Nour Aldeen Jalal*, Tamer Abdulbaki Alshirbaji and Knut Möller

Evaluating Convolutional Neural Network and Hidden Markov Model for Recognising Surgical Phases in Sigmoid Resection

Abstract: Surgical workflow analysis in laparoscopic surgeries has been studied widely during last years because of its various applications. For example, optimising the schedule of operating rooms (OR) and developing a context-aware system that supports surgical team during the intervention. Surgical phase recognition has been applied to various kinds of laparoscopic procedures, mainly of type cholecystectomy. Sigmoid resection procedures are considered more complex than cholecystectomy, and they have not been extensively studied. Therefore, the focus of this work is to study phase recognition in sigmoid resection. In this paper, a convolutional neural network (CNN) architecture and Hidden Markov Model (HMM) were evaluated for performing phase recognition in sigmoid resection videos. The CNN is an extension of a pre-trained model, and it was fine-tuned to perform the recognition. To consider the temporal aspect of the phase sequences, confidences obtained by the CNN were then provided into a HMM to release final classification. Experimental results show a low performance of the proposed method to recognise surgical phases in such complex procedures. Therefore, the dataset used for the evaluation was also reviewed, and statistics of each phase were generated.

Keywords: Phase recognition, sigmoid resection, convolutional neural network (CNN), Hidden Markov Model (HMM)

https://doi.org/10.1515/cdbme-2018-0099

1 Introduction

In recent years, analysing surgical workflow has gained an increasing interest in the research community. This direction is motivated by many potential applications that share a common goal which is improving and maintaining the quality of patient care. For instance, recognising surgical activities aims to develop a context-aware system that keeps all members of the surgical and anaesthesia team well informed about the surgical workflow, particular hazards, and any special situations that need to be considered. Moreover, modelling surgical workflows helps in optimising the operating room schedule, assessment of surgical skills and training surgeons.

Previous works have employed various features to perform phase recognition. Padoy et al. [1] used binary tool usage signals to carry out the recognition. In recent studies [2,3], laparoscopic videos have been used since they represent a very rich source of information by using image analysis techniques. Therefore, different analysis methods rely on performing the classification based on visual features were proposed. These visual features are either handcrafted [4] or automatically learnt using Neural Networks [3]. Surgical phase recognition has been applied to various kinds of surgeries, but laparoscopic procedures, mainly of type cholecystectomy [2,3], have been preferred. In contrast to cholecystectomy surgeries that were widely studied, laparoscopic sigmoid resection procedures are considered more complicated where the procedure can be performed in different strategies. Therefore, this work aims to evaluate the performance of a Convolutional Neural Network (CNN) with a Hidden Markov Model (HMM) in recognising surgical phases in sigmoid resection procedures.

In this paper, we address the problem of surgical phase recognition in laparoscopic videos. An extension of a CNN model called AlexNet [5] was fine-tuned to perform the classification. Then a HMM was employed to consider the temporal aspect of the procedure, and produce the final results. Finally, the results were discussed, the used dataset was reviewed, and observations were made

***Corresponding author: Nour Aldeen Jalal:** Furtwangen University, Institute of Technical Medicine, Germany, e-mail: ja@hs-furtwangen.de
Tamer Abdulnaki Alshirbaji, Knut Möller: Furtwangen University, Institute of Technical Medicine, Germany, e-mail: {abd, moe}@hs-furtwangen.de

2 Method

The methodology followed to accomplish this study is shown in **Figure 1**. At first, the CNN was fine-tuned using training dataset. Then it was used to perform the phase classification task. The confidences given by the CNN were passed to a Hidden Markov Model (HMM) to release the final results.

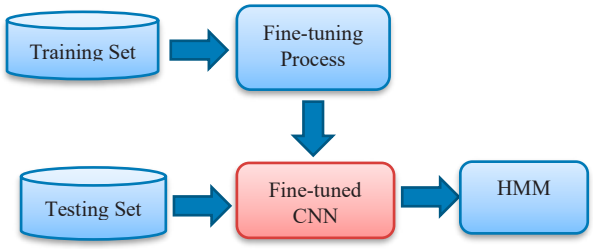

Figure 1: the methodology followed to accomplish this study

2.1 Dataset

The dataset used in this study includes a combination of sigmoid resection videos and device data. This dataset is a part of the dataset submitted for "Endoscopic Vision Challenge 2017 - Workflow Subchallenge" [6] which consists of three sub-datasets for three colorectal surgeries, specifically of type rectal resection, sigmoid resection or proctocolectomy. Sigmoid resection dataset contains eight videos for training and two videos for testing. The dataset was collected at the University Hospital of Heidelberg, and all surgeries are provided with phase annotation. The surgical phases defined in Sigma resection procedure are listed in **Table 1**.

2.2 CNN architecture

The used CNN architecture is an extension of AlexNet model [5]. The CNN architecture is designed by replacing the last fully connected layer fc8 in AlexNet model by a fully connected layer fc_phase to perform the phase classification. This layer contains ten nodes because ten phases are defined in the dataset.

2.3 Fine-Tuning the CNN

The aforementioned CNN architecture was fine-tuned using the previous dataset to adapt the transferred weights of AlexNet layers and train the weights of the new layer. The weights of the new layers were initialised randomly. The network was trained using gradient descent algorithm with softmax multinomial logistic loss function. The loss function is defined as:

$$L_P = \frac{-1}{N_i}\sum_{i=1}^{N_i}\sum_{p=1}^{N_p} i_p^i \log\left(\varphi\left(w_p^i\right)\right)$$

where i and $p \in \{1,\dots,10\}$ are the image and phase indices respectively, i_p^i is the ground truth corresponding to image i, w_p^i is the output of layer fc_phase, and φ is the softmax function.

2.4 Hidden Markov Model

The previous CNN model was designed to predict the surgical phase from the images of the procedure. To enhance recognition capacity, the temporal aspect of the surgical procedure needs to be considered. This was done by providing the phase confidences obtained by the CNN into a Hidden Markov Model (HMM). The surgical phases are the hidden states, and the confidences obtained by the CNN ($v \in R^{10}$) are the observations. Consequently, the established Hidden Markov Model consists of ten states representing ten surgical phases.

The following steps are performed to get a finite set of observations out of the confidences obtained by the CNN:

- At first, values of confidences are divided into ten intervals where this number was chosen empirically. This is required to get discrete values of observations that will be used as an input of the HMM.
- Then a finite set of observations (O) is produced from all possible values of observations in the training set.
- Finally, the observation matrix is computed by calculating the number of all observation's occurrences in each state (surgical phase).

2.5 Implementation

The CNN was trained using MATLAB Neural Network Toolbox (2017a). The pre-trained AlexNet model was loaded from the toolbox. The fine-tuning process was run on a laptop using NVIDIA GEFORCE 840M graphics card.

The HMM was trained using MATLAB to calculate the maximum likelihood of transition probabilities. For testing the Viterbi algorithm was used for offline recognition of the phase sequence.

3 Result

Videos of eight procedures were used for training the CNN model, and the remaining two videos were used for testing. The dataset was downsampled at 1 fps. The training set contains 71785 images while the testing set contains 25532 images. **Table 1** shows surgical phases with their durations in EndoVis dataset. Statistics of training and testing sets are also shown in **Figure 2**.

Table 1: List of phases in EndoVis and their durations (mean ± std)

Phase ID	Phase	Duration (Sec)
1	Preparation and orientation at abdomen	488 ± 456
2	Dissection of lymphnodes and blood vessels	1238 ± 1006
3	Retroperitoneal preparation to lower pancreatic border	916 ± 675
4	Mobilizing the sigmoid and the descending colon	2630 ± 1448
5	Mobilizing the splenic flexure	730 ± 568
6	Mobilizing the transverse colon	549 ± 295
7	Dissection and resection of rectum	978 ± 795
8	Preparing the anastomosis extraabdominally	2159 ± 679
9	Preparing the anastomosis intraabdominally	1294 ± 1080
10	Finishing the operation	424 ± 193

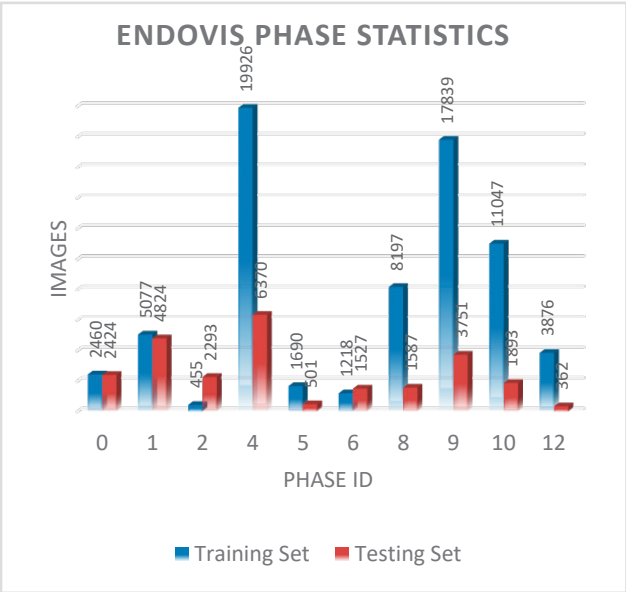

Figure 2: Number of images in the EndoVis dataset for each phase

Figure 3 shows the development of loss function during the training process. The loss starts around 3 and then decreases close to 0.5. After 15K iterations periodic patterns of loss values can be observed; this is because the training set was shuffled only one time at the beginning of the training process.

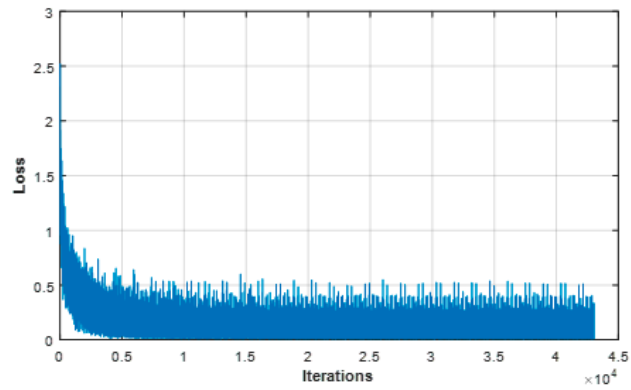

Figure 3: Training loss plotted over training iterations

The results for phase recognition before and after applying HMM in offline mode are shown in **Table 2**. Training loss plotted over training iterations

Table 2: Phase recognition result

Accuracy (%)	Video 9	Video 10
Before applying HMM	38	47
After applying HMM (Offline)	38	51

4 Discussion

In this study, we evaluated using a convolutional neural network (CNN) and a Hidden Markov Model (HMM) to recognise surgical phases in sigmoid resection videos. This method yields moderate classification results in detecting surgical phases in such complex procedures, even though a similar method proposed by Twinanda et al. [3] has proved high recognition performance in cholecystectomy procedures.

It can be noticed that the accuracy is too low for both tested videos before applying the HMM. Even after using the HMM to produce the final classification, the accuracy for video9 did not change, and it was slightly enhanced for video10. To figure out the reasons behind these low classification accuracies, we reviewed the dataset and phase annotation of all videos, and the following observations were made. In fact, Sigmoid resection procedures appear to be too

complicated where they can be performed in different strategies (see **Figure 4**). Therefore, some phases do not appear in some procedures, and the length of each phase is totally different from one operation to another. In addition, some phase transitions occur only in one or two procedures, and this affects training the HMM. Furthermore, the testing data (Video9 & Video10) contains phase transitions that do not appear in the training data. Therefore, these phase transitions are not considered when training the HMM, and the probability of them is zeros. These transitions are shown in **Table 3**.

Table 3: Phase transitions that appear only in testing data

Transition from phase '1' to phase '3'	Occur in Video9
Transition from phase '3' to phase '1'	
Transition from phase '2' to phase '1'	
Transition from phase '1' to phase '6'	Occur in Video10
Transition from phase '6' to phase '1'	

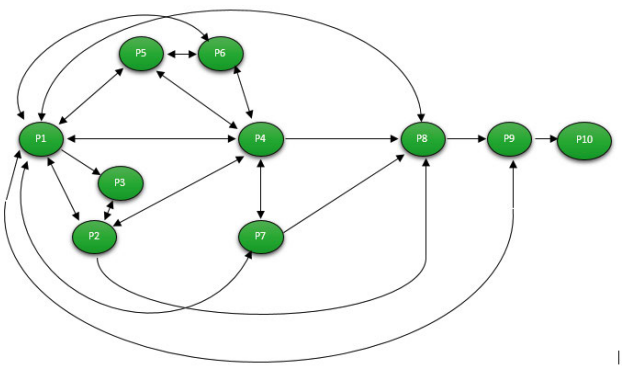

Figure 4: Representation of surgical phases and possible transitions in the dataset

Finally, there might be a mixing between surgical activities and surgical phases in the phase annotation of the dataset. For instance, some images after the extraabdominal anastomosis phase were annotated as belonging to the preparation and orientation phase. However, the surgeon inserts surgical tools into the abdominal cavity to continue preparing anastomosis intraabdominally, but this action should rather belong to the intraabdominally anastomosis phase than the preparation phase.

In summary, the EndoVis dataset contains videos for ten sigmoid resection procedures. This dataset is relatively small and therefore larger dataset should be collected to ensure enough training data. A deeper CNN (i.e. contains larger number of convolutional layers and nodes in the layers) may be utilized to improve classification accuracy, and another kind of HMM can be utilized.

Author's Statement

Research funding: This work was supported by the German Federal Ministry of Research and Education (BMBF under grant CoHMed/IntelliMed grant no. 03FH5I01IA). Conflict of interest: Authors state no conflict of interest. Informed consent: Informed consent is not applicable. Ethical approval: The conducted research is not related to either human or animals use.

References

[1] Padoy, N., Blum, T., Ahmadi, S.A., Feussner, H., Berger, M.O. and Navab, N., 2012. Statistical modeling and recognition of surgical workflow. Medical image analysis, 16(3), pp.632-641.

[2] Dergachyova, O., Bouget, D., Huaulmé, A., Morandi, X. and Jannin, P., 2016. Automatic data-driven real-time segmentation and recognition of surgical workflow. International journal of computer assisted radiology and surgery, 11(6), pp.1081-1089.

[3] Twinanda, A.P., Shehata, S., Mutter, D., Marescaux, J., de Mathelin, M. and Padoy, N., 2017. Endonet: A deep architecture for recognition tasks on laparoscopic videos. IEEE transactions on medical imaging, 36(1), pp.86-97.

[4] Dergachyova, O., Bouget, D., Huaulmé, A., Morandi, X. and Jannin, P., 2016. Automatic data-driven real-time segmentation and recognition of surgical workflow. International journal of computer assisted radiology and surgery, 11(6), pp.1081-1089.

[5] Krizhevsky, A., Sutskever, I. and Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105).

[6] Endoscopic Vision Challenge 2017 - Workflow Subchallenge [https://endovis.grand-challenge.org/].