Eric L. Wisotzky*, Lasse Renz-Kiefel, Sophie Beckmann, Sebastian Lünse, René Mantke, Anna Hilsmann, and Peter Eisert

Surgical Phase Recognition for different hospitals

https://doi.org/10.1515/cdbme-2023-1079

Abstract: Surgical phase recognition is an important aspect of surgical workflow analysis, as it allows an automatic analysis of the performance and efficiency of surgical procedures. A big challenge for training a neural network for surgical phase recognition is the availability of training data and the large (visual) variability in procedures of different surgeons. Hence, a network must be able to generalize to new data. In this paper, we present an adaptation of a Temporal Convolutional Network for surgical phase recognition in order to ensure the generalization of the network to new scenes with different conditions on the example of cholecystectomy. We used publicly available datasets of 104 surgeries from four different centers for training. The results showed that the network was able to generalize to new scenes and we obtained recognition results with accuracy up to 82% on our own six captured surgeries, performed in a different hospital. This performance is similar for test data from the hospitals of the training data, suggesting that the network can well generalize to new surgical rooms and surgeons. The findings have important implications for the development of automated surgical decision support systems that can be applied in a variety of real-world surgical settings.

Keywords: Phase Recognition, Surgical Phase, Temporal Convolutional Network, Cholecystectomy, Laparoscopic Surgery

1 Introduction

Surgical phase recognition is an important aspect of surgical workflow analysis, as it allows automatic analysis of the performance and efficiency of surgical procedures. Surgical workflow analysis is a key technology to connect different technological advancements in surgery with the aim of increasing the patients' safety and the surgical efficiency [14].

Surgical phases are high-level tasks that form an entire surgical procedure, e.g., dissecting the Calots' triangle to achieve critical view of safety in laparoscopic cholecystectomy (LC) [6]. With the increasing use of electronic health records and surgical video recordings, there is a wealth of data available that can help to improve patient care and be used to develop automated surgical phase recognition systems [14, 15].

The analysis of laparoscopic operations is attracting increasing interest in the detection of surgical phases, since they represent the standard of care in certain routine operations (e.g., LC) and video collection is easy and can be performed in a more or less standardized manner [3]. Further, laparoscopy is challenging due to the different hand-eye coordination and thus requires a lot of training. Here, an objective assessment of surgical skills and competencies could improve training and assistance outcome.

However, the availability of surgical videos for machine learning (ML) tasks is poor due to time-consuming and laborious annotations [6, 14]. Further, standardized phase definitions are missing, not only for LC. As a result, only few public available data for LC exists. The Cholec80 dataset contains videos of 80 LCs from a single center [13] and has widely been used for phase recognition training [5, 6, 8]. The Heidelberg Colorectal dataset contains 30 videos from three hospitals [14].

The most common network architectures for phase recognition are recurrent neural networks (RNN), temporal convolutional networks (TCN) and transformer models [5, 6, 8]. Most methods use ResNet50 for (spatial) feature extraction and are trained and evaluated on the Cholec80 dataset, which is a small dataset in terms of ML [5, 6, 8, 14]. In addition, no generalisation can be made due to the focus on data from only one or few sources, i.e. one center and/or few surgeons, respectively [1].

Therefore, model adaptation is essential for fast deployment in other medical centers [1]. Adapting models to a new unseen medical center can be done using a finetuning approach, in which the learning process continues on a relatively

***Corresponding author: Eric L. Wisotzky, Sophie Beckmann, Peter Eisert,** Fraunhofer Heinrich-Hertz-Institute HHI & Humboldt-University, Berlin, Germany, e-mail: eric.wisotzky@hhi.fraunhofer.de
Lasse Renz-Kiefel, Anna Hilsmann, Computer Vision & Graphics, Fraunhofer Heinrich-Hertz-Institute HHI, Germany
Sebastian Lünse, Brandenburg Medical School, Department of Surgery, University Hospital Brandenburg, Germany
René Mantke, Faculty of Health Sciences, Joint Faculty of the Brandenburg University of Technology Cottbus - Senftenberg, the Brandenburg Medical School Theodor Fontane and the University of Potsdam & Department of Surgery, University Hospital Brandenburg, Germany

small number of new samples [7]. Still, the model has to train on comparatively small datasets.

In this work, we aim to analyze and demonstrate the feasibility and generalizability of TCNs in real-world surgical settings. We use a standard and adapted TCN approach, both trained on the introduced publicly available data. We applied these networks to data from a completely different hospital. The results of our study could have significant implications for surgical workflow analysis, as well as for the development of automated surgical decision support systems.

2 Methods

2.1 Network

We base our work on the TCN-based surgical phase recognition pipeline TeCNO [4]. The visual features are extracted using a ResNet50 as backbone and refined within a 2-stage TCN. As originally described in [4], the ResNet50 is trained frame-wise without any temporal context. The extracted features of $(T+1)$ frames $(x_{t-T}, \ldots, x_t)$ are then included into the first layer of stage 1 of the TCN to predict the surgical phase at time step t.

Since the temporal progression of a surgery plays an essential role for phase recognition, we integrate the temporal context as early as possible in the pipeline. To do so, we extended the input for the ResNet50 to include two images $[I_t, I_{t-1}]$ at time steps t and $t-1$ instead of the current image I_t at time step t alone. This input is given in the form (u, v, c) with u and v being the width and height of the frame and c holding the color channels of both sequential images $[r_t\ g_t\ b_t\ r_{t-1}\ g_{t-1}\ b_{t-1}]$.

2.2 Dataset and Training

Our aim is to analyze the possibility of training surgical phase recognition methods on few public available datasets and apply these on completely different data, i.e. data from another center and different surgeons. To achieve that, we used the available Cholec80 dataset with 80 LC videos (from one center and 13 surgeons) and the HeiChole dataset with 24 LC videos (from three clinics). Further, we captured six LC videos at the Brandenburg Medical School Theodor Fontane, Germany (MHB). The age of the six patients was 48.5 ± 15.8 yrs, two were female and four male.

We trained TeCNO and our adapted TeCNO with four different dataset options. The different options with their specified amount of videos in training, validation and testing set is presented in Tab. 1. The two options no. 0 are without our MHB data representing the literature-based analysis and are only for comparision. They do not count into the four dataset options.

The selected frame-rate was set to 1 fps. For training, we used the Adam optimizer with an initial learning rate of $5\mathrm{e}{-4}$ and up to 10 epochs for the ResNet50 and an initial learning rate of $7\mathrm{e}{-5}$ and up to 14 epochs for the TCN. The best model was then selected by its performance on the validation set.

Tab. 1: Dataset partitioning between training, validation and test sets. In total three options of dataset partitioning were analyzed. Partitioning No. 0 represents the literature-based analysis without our own MHB data.

No.	Training Data	Validation Data	Test Data
0a.	40 Cholec80	8 Cholec80	32 Cholec80
0b.	40 Cholec80, 12 HeiChole	8 Cholec80, 3 HeiChole	32 Cholec80, 8 HeiChole
1.	64 Cholec80	16 Cholec80	6 MHB
2.	64 Cholec80, 18 HeiChole	16 Cholec80, 6 HeiChole	6 MHB
3.	No. 2 fine-tuned with 2 MHB	1 MHB	3 MHB
4.	64 Cholec80, 18 HeiChole, 2 MHB	16 Cholec80, 6 HeiChole, 1 MHB	3 MHB

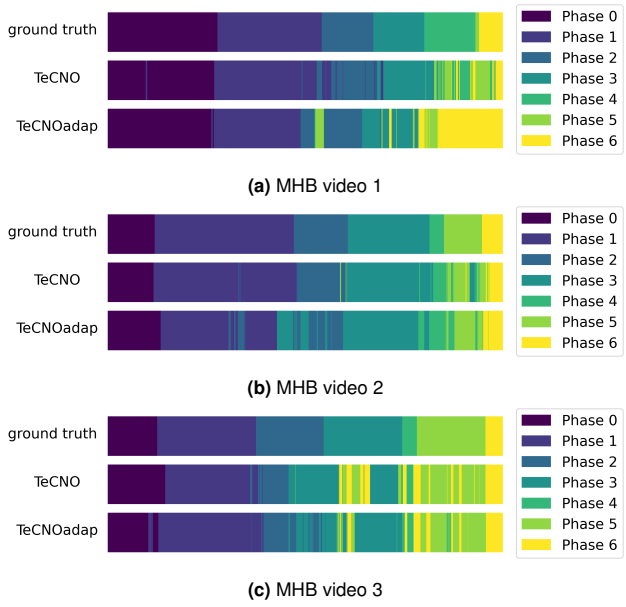
3 Results

The results in terms of accuracy, precision and recall compared to the annotated ground truth is stated in Tab. 2. As for each data partitioning a different number of MHB data was put in the testing set, we performed all training options with all possible MHB data set arrangements and averaged the results. Thus, every MHB video was at least once in the testing set.

Overall, the highest results in terms of accuracy and precision could be achieved by including some of our MHB data into the training set. The fine-tuning option does not lead to much better results than just training without our MHB data. The difference of the overall performance between the initial and the adapted TeCNO model is statistically insignificant. However, we noticed that the network performance varied for individual videos. Beside the increased accuracy and precision of our adaptation compared to initial TeCNO, we noted that phase flickering, i.e., an incorrectly predicted phase for only a single or very few frames, could be reduced. To analyze this, we use phase plots to visualize which frame was predicted to which phase, see Fig. 1.

Tab. 2: The results in terms of accuracy, precision and recall compared to the annotated ground truth data. TeCNO refers to the surgical phase recognition pipeline introduces in [4] and TeCNOadap refers to our adapted network. The used data partitioning is as described in Tab. 1.

Method	Data Partitioning	Acc.	Prec.	Rec.
TeCNO	No. 1	0.670 ± 0.01	0.669 ± 0.03	0.652 ± 0.01
TeCNO	No. 2	0.767 ± 0.01	0.794 ± 0.02	0.680 ± 0.01
TeCNO	No. 3	0.774 ± 0.08	0.764 ± 0.09	0.747 ± 0.11
TeCNO	No. 4	0.804 ± 0.03	0.809 ± 0.04	0.729 ± 0.04
TeCNOadap	No. 2	0.764 ± 0.01	0.794 ± 0.01	0.702 ± 0.02
TeCNOadap	No. 3	0.765 ± 0.09	0.764 ± 0.07	0.721 ± 0.10
TeCNOadap	No. 4	0.791 ± 0.04	0.804 ± 0.04	0.732 ± 0.04



(a) MHB video 1

(b) MHB video 2

(c) MHB video 3

Fig. 1: The phase plots of three MHB videos. For comparison the accuracies of the single results are (a) Acc. TeCNO = 0.81, Acc. TeCNOadap = 0.73. (b) Acc. TeCNO = 0.92, Acc. TeCNOadap = 0.82. (c) Acc. TeCNO = 0.74, Acc. TeCNOadap = 0.83.

The visualization shows that especially phase 0 - 'Preparation' and phase 1 - 'Calot Triangle Dissection' could be recognized with a high accuracy. The following phases were more difficult for both models to predict as more phase flickering is observed. This is true especially for the final stages of the surgery onwards from phase 4 - 'Galbladder Packaging'.

Comparing the performance of both methods on the different videos individually, it can be seen that the performance of the TeCNO model varies quite heavy in between the videos (recognition accuracies from 0.74 to 0.92). The adapted TeCNO model shows lower deviation on the analyzed videos (recognition accuracies 0.73 to 0.83).

4 Discussion and Conclusion

The results of our study suggest that a TCN-based network trained on publicly available data with appropriate variability could adapt to data from a different hospital with different surgeons. This demonstrates the feasibility and generalizability of our approach in real-world surgical settings.

However, it is important to note that only one available dataset seems not sufficient to train a reliable surgical phase recognition system. There is a need for larger datasets with larger variability since there may not be enough variation among different clinics and surgeons in one single dataset to account for the nuances of different surgical procedures. Thus, it can be concluded that more diverse training data is urgently needed to allow sufficient generalization to unseen data. Further, if the diversity of the initial training data is not sufficient, it seems that fine-tuning with only few own clinical data cannot compensate this drawback.

In our study, we used 104 videos from four different medical centers. We find that with increasing variation of the training data the transferability of the models to new unseen data from a different hospital is increased. Moreover, fine-tuning the models on a small amount of unseen data from a different clinic can increase the performance even further. The results are then comparable to the literature [4–6], where the network was evaluated on data from the same origin as the training data. This suggests that, if the training data are not diverse enough, a relatively small amount of data from a different clinic or surgeon may be needed to achieve good results with accuracies up to 82% (vs. accuracy of $81\% - 87\%$ for literature-based analysis of data partitioning No. 0a and 0b). Aiming for a broad clinical application, however, would mean that fine-tuning on own clinical data and or even individual surgeons would not be necessary.

The appearing failures in the phase recognition or occurring discrepancies to the ground truth can have different causes. On the one hand, the individual way in which a standard operation is performed slightly varies from surgeon to

surgeon. On the other hand, the recognition errors are caused by instruments or instrument manufacturers different to those used in the 'foreign' clinic. Further, the prediction errors occur in 'pausing' situations. These faults could be addressed with a more specific definition of the situations, i.e., the introduction of a general rest and out-of-body phase could make the detection more stable and accurate.

In addition, we will investigate deeply the reason of the higher deviation between the individual recognition results for TeCNO in comparision to the adapted TeCNO. This will include a deep comparison of the individual videos where high differences were observed.

Overall, our findings have important implications for the development of automated surgical decision support systems. By using publicly available data from multiple sources and applying our TCN-based network to data from different hospitals, we can develop more robust and reliable surgical phase recognition systems that can be applied in a variety of real-world surgical settings. This would support procedure time prediction [2], surgical management [12] and continuous surgical education [16]. Further, it could be combined with early context-sensitive warnings and objective patient assessments like blood-flow [9, 11], vital sign [10] or hyperspectral tissue analysis [17].

Author Statement

Research funding: The author state no funding involved. Conflict of interest: Authors state no conflict of interest. Informed consent: Informed consent has been obtained from all individuals included in this study. Ethical approval: The research related to human use complies with all the relevant national regulations, institutional policies and was performed in accordance with the tenets of the Helsinki Declaration, and has been approved by the Brandenburg ethics board.

References

[1] Bar O, Neimark D, Zohar M, Hager GD, Girshick R, Fried GM, et al. Impact of data on generalization of AI for surgical intelligence applications. Scientific reports, 2020, 10(1):22208.

[2] Bodenstedt S, Wagner M, Mündermann L, Kenngott H, Müller-Stich B, Breucha M, et al. Prediction of laparoscopic procedure duration using unlabeled, multimodal sensor data. International Journal of Computer Assisted Radiology and Surgery, 2019, 14:1089-1095.

[3] Cheng K, You J, Wu S, Chen Z, Zhou Z, Guan J, et al. Artificial intelligence-based automated laparoscopic cholecystectomy surgical phase recognition and analysis. Surgical endoscopy, 2022, 36(5):3160-3168.

[4] Czempiel T, Paschali M, Keicher M, Simson W, Feussner H, Kim ST, et al. Tecno: Surgical phase recognition with multi-stage temporal convolutional networks. In Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, 2020, pp. 343-352

[5] Czempiel T, Sharghi A, Paschali M, Navab N, Mohareri O. Surgical Workflow Recognition: From Analysis of Challenges to Architectural Study. In Proceedings Computer Vision–ECCV, 2022, pp. 556-568

[6] Garrow CR, Kowalewski KF, Li L, Wagner M, Schmidt MW, Engelhardt S, et al. Machine learning for surgical phase recognition: a systematic review. Annals of surgery, 2021, 273(4):684-693.

[7] Girshick R, Donahue J, Darrell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. Proc IEEE/CVF Conf Comput Vision Pattern Recognit (CVPR), 2014, 580–587.

[8] Hartwig R, Berlet M, Czempiel T, Fuchtmann J, Rückert T, Feussner H, et al. Bildbasierte Unterstützungsmethoden für die zukünftige Anwendung in der Chirurgie. Die Chirurgie, 2022, 93(10):956-965.

[9] Kossack B, Wisotzky EL, Eisert P, Schraven SP, Globke B, Hilsmann A. Perfusion assessment via local remote photoplethysmography (rPPG). Proc IEEE/CVF Conf Comput Vision Pattern Recognit (CVPR), 2022, pp. 2192-2201.

[10] Kossack B, Wisotzky EL, Hilsmann A, Eisert P. Automatic region-based heart rate measurement using remote photoplethysmography. Proc IEEE/CVF Int Conf Comput Vision (CVPR), 2021, pp. 2755-2759.

[11] Schraven SP, Kossack B, Strüder D, Jung M, Skopnik L, Gross J, et al. Continuous intraoperative perfusion monitoring of free microvascular anastomosed fasciocutaneous flaps using remote photoplethysmography. Sci Rep, 2023, 13(1):1532.

[12] Tanzi L, Piazzolla P, Vezzetti E. Intraoperative surgery room management: A deep learning perspective. Int J Med Rob Comput Assisted Surg, 2020, 16(5):1-12.

[13] Twinanda AP, Shehata S, Mutter D, Marescaux J, de Mathelin M, Padoy N. EndoNet: A Deep Architecture for Recognition Tasks on Laparoscopic Videos. IEEE Trans Med Imaging, 2017, 36:86–97. doi: 10.1109/TMI.2016.2593957

[14] Wagner M, Müller-Stich BP, Kisilenko A, Tran D, Heger P, Mündermann L, et al. Comparative validation of machine learning algorithms for surgical workflow and skill analysis with the heichole benchmark. Med Image Anal, 2023, 86:102770.

[15] Wisotzky EL, Rosenthal JC, Eisert P, Hilsmann A, Schmid F, Bauer M, et al. Interactive and multimodal-based augmented reality for remote assistance using a digital surgical microscope. Conf Virtual Reality and 3D User Interfaces (VR). IEEE, 2019, pp. 1477-1484.

[16] Wisotzky EL, Rosenthal JC, Meij S, van den Dobblesteen JJ, Arens P, Hilsmann A, et al. Telepresence for surgical assistance and training using eXtended reality (during and after pandemic periods). J Telemed Telecare, 2023, in publication. doi: 10.1177/1357633X231166226

[17] Wisotzky EL, Uecker FC, Arens P, Dommerich S, Hilsmann A, Eisert P. Intraoperative hyperspectral determination of human tissue properties. J Biomed Opt, 2018, 23(9):091409.