Assisted Annotation of Surgical Videos Using Deep Learning

Gurvan Lecuyer, Martin Ragot, Nicolas Martin, Laurent Launay, Pierre Jannin

International Journal of Computer Assisted Radiology and Surgery

Assisted Annotation of Surgical Videos Using Deep Learning

--Manuscript Draft--

Manuscript Number:	CARS-D-19-00052
Full Title:	Assisted Annotation of Surgical Videos Using Deep Learning
Article Type:	Original Article
Keywords:	Assisted annotation; Surgical Workflow; Phase recognition; Deep Learning; User test
Corresponding Author:	Gurvan Lecuyer IRT b<>com Rennes, FRANCE
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	IRT b<>com
Corresponding Author's Secondary Institution:	
First Author:	Gurvan Lecuyer
First Author Secondary Information:	
Order of Authors:	Gurvan Lecuyer
	Martin Ragot
	Nicolas Martin
	Laurent Launay
	Pierre Jannin
Order of Authors Secondary Information:	
Funding Information:	
Abstract:	Training deep learning algorithms requires a huge amount of labeled data. Annotation of medical data is a challenging task, as it needs specific knowledge. In this paper, we propose an assistance system to annotate the phase in surgical videos. A convolutional neural network (CNN) was trained to perform an initial frame-by-frame phase recognition with a post-processing method to produce a pre-annotation. We conducted a user study to validate the proposed assistance system. Different measurements were recorded to objectively assess the proposed system: annotation time, accuracy of the annotations performed by the participants and a subjective evaluation based on the Technology Acceptance Model (TAM 3) questionnaire. Two metrics in particular were assessed: perceived usefulness and perceived ease of use of the system. 31 volunteers participated to this study. The results showed that the assistance system significantly improved the annotation accuracy while it had no influence on the annotation time. The results of the questionnaire showed that the participants found the assistance system more useful than the manual system whereas the perceived ease of use of both systems were similar.

Noname manuscript No.
(will be inserted by the editor)

Assisted Annotation of Surgical Videos Using Deep Learning

Gurvan Lecuyer · Martin Ragot · Nicolas Martin · Laurent Launay · Pierre Jannin

Received: date / Accepted: date

Abstract
Purpose Training deep learning algorithms requires a huge amount of labeled data. Annotation of medical data is a challenging task, as it needs specific knowledge. In this paper, we propose an assistance system to annotate the phase in surgical videos.
Methods A convolutional neural network (CNN) was trained to perform an initial frame-by-frame phase recognition with a post-processing method to produce a pre-annotation. We conducted a user study to validate the proposed assistance system. Different measurements were recorded to objectively assess the proposed system: annotation time, accuracy of the annotations performed by the participants and a subjective evaluation based on the Technology Acceptance Model (TAM 3) questionnaire. Two metrics in particular were assessed: perceived usefulness and perceived ease of use of the system.
Results Thirty one volunteers participated to this study. The results showed

Gurvan Lecuyer
IRT b< >com 1219 avenue des Champs Blancs, 35510 Cesson-Sevigne
Univ. Rennes, INSERM, LTSI-UMR 1099, F-35000 Rennes, France
E-mail: gurvan.lecuyer@b-com.com

Martin Ragot
IRT b< >com 1219 avenue des Champs Blancs, 35510 Cesson-Sevigne
E-mail: martin.ragot@b-com.com

Nicolas Martin
IRT b< >com 1219 avenue des Champs Blancs, 35510 Cesson-Sevigne
E-mail: nicolas.martin@b-com.com

Laurent Launay
IRT b< >com 1219 avenue des Champs Blancs, 35510 Cesson-Sevigne
E-mail: laurent.launay@b-com.com

Pierre Jannin
Univ. Rennes, INSERM, LTSI-UMR 1099, F-35000 Rennes, France
E-mail: pierre.jannin@univ-rennes1.fr

that the assistance system improved the annotation accuracy by 5% while it had no influence on the annotation time. The results of the questionnaire showed that the participants found the assistance system more useful than the manual system whereas the perceived ease of use of both systems were similar.
Conclusion The proposed assistance system improved significantly the annotation accuracy of the users when compared with the manual annotation.

Keywords Assisted annotation · Surgical workflow · Phase recognition · Deep Learning · User test

1 Introduction

From the microscope used to perform a cataract operation to the laparoscope used in Minimally Invasive Surgical (MIS) procedures, video in the Operating Room (OR) is a key sourcec of information. With the growing use of recent Machine Learning approaches and surgical data science, situation-awareness methods have been developed for intra-operative assistance, surgical education and OR management. For situation awareness, automatic recognition of activities in the OR is a technological deadlock. Recently, some methods have been proposed for cataract [18, 20], heart surgery [2], laparoscopic procedures and especially cholecystectomy interventions [7, 15, 19, 25, 29].
The development of such Machine Learning-based based approaches requires an extensive amount of precisely annotated data for learning. Different factors make it difficult such as patient privacy, medical staff approval, time-consuming manual annotations and high inter-patient and inter-surgeon variability.
Over the few years, some studies attempted to tackle this issue using unsupervised learning [4], or self-supervised learning [11]. However, these models still need annotated data at some points of the training steps or, at least, for testing purposes.
Assisted annotation tools have been studied for several tasks such as text annotation [6, 14, 23, 24, 28], audio annotation [3, 26], genome annotation [10, 21], medical image segmentation [5, 9, 16] and surgical workflow [12].
In this paper, we propose a new approach to facilitate annotation by using a convolutional neural network to infer labels in order to pre-annotate unlabeled data. We conducted a user study to assess the usefulness and efficiency of the proposed assistance system for the surgical phase annotation task.

2 Materials and Method

2.1 *Surgery Workflow Toolbox* [Annotate]

Two experiments were conducted using the same software named *Surgery Workflow Toolbox* [Annotate]. The latter is based on the work of Garraud et

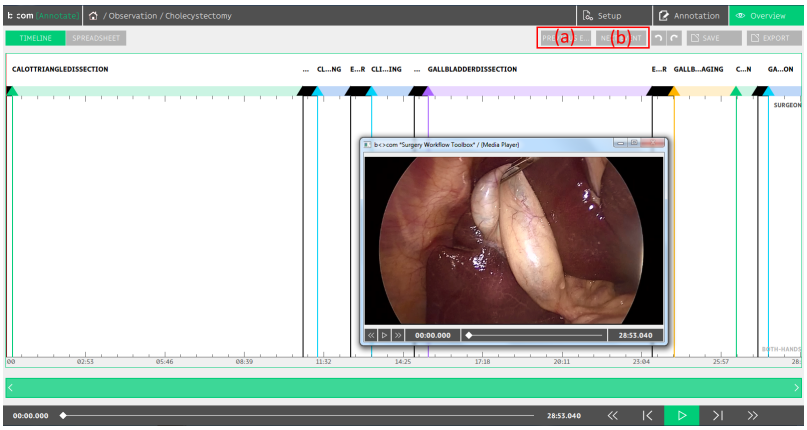

Fig. 1 User interface of [Annotate]. The two added buttons are shown in the red rectangle where (a) is the button for navigating to the previous transition and (b) is the button for navigating to the next transition

al. [12], in the remaining of this paper, we will refer to the software as [Annotate]. This software was developed by b<>com Technology Research Institute and enables surgeons to annotate surgical videos including phases, steps and activities of surgeries based on ontologies [13]. The user interface of the software is shown in figure 1.
For the purpose of this paper, new features were added to [Annotate] to allow assisted annotation. A pre-annotation is performed using a VGG19 architecture [22] trained with our data. Two buttons were added to navigate between the phases predicted by the neural network as it can be seen inside the red rectangle in figure 1. These buttons allowed the users to go to the beginning of the next or previous phase suggested by the assistance system, to quickly jump from a transition to another one.

2.2 Data

We used videos of cholecystectomy procedures for the user study. Videos were annotated by a single person with the name of the phases. The cholecystectomy procedure is composed of seven phases: preparation, calot triangle dissection, clipping and cutting, gallbladder dissection, gallbladder packaging, cleaning and coagulation, and gallbladder retraction. We chose two videos with similar duration to conduct our study. Videos A and B last respectively 17min17s and 17min26s. The details of both videos are summarized in the table 1.

2.3 Network Architecture

We chose to use the VGG19 architecture since it displays good results on the ImageNet classification task [17] while being fast to train compared to ResNet

Phases	Video A duration (s)	Video B duration (s)
Preparation	21	35
Calot triangle dissection	461	454
Clipping and cutting	79	80
Gallbladder dissection	327	292
Gallbladder packaging	54	66
Cleaning and coagulation	46	56
Gallbladder retraction	49	63
Total duration	1037	1046

Table 1 Duration of each phase and total duration of the videos A and B

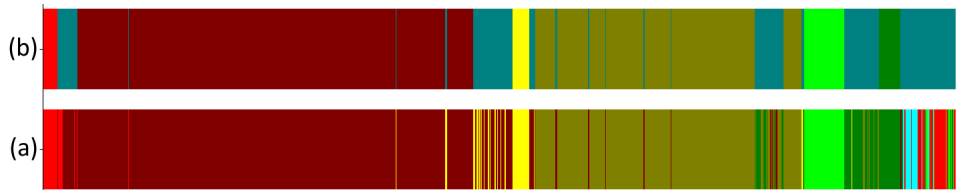

Fig. 2 Visualization of the frame-by-frame phase prediction, each color representing a phase. (a) Prediction computed by the VGG19 architecture. (b) Prediction after passing the error detector (the errors are represented in gray).

or InceptionV3. The architecture is composed of 6 convolutional layers, each followed by a maxpool layer. Two fully-connected layers of size 4,096 and a softmax layer complete the architecture. For the phase recognition task, we modified the size of the softmax output to 7; one for each possible class. We implemented this network using the framework Keras [8], with TensorFlow as backend [1].
The fully-connected layers were randomly initialized and two dropout layers of 0.5 were added for the training step. The previous layers were initialized with the weights obtained with the ImageNet classification task. All the layers were then trained using a learning rate set to 0.001. The training was achieved by using batches of 32 frames. The frames were resized to a resolution of 224x224 resolution as required by the input layer of VGG19.

2.4 Noise detector

A noise detector was implemented in order to increase the accuracy of the network. A threshold value T was defined with the assumption that a surgical phase has a minimum duration which cannot be under T. A subset of successive frames predicted with the same label was considered as a detected phase. If a detected phase lasts less than T, then it was considered as noise, as illustrated in gray in the figure 2. It has to be noted that a residual part of the misclassified frames are not detected while they should have. The threshold value T was dynamically computed for each video and it was based on the total length of the video.

Error category	Average highest confidence value (%)
Detected noise	77.58± 20.23
Not-detected noise	93.08± 12.85
Good prediction	95.84± 10.98

Table 2 Average highest confidence value computed by the network for the different categories of frames.

2.5 Post-processing

We developed a post-processing method to clean the detected noise. This approach is based on the confidence value computed by the network. As it can be seen in the table 2, the average highest confidence-value computed by the network is lower for the frames that were detected as noise than for the well-classified frames. Based on this observation, we used a sliding window w to infer temporal coherence, by locally summing the confidence-value vector computed by the network of the frames around each frame detected as noise. The label corresponding to the highest value of the summed vector was used as the final label. In [30], Zia et al. developed a similar process but the authors used a median filter in order to avoid local prediction incoherence.

The proposed post-processing method was developed to provide an assistance system for annotation. For each detected noise, the participants were asked to check the frames that raised an issue in order to correct the suggested annotation. The objective was to decrease the number of phases detected as noise in order to decrease the number of alerts for the users while keeping a high level of confidence in the prediction.

To provide an efficient assistance, the noise was corrected only if the certainty in the prediction was sufficient. Remaining detected noise was left to the user for checking. The uncertainty was defined as a threshold value, denoted as D. If the difference between the highest value of the summed vector and the second highest value of this vector was lower than D, the detected noise was unaffected so as to be corrected. Otherwise, the label corresponding to the highest confidence value was defined as the label. The final label attribution follows equation 1.

$$L_t = \begin{cases} max(\sum_{n=t-w}^{t+w} C_n) & \text{if } V_1 - V_2 < D \\ Error & \text{else} \end{cases} \quad (1)$$

With $0 \leq n \leq N$and $0 \leq t \leq N$, where t is the current frame, w is the window size, C_n is the outputted confidence value vector for the frame n, V_1 is the highest value of the sum $\sum_{n=t-w}^{t+w} C_n$ and V_2 the second one.

	Group A1	Group B1	Group A2	Group B2
First annotation	Video A manual	Video B assisted	Video B manual	Video A assisted
Second annotation	Video B assisted	Video A manual	Video A assisted	Video B manual

Table 3 The four groups of participants of the user study

3 User study

3.1 Objective

The objective was to assess the assistance system for a task of phase annotation when performed by non-expert users. Our hypothesis is that the proposed functionality will increase annotation accuracy and reduce annotation time. For accuracy, our reference was the annotation provided with Cholec80 database.

3.2 Participants

21 volunteers (5 women, 17 men, age 36± 11 years) took part to the study. 8 participants already used [Annotate] once. None of them had knowledge about the cholecystectomy procedure.

3.3 Experimental design

In order to test our approach, the participants were asked to perfom two annotations: one using the manual system, called "manual annotation", and one using the proposed assistance system, called "assisted annotation".
During the manual annotation task, the participants had to annotate the video from scratch. With assistance, [Annotate] displayed the different phases in the timeline. The participants were asked to check the proposed annotation and, in the case of an error, they had to modify or adjust the transition or the label of the phase.
Every participant performed annotations with both systems in randomized order. The difference of order was used to balance the two tasks, which allowed to limit the learning bias. Half of the participants (n=16) started to annotate without assistance while the other half (n=15) started to annotate with assistance. Two videos were used to assess our hypothesis, which led to four possible sets of tests, which are summarized in the table 3.

3.4 Collected data and metrics

After each annotation, the participants were asked to answer a short questionnaire. The latter contained eight questions taken from the Technology Acceptance Model (TAM 3) proposed by Venkatesh et al. in [27]. These items

TAM3 Dimension	Items
Perceived usefulness	Using the system improves my performance in my job Using the system in my job increases my productivity Using the system enhances my effectiveness in my job I find the system to be useful in my job
Perceived ease of use	My interaction with the system is clear and understandable Interacting with the system does not require a lot of my mental effort I find the system to be easy to use I find it easy to get the system to do what I want it to do

Table 4 Questions used to assess both perceived usefulness and perceived ease of use

correspond to the dimensions of Perceived Usefulness (PU) and Perceived Ease of Use (PEOU). They are depicted in table 4. [Annotate] targetted user are health professional, the participants were asked to imagine themselves as surgeon to answer the questionnaire. The questions were randomly sorted when presented to the participants. The answers of the questionnaire were used to assess how the assistance was perceived by the users in term of ease of use and usefulness. The time taken by the user to complete each annotation was measured via the log files of [Annotate]. Finally, the annotations made by the users were used to control that the participants had understood the task and to objectively measure their performance. The recorded annotations were compared with reference provided by the Cholec80 database. Two metrics were used to assess the performance of the users: accuracy and transition delay. Accuracy was measured using a frame-by-frame comparison between the reference and the annotation performed by the user. The transition delay is the sum of the timespan between the transition time of the reference and the transition time annotated by the user for each transition of the procedure.

3.5 Procedure

The user test included three steps. During the first step, all the participants received a presentation of the cholecystectomy procedure including details about the seven phases they would have to label. For this purpose, two documents were provided to them: one where each phase was explained, and the other showed illustrations of the instruments used to perform the surgery. The participants were allowed to refer to these documents during the annotation task.
To perform the first annotation, the software interactions were introduced to the participants. Then they were asked to perform the annotation (respectively with or without assistance). At the end of this first task, participants were asked to fill-in the questionnaire.
To perform the second annotation, the software interactions were introduced to the participants. Then, they were asked to perform the annotation (respectively without or with assistance). At the end of this second task, participants were asked, once again, to fill-in the questionnaire.
For each annotation task, the participants were left alone in the office. In-

		Video A	Video B
Prediction	Well classified	87.17%	84.78%
	Noise	0%	0%
	Misclassified	12.83%	15.22%
Post-processing	Well classified	80.81%	81.15%
	Noise	17.26%	17.22%
	Misclassified	1.93%	1.63%

Table 5 Repartition of the frames in the three categories after prediction and post-processing

formed consent was obtained from all individual participants at the beginning of the test.

4 Results

4.1 Assistance system

The network reached a mean accuracy of 77.9+/-11.7% on all the database. For both videos A and B, the threshold value T for error detection was set to 15. The window w of the post-processing method achieved the best results with w=50. For videos A and B,respectively 87.17% and 84.78% of the frames were well classified, which means that respectively 12.83% and 15.22% of the frames were misclassified. The frames of videos A and B were respectively divided into 126 and 115 detected phases.
After using the post-processing method, three categories of frames were exhibited: well-classified frames, frames detected as noise and misclassified frames. For video A, 80.81% of the frames were well classified, 19.47% of the frames were left as detected noise and 1.92% of the frames were misclassified. For the video B, 81.15% of the frames were well classified, 17.22% of the frames were detected as noise and 1.63% of the frames were misclassified. The frames of the video A were divided in 11 detected phases among which 6 were detected noise. For the video B, the frames were divided in 12 detected phases and 6 of them were detected-noise.
After the post-processing method, for both videos A and B, the well-classified frames decreased respectively by 6.36% and 3.63%.On the other hand, the misclassified frames decreased respectively by 10.9% and 13.59%. Some of the well-classified frames were detected as noise but most of the misclassified frames were identified.
Prediction performance for both videos are summarized in the table 5 and a display of the result is presented in figure 3 for video A and figure 4 for video B.

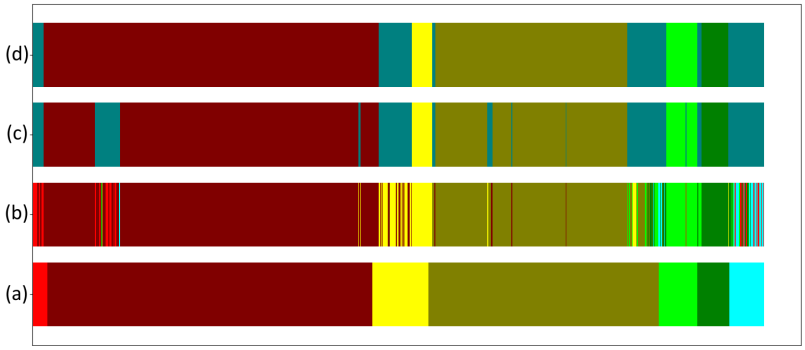

Fig. 3 Frame-by-frame results for video A, each color corresponds to a phase label, the detected errors are in gray. (a) reference (b) prediction, (c) error detector, (d) post-processing

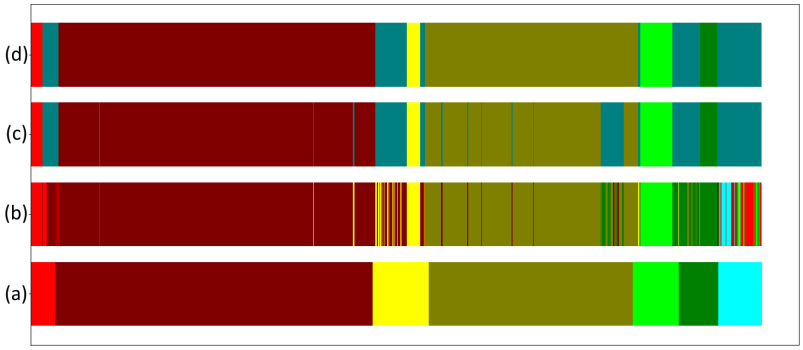

Fig. 4 Frame-by-frame results for video B, each color corresponds to a phase label, the detected errors are in gray. (a) reference, (b) prediction, (c) error detector, (d) post-processing

4.2 User study

4.2.1 Objective measurements

As a within-subject design was used, both conditions were performed by each participant (i.e., repeated measures). Generalized Linear Mixed Model (GLMM) analyses were used for each dependent variable. For each analysis, the participant was considered as a random factor and the independent variable, the condition, as a within-subject factor. The statistical analysis was performed using R software.
The GLMM showed a main effect on accuracy $F_{1,28} = 12.23$, $p = 0.002$, $\eta_p^2 = 0.30$. The percentage of misclassified frame is more important using manual tool ($M = 0.04$, $SD = 0.02$) than using assisted tool ($M = 0.1$, $SD = 0.11$).
The GLMM also showed a main effect on the transition delay $F_{1,28} = 14.49$, $p = 0.0007$, $\eta_p^2 = 0.34$. The annotation is more precise using assisted tool ($M = 36.28$, $SD = 21.49$) than using manual tool ($M = 78.07$, $SD = 60.78$). The figure 5 shows a boxplot on the averaged transition delay resulting from the

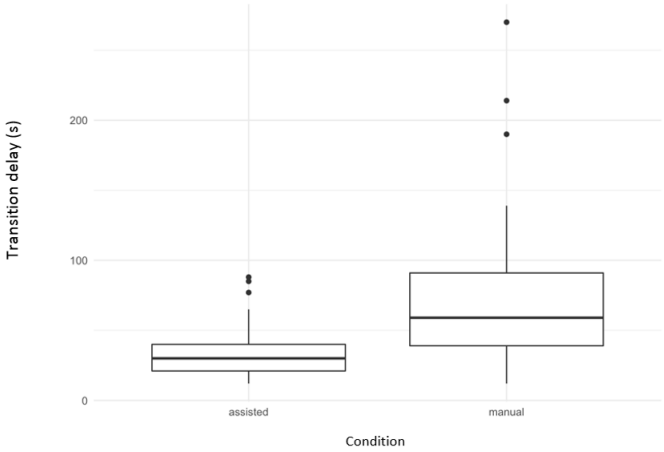

Fig. 5 Transition delay for both manual and assisted annotation

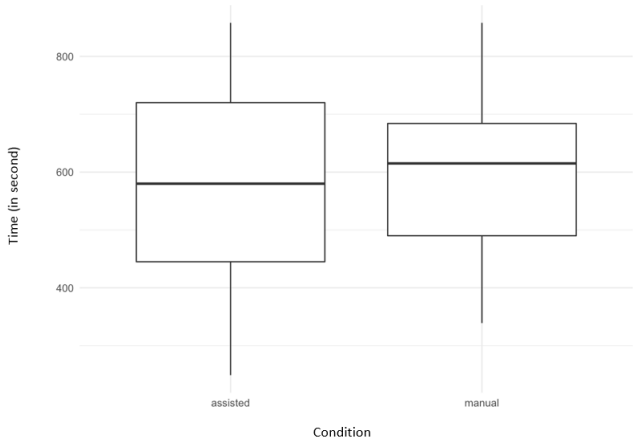

Fig. 6 Completion time for both manual and assisted annotation

annotation with both systems.
Over the 31 participants, 4 phases were completely missed when using the manual system while only 1 phase was missed when performing the annotation with the assistance system.

The GLMM showed no main effect on the completion time $F_{1,28} = 1.08$, $p = 0.31$, $\eta_p^2 = 0.04$. The completion time does not significantly differ between assisted condition ($M = 578.72$, $SD = 157.72$) and manual condition ($M = 605.69$, $SD = 138.17$). The completion time for both tasks is represented in the boxplot 6.

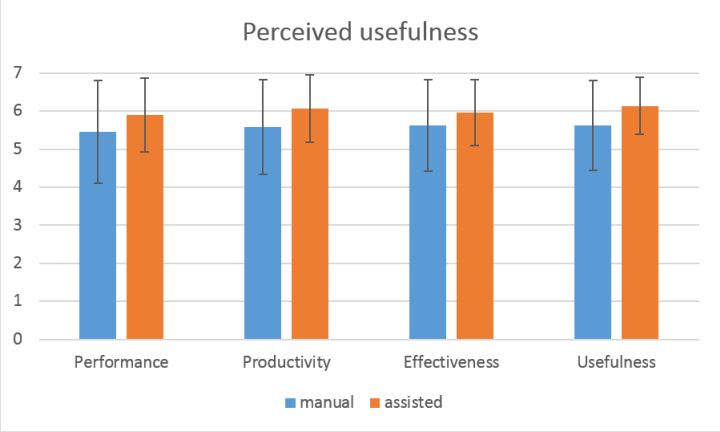

Fig. 7 Evaluation of the perceived usefulness of the two systems

4.2.2 Subjective measurements

Perceived usefulness
After the videos annotation, each participant was asked to assess two subjective dimensions of their interaction with the software: perceived usefulness and perceived ease of use. These theoretical constructs represent key factors to predict behavorial intention (i.e., intention to use the software). First, in terms of perceived usefulness, observable differences (see Figure 7 for the overall ratings of the four items) can be highlighted between the two experimental conditions (i.e., manual annotation vs. assisted annotation). However, no significant difference can be reported ($F_{1,27} = 3.10$, $p = 0.09$, $\eta_p^2 = 0.10$). In parallel, perceived ease of use was evaluated (see Figure 8 for the overall ratings of the four items). No significant difference can be observed between the manual annotation and assisted annotation ($F_{1,27} = 0.18$, $p = 0.68$, $\eta_p^2 = 0.007$).

4.3 Discussion

The protocol used to conduct the user test permitted to validate the hypothesis of improvements of accuracy and transition delay when using the assistance system. However, the second annotation-duration-reduction hypothesis was not verified.
The significant improvement of the annotation accuracy is an important achievement since the training of learning algorithms needs precisely annotated data. Noisy labeled data will be less efficient to train a network. The proposed approach reduced the mean transition delay by 41.79 seconds and increased the accuracy by 5%.
The participants had no time-constraint to perform the annotation task. In

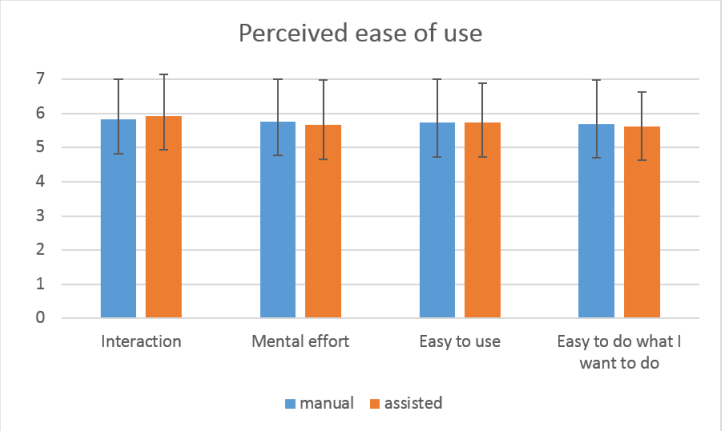

Fig. 8 Evaluation of the perceived ease of use of the two systems

the introduced context, users were asked to imagine themselves as a surgeon recruited to perform annotation on daily basis. The proposed user interface enhanced the annotation process but required some time to master the basics, which can reduced the potential gain in terms of annotation time when used for the first time. This condition can be tested by a repeated use of the software.

The participants were all non-experts. It can be assumed that the difference in term of usefulness could have been better highlighted with a public of experts. With the statistically significant improvement of the annotation accuracy, we can assum that because of the non-expertise of the participants, they did not realize that the assistance system actually helped them.

We can also make the hypothesis that the software and the proposed user interface can only help to improve annotation accuracy which is already a significant improvement. We can expect an improvement in the perceived ease of use with a rework of the ergonomy and/or of the tutorial of the assistance system. Indeed, the software with and without the assistance remains almost the same. The assistance system only consisted in the addition of two buttons that allowed the user to navigate between the predicted transition. All the other interactions remained the same.

It can be assumed that gain of time can be expected for more complex annotation, for lower granularity levels, such as surgical activities.

Three axes of improvements have been identified to enhance our assistance system: the navigation between pre-annotated phases, the highlight of the error and the absence of suggestion to solve the detected noise.

It should be noted that the comparison of the accuracy of the annotation was made using the reference provided by the Cholec80 database. This annotation has been made by only one person and cannot be considered as an absolute ground truth.

4.4 Conclusion

In this paper, we introduced an assistance system based on Deep Learning to annotate the phases of cholecystectomy procedure. The assistance system pre-annotated the videos and highlighted uncertain areas of the annotation, in order to indicate to the user some critical moments in the video. The proposed assistance system was compared with a manual annotation system and evaluated through an objective and subjective measurements. Thirty one volunteers, all non-experts, participated to the study. On average, the assistance system showed significant annotation-accuracy improvements. In future works, inter-user variability of the annotation will be investigated for other surgial procedures.

Compliance with ethical Standards

Funding This study was supported by French state funds maneged by ANR under the reference ANR-10-AIRT-07.

Conflict of interest Gurvan Lecuyer, Martin Ragot, Nicolas Martin, Laurent Launay and Pierre Jannin declare that they have no conflict of interest

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

1. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, Corrado GS, Davis A, Dean J, Devin M, Ghemawat S, Goodfellow I, Harp A, Irving G, Isard M, Jia Y, Jozefowicz R, Kaiser L, Kudlur M, Levenberg J, Mané D, Monga R, Moore S, Murray D, Olah C, Schuster M, Shlens J, Steiner B, Sutskever I, Talwar K, Tucker P, Vanhoucke V, Vasudevan V, Viégas F, Vinyals O, Warden P, Wattenberg M, Wicke M, Yu Y, Zheng X (2015) TensorFlow: Large-scale machine learning on heterogeneous systems. URL https://www.tensorflow.org/, software available from tensorflow.org
2. Alhrishy M, Toth D, Narayan SA, Ma Y, Kurzendorfer T, Rhode K, Mountney P (2018) A machine learning framework for context specific collimation and workflow phase detection
3. Basile V, Bos J, Evang K, Venhuizen N (2012) A platform for collaborative semantic annotation. In: Proceedings of the Demonstrations at the 13th

Conference of the European Chapter of the Association for Computational Linguistics, Association for Computational Linguistics, pp 92–96
4. Bodenstedt S, Wagner M, Katić D, Mietkowski P, Mayer B, Kenngott H, Müller-Stich B, Dillmann R, Speidel S (2017) Unsupervised temporal context learning using convolutional neural networks for laparoscopic workflow analysis. arXiv preprint arXiv:170203684
5. Bortsova G, Dubost F, Ørting S, Katramados I, Hogeweg L, Thomsen L, Wille M, de Bruijne M (2018) Deep learning from label proportions for emphysema quantification. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, pp 768–776
6. Cejuela JM, McQuilton P, Ponting L, Marygold SJ, Stefancsik R, Millburn GH, Rost B (2014) tagtog: interactive and text-mining-assisted annotation of gene mentions in plos full-text articles. Database 2014
7. Chen Y, Sun QL, Zhong K (2018) Semi-supervised spatio-temporal cnn for recognition of surgical workflow. EURASIP Journal on Image and Video Processing 2018(1):76
8. Chollet F. (2015) Keras. https://github.com/fchollet/keras
9. Deng Y, Sun Y, Zhu Y, Zhu M, Yuan K (2018) A strategy of mr brain tissue images' suggestive annotation based on modified u-net. arXiv preprint arXiv:180707510
10. Falk MJ, Shen L, Gonzalez M, Leipzig J, Lott MT, Stassen AP, Diroma MA, Navarro-Gomez D, Yeske P, Bai R, Boles RG, Brilhante V, Ralph D, DaRe JT, Shelton R, Terry SF, Zhang Z, Copeland WC, van Oven M, Prokisch H, Wallace DC, Attimonelli M, Krotoski D, Zuchner S, Gai X. (2015) Mitochondrial disease sequence data resource (mseqdr): a global grass-roots consortium to facilitate deposition, curation, annotation, and integrated analysis of genomic data for the mitochondrial disease clinical and research communities. Molecular genetics and metabolism 114(3):388–396
11. Funke I, Jenke A, Mees ST, Weitz J, Speidel S, Bodenstedt S (2018) Temporal coherence-based self-supervised learning for laparoscopic workflow analysis. arXiv preprint arXiv:180606811
12. Garraud C, Gibaud B, Penet C, Gazuguel G, Dardenne G, Jannin P (2014) An ontology-based software suite for the analysis of surgical process model. In: Proceedings of Surgetica, pp 243–245
13. Gibaud B, Forestier G, Feldmann C, Ferrigno G, Gonçalves P, Haidegger T, Julliard C, Katić D, Kenngott H, Maier-Hein L, März K, de Momi E, Nagy DA, Nakawala H, Neumann J, Neumuth T, Rojas Balderrama J, Speidel S, Wagner M, Jannin P. (2018) Toward a standard ontology of surgical process models. International journal of computer assisted radiology and surgery 13(9):1397–1408
14. Gobbel GT, Garvin J, Reeves R, Cronin RM, Heavirland J, Williams J, Weaver A, Jayaramaraja S, Giuse D, Speroff T, Brown SH, Xu H, Matheny ME. (2014) Assisted annotation of medical free text using raptat. Journal of the American Medical Informatics Association 21(5):833–841

15. Jin Y, Dou Q, Chen H, Yu L, Qin J, Fu CW, Heng PA (2018) Sv-rcnet: Workflow recognition from surgical videos using recurrent convolutional network. IEEE transactions on medical imaging 37(5):1114–1126
16. Koch LM, Rajchl M, Bai W, Baumgartner CF, Tong T, Passerat-Palmbach J, Aljabar P, Rueckert D (2018) Multi-atlas segmentation using partially annotated data: methods and annotation strategies. IEEE transactions on pattern analysis and machine intelligence 40(7):1683–1696
17. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In: Pereira F, Burges CJC, Bottou L, Weinberger KQ (eds) Advances in Neural Information Processing Systems 25, Curran Associates, Inc., pp 1097–1105, URL http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf
18. 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 Transactions on Biomedical Engineering 59(4):966–976
19. Loukas C (2018) Surgical phase recognition of short video shots based on temporal modeling of deep features. arXiv preprint arXiv:180707853
20. Primus MJ, Putzgruber-Adamitsch D, Taschwer M, Münzer B, El-Shabrawi Y, Böszörmenyi L, Schoeffmann K (2018) Frame-based classification of operation phases in cataract surgery videos. In: International Conference on Multimedia Modeling, Springer, pp 241–253
21. Rouzé P, Pavy N, Rombauts S (1999) Genome annotation: which tools do we have for it? Current opinion in plant biology 2(2):90–95
22. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. CoRR abs/1409.1556
23. South BR, Shen S, Leng J, Forbush TB, DuVall SL, Chapman WW (2012) A prototype tool set to support machine-assisted annotation. In: Proceedings of the 2012 Workshop on Biomedical Natural Language Processing, Association for Computational Linguistics, pp 130–139
24. Stenetorp P, Pyysalo S, Topić G, Ohta T, Ananiadou S, Tsujii J (2012) Brat: a web-based tool for nlp-assisted text annotation. In: Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics, Association for Computational Linguistics, pp 102–107
25. Twinanda AP, Shehata S, Mutter D, Marescaux J, De Mathelin M, Padoy N (2017) Endonet: A deep architecture for recognition tasks on laparoscopic videos. IEEE transactions on medical imaging 36(1):86–97
26. Tzanetakis G, Cook PR (2000) Experiments in computer-assisted annotation of audio. Georgia Institute of Technology
27. Venkatesh V, Bala H (2008) Technology acceptance model 3 and a research agenda on interventions. Decision sciences 39(2):273–315
28. Ye C, Coco J, Epishova A, Hajaj C, Bogardus H, Novak L, Denny J, Vorobeychik Y, Lasko T, Malin B, Fabbri D. (2018) A crowdsourcing framework for medical data sets. AMIA Summits on Translational Science Proceedings 2017:273

29. Yengera G, Mutter D, Marescaux J, Padoy N (2018) Less is more: Surgical phase recognition with less annotations through self-supervised pre-training of cnn-lstm networks. arXiv preprint arXiv:180508569
30. Zia A, Hung A, Essa I, Jarc A (2018) Surgical activity recognition in robot-assisted radical prostatectomy using deep learning. arXiv preprint arXiv:180600466