Current Directions in Biomedical Engineering 2019;5(1):405-408

Nour Aldeen Jalal*, Tamer Abdulbaki Alshirbaji and Knut Möller

Predicting Surgical Phases using CNN-NARX Neural Network

Abstract: Online recognition of surgical phases is essential to develop systems able to effectively conceive the workflow and provide relevant information to surgical staff during surgical procedures. These systems, known as context-aware system (CAS), are designed to assist surgeons, improve scheduling efficiency of operating rooms (ORs) and surgical team and promote a comprehensive perception and awareness of the OR. State-of-the-art studies for recognizing surgical phases have made use of data from different sources such as videos or binary usage signals from surgical tools. In this work, we propose a deep learning pipeline, namely a convolutional neural network (CNN) and a nonlinear autoregressive network with exogenous inputs (NARX), designed to predict surgical phases from laparoscopic videos. A convolutional neural network (CNN) is used to perform the tool classification task by automatically learning visual features from laparoscopic videos. The output of the CNN, which represents binary usage signals of surgical tools, is provided to a NARX neural network that performs a multistep-ahead predictions of surgical phases. Surgical phase prediction performance of the proposed pipeline was evaluated on a dataset of 80 cholecystectomy videos (Cholec80 dataset). Results show that the NARX model provides a good modelling of the temporal dependencies between surgical phases. However, more input signals are needed to improve the recognition accuracy.

Keywords: Surgical phase recognition, convolutional neural network (CNN), NARX neural network, surgical process modelling, cholecystectomy.

https://doi.org/10.1515/cdbme-2019-0102

***Corresponding author: Nour Aldeen Jalal:** Institute of Technical Medicine, Furtwangen University, Jakob-Kienzle-Straße 17, Villingen-Schwenningen, Germany, e-mail: ja@hs-furtwangen.de
Tamer Abdulbaki Alshirbaji (abd@hs-furtwangen.de), **Knut Möller** (moe@hs-furtwangen.de): Institute of Technical Medicine, Furtwangen University, Villingen-Schwenningen, Germany

1 Introduction

With the rising number of medical devices and complexity of technology in operating rooms (OR), intelligent systems are strongly required to be adopted in the surgical environment to compensate the complexity of surgical workflow and streams of data coming from medical devices. The potential applications of such intelligent systems, known as context-aware systems (CASs), are to provide the surgical team with contextual-awareness of the current situation, to support decision-making and to optimize OR management. Modelling surgical procedures is a key motivation of CASs and is an essential step for developing intelligent technologies that can provide assistance to the surgical team in future ORs.

Many works have been done in this area to establish automatic recognition of surgical phases. In this context, state-of-the-art studies have made use of data from different sources inside the OR such as videos [1-3], dedicated sensors [4], or binary tool signals [5,6]. Furthermore, various approaches have been introduced to carry out the recognition of the surgical workflow. Indeed, the majority of previous work focused on using visual features (i.e. video data) or tool binary signals. In visual-based methods, visual features are either handcrafted [1] or automatically learned by convolutional neural networks (CNNs) [2,3]. While tool binary signals are provided into a machine learning method like Hidden Markov Model (HMM) [5,6] or Dynamic Time Warping (DTW) [6] to recognize a particular phase from these signals. However, these binary signals are generally obtained via manual annotation of laparoscopic videos or by installing additional sensors.

In this paper, we present a method that combines a CNN with a nonlinear autoregressive network with exogenous inputs (NARX) to perform the phase recognition task. The CNN model was used to perform tool classification by automatically learning visual features from laparoscopic videos. The tool usage signals generated by the CNN were then provided as input to the NARX neural network for performing a multistep-ahead prediction of the surgical phases of cholecystectomy procedures. The proposed method was finally evaluated on a dataset of 80 cholecystectomy videos.

2 Method

This study relies on formulating the problem of classifying surgical phase as a multistep-ahead prediction task using a dynamic neural network (i.e. NARX neural network). The methodology followed to carry out this work is shown in Figure 1. The first step was selecting an efficient NARX neural network architecture based on several trials and using manual annotations of tool binary signals. Then, a CNN was trained to detect the tool presence in laparoscopic images. Once the CNN was trained, the tool presence probabilities obtained by the CNN were passed to the NARX neural network to predict the surgical phases.

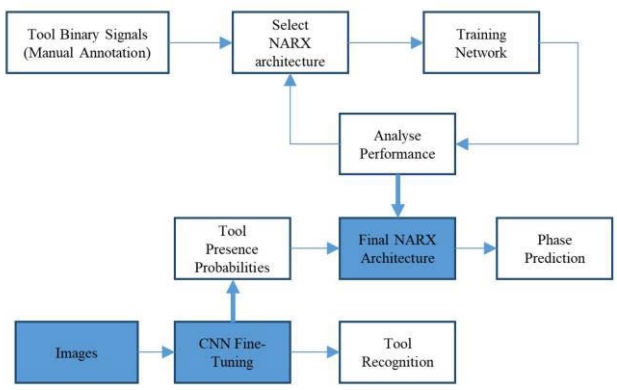

Figure 1: The full pipeline of the proposed method

2.1 Data

A dataset of 80 cholecystectomy procedures (cholec80 dataset [2]) was used for evaluation. This dataset contains videos with corresponding manual annotation of surgical phase and tool presence. Seven phases were defined, and seven surgical tools were used in this dataset.

2.2 NARX Neural Network

2.2.1 Data preparation

The first seventy surgeries were used for training the network while the left ten procedures were used for evaluating the network performance. Input signals were formatted as 7x1 vectors representing the tool binary signals, and targets were represented by the phase number and matched to the corresponding input at each time step. Both input signals and targets were normalized to the range [-1, +1].

2.2.2 Network architecture

A pilot study was conducted to select the optimal network architecture. A two-layer network (i.e. one hidden layer and one output layer) was chosen to perform phase prediction. The hidden layer size was initially set to ten neurons. In addition, the tapped delay lines (TDLs) for both inputs and outputs were set to one.

Several trials were conducted for seeking the optimal number of hidden neurons and TDL lengths by analysing network performance. First, the TDLs were kept to one, and the network was retrained with different number of hidden neurons. By comparing the following metrics: effective number of parameters, sum of squared errors and sum of squared parameters, the number of hidden neurons was finally set to twelve. Second, the network with twelve hidden neurons was retrained against various TDL lengths, and the optimal length was similarly determined by investigating these metrics: autocorrelation, cross-correlation and sum squared errors. Consequently, the NARX network with twelve hidden neurons and three TDL length was selected as a final architecture (see Figure 2).

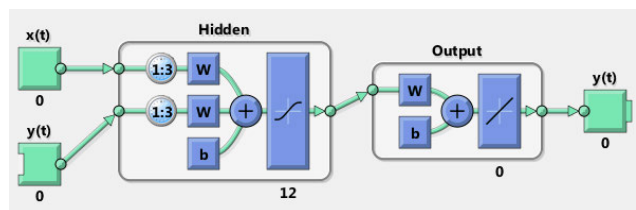

Figure 2: NARX network training architecture

2.2.3 Network training

The NARX neural network was trained using the series-parallel architecture where the true targets (i.e. phases from annotations) are provided into the network instead of feeding back the estimated phases.

The Gauss-Newton approximation to the Bayesian Regularization algorithm [9] was used to train the network. The objective function is represented by:

$$F = \beta E_D + \alpha E_W \tag{1}$$

Where $\boldsymbol{E_D}$ and $\boldsymbol{E_W}$ represent the sum squared error and the sum of squares of the network parameters respectively, and $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ are objective function parameters.

2.3 CNN model

Based on our work [7], the pre-trained model Alexnet [8] was retrained to perform the tool classification task based on the method proposed in [7]. The first 40 videos were used for training, and the left 40 videos were used for testing the CNN.

The trained CNN model was then employed to extract tool presence probabilities from the entire dataset. These probabilities were then provided into the NARX neural network to perform phase prediction.

3 Result

Due to random initialization of network weights, ten networks were trained, and the network with best performance was used to release final results. Tables 1 and 2 show the accuracy and precision of the NARX neural network when binary signals from manual annotation were fed to the network to perform phase prediction.

Table 1: Prediction accuracy (%) of the validation set using tool binary signals obtained by manual annotation

OP	P1	P2	P3	P4	P5	P6	P7
71	100	99.7	52.4	0	8.4	NaN	100
72	42.5	95.2	50.4	87.1	97.4	79.7	100
73	100	99.5	57.6	57.8	31.9	57.1	0
74	100	100	72.0	86.8	74.5	0	0
75	100	98.8	35.7	81.2	52.8	NaN	0
76	59.2	73.9	28.6	100	69.5	90.6	48.3
77	100	99.8	69.0	99.6	72.6	59.2	100
78	100	97.5	43.7	0	31.0	0	0
79	100	97.7	65.6	98.6	69.6	0	56.5
80	100	99.7	68.1	82.1	41.3	0	0

Tables 3 and 4 show accuracies and average precisions of the seven surgical phases defined in the dataset. The phase prediction result of surgery 77 (best video) are shown in Figure 3.

Table 2: Precision of surgical phases, average precision and average accuracy (%) of the validation set using tool binary signals obtained by manual annotation

	Precision							Av. Prec.	Av. Acc.
	P1	P2	P3	P4	P5	P6	P7		
Mean	96.2	82.5	32.9	70.9	58.0	39.8	81.3	65.9	78.9
Std	3.1	16.5	33.2	38.5	41.9	48.8	17.9	16.2	12.3

Table 3: Prediction accuracy (%) of the validation set using tool probabilities obtained by the CNN

OP	P1	P2	P3	P4	P5	P6	P7
71	93.9	100	13.1	73.0	21.5	NaN	100
72	50.7	82.2	86.7	66.7	91.0	91.4	100
73	100	78.8	100	66.8	25.5	0	0
74	91.6	100	0	20.0	87.2	0	0
75	78.1	100	0	70.3	69.5	NaN	0
76	88.8	77.8	0	91.0	100	92.7	100
77	61.2	100	0	60.5	93.1	80.6	100
78	100	95.8	0	0	7.9	0	0
79	100	99.2	50.7	82.3	48.0	0	52.3
80	77.7	100	0	0	0	0	0

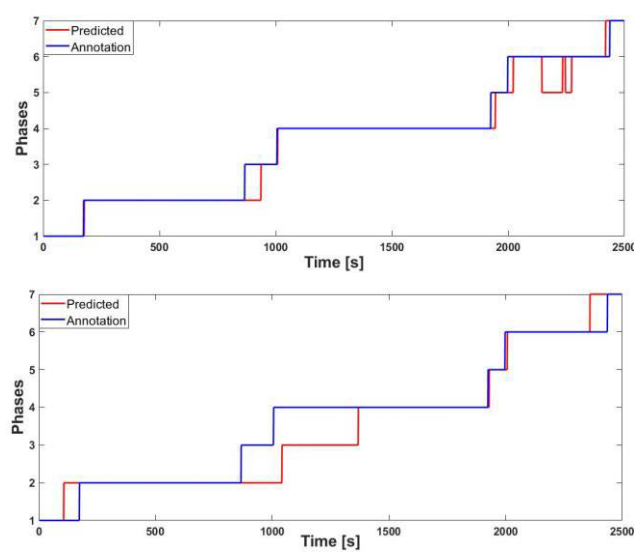

Figure 3: Predicted phases against ground truth of surgery 77. Top figure shows predictions when tool manual annotations were used, bottom figure shows results of the CNN and NARX combination

Table 4: Precision of surgical phases, average precision and average accuracy (%) of the validation set using tool probabilities obtained by the CNN

	Precision							Av. Prec.	Av. Acc.
	P1	P2	P3	P4	P5	P6	P7		
Mean	95.3	82.5	11.7	70.6	40.1	37.2	69.2	58.1	69.3
Std	8.4	14.8	18.7	39.0	39.7	50.3	32.2	29.3	13.1

4 Discussion

In this work, we tested our hypothesis that NARX neural network can learn tool presence patterns in surgical phases and model temporal dependencies between them. Therefore, the NARX neural network was firstly evaluated on tool binary signals obtained by manual annotation of laparoscopic videos. Then, a convolutional neural network was employed to automatically generate tool usage signals from videos and combined with the NARX model.

Prediction accuracies vary for the different surgeries, where they are high for some procedures like surgeries 72 and 77 and low for others like surgery 78 (see Table 1). More precisely, it can be noticed from Table 1 and Table 2 that prediction accuracies of phases 1 and 2 for almost all surgeries are above 95%, and their average precisions are 96% and 82% respectively. In contrast, the NARX seems to have difficulties in predicting the phases 3, 6 and 7. Similarly, the prediction accuracies of phase 5 changes between 8% and 97%, that indicates the NARX is not able to precisely predict this phase for all surgeries.

After combining the NARX with the CNN model, accuracies and precisions for all phases generally decreased. Yet, phases 1 and 2 are still precisely predicted with average precision 95% and 82% relatively. As expected, Prediction accuracy and precision for phase 3 decreased since the CNN model exhibits low ability to identify tools that appear in phase 3 (i.e. scissor and clipper) [7]. Additionally, phases 6 and 7 are still either precisely predicted or not predicted at all. To sum up, average accuracy and precision after combining NARX with CNN are 69% and 58% respectively compared to 79% and 66% using manual annotations.

A limitation of the study was that the CNN-NARX was not evaluated with different splitting of training and testing sets because training the CNN is a time consuming process. In summary, the obtained prediction accuracy shows that using NARX neural network for modelling surgical phases is promising. Nevertheless, tool presence signals alone seem to be not sufficient to obtain perfect performance, and thus additional signals need to be provided to the NARX network.

Author 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] 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.

[2] 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.

[3] Bodenstedt, S., Wagner, M., Katić, D., Mietkowski, P., Mayer, B., Kenngott, H., Müller-Stich, B., Dillmann, R. and Speidel, S., 2017. Unsupervised temporal context learning using convolutional neural networks for laparoscopic workflow analysis. arXiv preprint arXiv:1702.03684.

[4] Neumuth, T. and Meißner, C., 2012. Online recognition of surgical instruments by information fusion. International journal of computer assisted radiology and surgery, 7(2), pp.297-304.

[5] Blum, T., Padoy, N., Feußner, H. and Navab, N., 2008. Modeling and online recognition of surgical phases using hidden markov models. Medical image computing and computer-assisted intervention–MICCAI 2008, pp.627-635.

[6] 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.

[7] Alshirbaji, T.A., Jalal, N.A. and Möller, K., 2018. Surgical Tool Classification in Laparoscopic Videos Using Convolutional Neural Network. Current Directions in Biomedical Engineering, 4(1), pp.407-410.

[8] 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).

[9] Foresee, F.D. and Hagan, M.T., 1997, June. Gauss-Newton approximation to Bayesian learning. In Proceedings of International Conference on Neural Networks (ICNN'97) (Vol. 3, pp. 1930-1935). IEEE