Needle & thread	0.92	0.84
Scissors	0.46	0.10
Irrigator	0.82	0.53
Weighted average	**0.88**	**0.79**
Phases		
Trocars & tools insertion	0.81	0.91
Uterus dissection	0.87	0.89
Uterus separation from the vagina	0.69	0.83
Uterus retrieval: transvaginal	0.42	0.27
Uterus retrieval: morcellation	0.78	0.55
Vaginal cuff closing	0.80	0.84
Final check & irrigation	0.71	0.48
Closing & desufflation	0.39	0.08
Weighted average	**0.78**	**0.79**
Additional phase: Bleeding	0.00	0.00

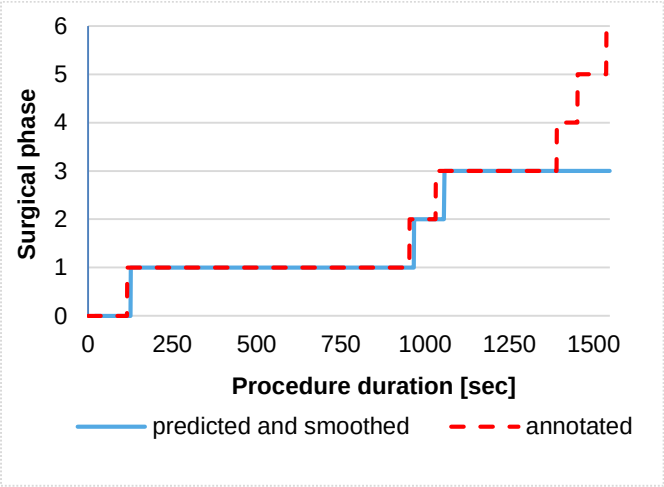
Table

	LC network		TLH network	
Tools	**Precision**	**Recall**	**Precision**	**Recall**
Grasper	0.74	0.82	0.85	0.67
Scissors	0.03	0.03	0.65	0.09
Hook	0.35	0.07	0.04	0.01
Irrigator	0.51	0.41	0.47	0.70
Weighted average	**0.63**	**0.63**	**0.76**	**0.52**

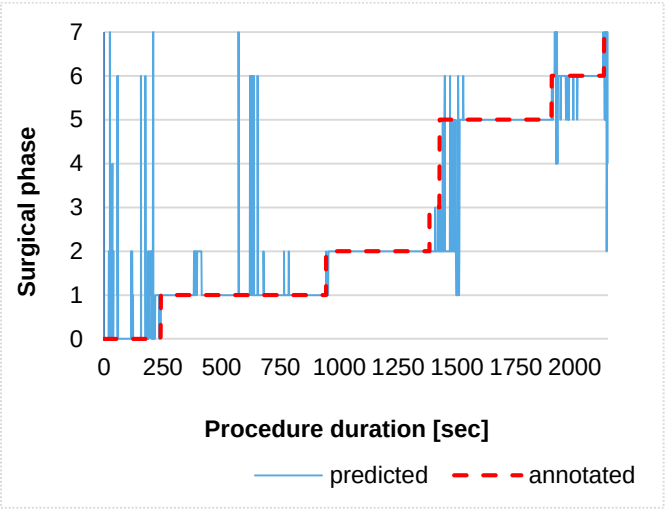
Figure


Surgical phase
6
5
4
3
2
1
0
0
250
500
750
1000
1250
1500
Procedure duration [sec]
predicted
annotated

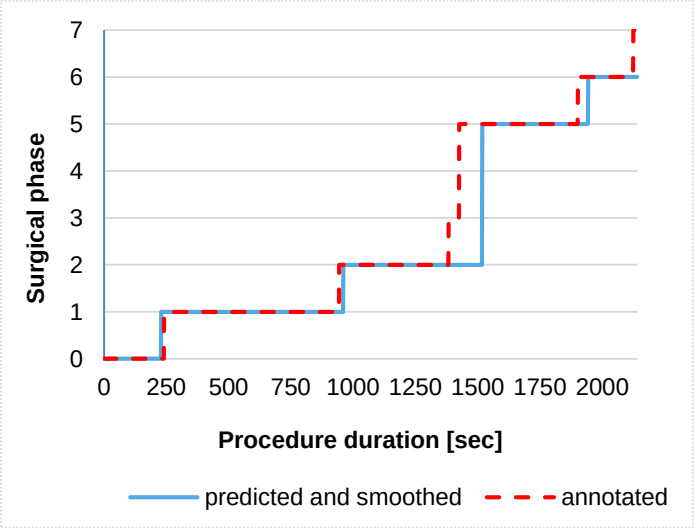
Figure


Surgical phase
6
5
4
3
2
1
0
0
250
500
750
1000
1250
1500
Procedure duration [sec]
predicted and smoothed
annotated

Figure


Surgical phase
7
6
5
4
3
2
1
0
0
250
500
750
1000
1250
1500
1750
2000
Procedure duration [sec]
predicted
annotated

Figure


Surgical phase
7
6
5
4
3
2
1
0
0
250
500
750
1000
1250
1500
1750
2000
Procedure duration [sec]
predicted and smoothed
annotated