Cross-dataset evaluation of a CNN-based approach for surgical tool detection

T. Abdulbaki Alshirbaji[1,2*], N. A. Jalal[1,2], P. D. Docherty[1,3], T. Neumuth[2] and K. Möller[1]

[1] *Institute of Technical Medicine (ITeM), Furtwangen University, Villingen-Schwenningen, Germany*
[2] *Innovation Centre Computer Assisted Surgery (ICCAS), University of Leipzig, Leipzig, Germany*
[3] *Department of Mechanical Engineering, University of Canterbury, Christchurch, New Zealand*
[*] *Corresponding author, email: tamer.abdulbaki.alshirbaji@hs-furtwangen.de*

Abstract: Surgical tool detection is a key component for analysing surgical workflow and operative activities. The power of deep learning approaches has been widely investigated for processing and recognising the content of laparoscopic images. However, proposed methods, so far, were trained and evaluated using data acquired from a single source. In this work, we evaluate the performance of a convolutional neural network (CNN) model to detect surgical tools in images obtained from different sources. The evaluation results show a drop in the model performance when the evaluation set and training set are not from the same source.

I. Introduction

The evolution of advanced medical technologies has promoted active research to integrate data and extract prerequisite knowledge for cognitive understanding of computer-assisted interventions (CAIs). In light of the progressing research, the future of operating rooms has been envisaged as an advanced cooperative surgical environment [1]. Surgical workflow recognition is an essential step in this direction.

Surgical tool detection is a key component to recognise surgical workflow and analyse surgical activities. Convolutional neural networks (CNNs) were widely employed for detecting surgical tools in laparoscopic images, as the CAIs have eased acquiring video signal of laparoscopic interventions. Twinanda et al. adapted the AlexNet model to learn visual features for recognising surgical tools and phases [2]. However, the imbalanced distribution of tools has a negative impact on training the model. This problem was tackled in [3] using resampling technique and loss-sensitive learning. Some other works proposed to leverage temporal dependencies across neighbouring frames. Chen et al. built a 3D CNN model to learn features across short video clips [4]. CNN and LSTM were employed to encode spatio-temporal information across video clips [5] and the entire video [6, 7].

Although abundant approaches have been proposed for surgical tool detection, performance robustness on different datasets has not been investigated yet. However, a drop in the CNN performance for tool segmentation (different but related task) was reported in [8] when a CNN model tested on data acquired from a different site.

In this work, we evaluate the generalisation ability of a CNN model, namely VGG-16 [9], on images from different datasets for the surgical tool detection task. After training the model on a dataset, the classification performance was evaluated on a different dataset recorded at another hospital and contained a different type of procedure. Another experiment was conducted to investigate if the size of training data can affect detection performance on data obtained from another source.

II. Material and methods

II.I. Datasets

Two datasets recorded at different hospitals were used in this work. Each of the two datasets contains multiple videos of one procedure type. The first dataset is Cholec80 which consists of 80 videos of cholecystectomy procedures. It also includes labels of surgical tools at rate of 1 frame per second (fps). The second dataset, termed Gyna05, consists of 5 videos of gynaecologic procedures. The videos were labelled for surgical tools at a rate of 1 fps. The surgical tools differ in the datasets. Nevertheless, four surgical tools appear in both datasets. These tools, termed as target tools, are grasper, scissors, irrigator and bag.

II.II. CNN-model

The VGG-16 model was initially pre-trained on ImageNet dataset. Then, it was adapted to classify target tools. The last layer of the model was replaced by another fully-connected layer with 4 nodes. This modification in the model architecture is similar to Twinanda adaptation of AlexNet model [2]. Datasets have imbalanced distribution of surgical tools. To reduce the effect of imbalanced data on model training, loss-sensitive learning approach was employed as in [3]. Losses of the target tools were weighted as following:

$$w_t = \frac{Class_{majority}}{Class_t} \quad (1)$$

where w_t is weight for tool t, $Class_{majority}$ is the number of majority class samples and $Class_t$ is the number of tool

t samples. The loss of tool *t* was computed using the binary cross-entropy function.

Two experiments were conducted in this study. In the first experiment, 40 videos of Cholec80 were used for training the model. In the second experiment, more data, namely 75 videos of Cholec80, were used for training. Gyna05 dataset was used as an evaluation set in both experiments. For comparison purpose, the model was also evaluated on the remaining data of Cholec80 in each experiment.

III. Results and discussion

This work evaluates the performance of a CNN model to classify surgical tools in datasets containing procedures of different types and recorded at different surgical sites. The VGG-16 model was used as an example case in this study.

The experimental results show good classification performance on images from the same dataset used for training the model. On the other hand, classification performance, as shown in Fig. 1, dropped when the model evaluated on a different dataset. Increasing the size of training data did not improve the classification performance on Gyna05 dataset, as shown in Fig. 2.

Nevertheless, the model has a high classification performance for grasper in Cholec80 and Gyna05. This indicates that the model is well trained and generalised for the grasper. This is due to the fact that grasper is used in almost the entire procedure and therefore it appears in the majority of video frames. On the other hand, the scissors, which are used for a short time to perform specific surgical activities, appear in a few frames. Thus, the training data was not enough to learn general features for the scissors, as it is failed to be classified in Gyna05 dataset. Although detection performance for the scissors is improved using the loss-sensitive approach on Cholec80 [3], this approach did not help with the generalisation issues of the model.

The irrigators used in both datasets had a similar visual appearance. However, one of the trocars that appeared in Gyna05, has a similar appearance to the irrigator. For this reason, the model misclassified images showing this trocar as an irrigator.

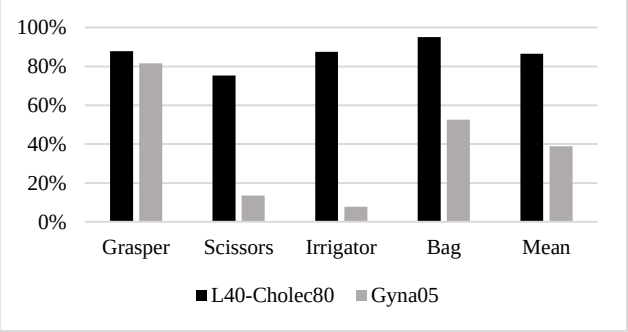

Figure 1: Results of the 1st experiment. It shows average precision (AP) of the target tools on the last 40 videos of Cholec80 and on Gyna05 dataset.

To develop any intra-operative system based on surgical tool detection, it is a necessity to identify the tools accurately in different surgical sites and types. Therefore, more investigations are required to discover possible directions to improve generalization across different datasets. Future work would examine training the model on a mixture of data acquired from many sources and explore the impact on model generalisation.

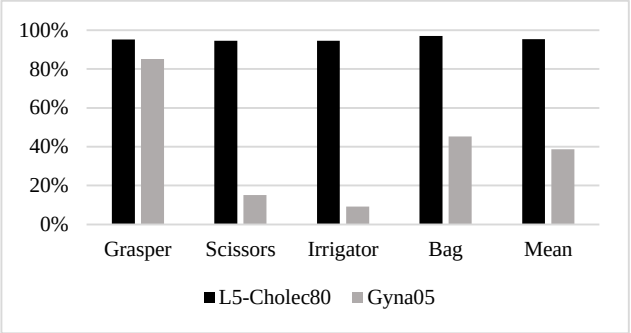

Figure 2: Result of the 2nd experiment. It shows AP of the target tools on the last 5 videos of Cholec80 and on Gyna05 dataset.

IV. Conclusions

In this paper, we evaluate the performance of a CNN model to detect surgical tools in images acquired from different sources. The model was not able to generalise for under-presented tools e.g. scissors, despite the good generalisation capability achieved for over-presented tools.

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. 13FH5I01IA and 13FH5I05IA). Conflict of interest: Authors state no conflict of interest. Informed consent: Informed consent is not applicable. Ethical approval: The research is not related to either human or animals use.

References

[1] Stauder, R., Ostler D., Vogel T., Wilhelm D., Koller S., Kranzfelder M., Navab N.: Surgical data processing for smart intraoperative assistance systems. Innovative surgical sciences 2 (3), 145-152 (2017).

[2] Twinanda, A. P., 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 36 (1), 86-97 (2016).

[3] Alshirbaji, T. A., Jalal N. A., Möller K.: Surgical tool classification in laparoscopic videos using convolutional neural network. Current Directions in Biomedical Engineering 4 (1), 407-410 (2018).

[4] Chen, W., Feng J., Lu J., Zhou J.: Endo3d: online workflow analysis for endoscopic surgeries based on 3d cnn and lstm. In: OR 2.0 Context-Aware operating theaters, computer assisted robotic endoscopy, clinical image-based procedures, and skin image analysis. Springer, pp 97-107 (2018).

[5] Jalal, N. A., Alshirbaji T. A., Docherty P. D., Neumuth T., Möller K.: Surgical Tool Detection in Laparoscopic Videos by Modeling Temporal Dependencies Between Adjacent Frames. In: European Medical and Biological Engineering Conference, pp. 1045-1052, Springer (2020).

[6] Mishra, K., Sathish R., Sheet D.: Learning latent temporal connectionism of deep residual visual abstractions for identifying surgical tools in laparoscopy procedures. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 58-65, (2017).

[7] Abdulbaki Alshirbaji, T., Jalal N. A., Möller K.: A convolutional neural network with a two-stage LSTM model for tool presence detection in laparoscopic videos. Current Directions in Biomedical Engineering 6 (1), (2020).

[8] Ross, T., Zimmerer D., Vemuri A., Isensee F., Wiesenfarth M., Bodenstedt S., Both F., Kessler P., Wagner M., Müller B.: Exploiting the potential of unlabeled endoscopic video data with self-supervised learning. International journal of computer assisted radiology and surgery 13 (6), 925-933 (2018).

[9] Simonyan, K., Zisserman A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:14091556, (2014).