Proceedings on Automation in Medical Engineering
Proc AUTOMED, Vol 2, No 1 (2023): Article Ref 757
www.journals.infinite-science.de/automed/article/view/757

Infinite Science Publishing

Fusing CNN features to improve generalisability for surgical tool classification

T. Abdulbaki Alshirbaji[1,2*], N. A. Jalal[1,2], 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*
[*] *Corresponding author, email: tamer.abdulbaki.alshirbaji@hs-furtwangen.de*

Abstract: Surgical tool classification is a fundamental component for understanding surgical workflow of laparoscopic interventions. Image-based approaches using convolutional neural networks (CNN) have been prominent with availability of computing infrastructures and achieved high performance. However, such approaches need to be assessed in terms of robustness and generalisability to new data sources. Previous works have revealed low generalisation performance of CNN base models. This work proposes a method to enhance CNN generalisability by fusing features from multiple intermediate layers. Experimental results showed good improvement in generalisation performance on data obtained from new clinics and unseen types of procedures.

I. Introduction

Surgical workflow analysis has been an active research area as it is a major key to transfer current operating theatre to the next advanced generation. Cognitive understanding of surgical processes, which is an aim of analysing surgical workflow, enables developing cooperative assistant surgical systems. Such systems have the potential to assist surgical team to improve surgical outcome and patient care.

Identifying surgical tools being in use at every time step is fundamental for recognising performed surgical activities and, therefore, understanding the surgical workflow. Several techniques have been introduced for detecting surgical tools. Early techniques were based on installing radio-frequency identification (RFID) tags on surgical tools. Tool usage signals were acquired from special sensors installed on trocars [1]. Recent approaches were based on analysing laparoscopic videos. Such approaches are preferable since laparoscopic videos are acquired without any need to install additional devices that might interfere human operators inside the operating theatre.

Initially, hand-crafted visual features, such as colour [2] and gradient [3] features, have been used for classifying surgical tools in laparoscopic images. With the success of deep learning approaches in computer vision tasks, convolutional neural networks (CNN) have been proposed for analysing the content of laparoscopic images. Twinanda et al. proposed a CNN architecture called EndoNet for recognising surgical tools and surgical phases in cholecystectomy images [4]. In other studies, CNN performance was improved by employing loss-sensitive learning [5] and by incorporating temporal dependencies [6].

The reported results of CNN-based approaches reflect high capability of CNN to classify surgical tools in laparoscopic images. However, model training and evaluation was conducted on images from a single dataset. Therefore, generalisation capability of CNN model to images from different datasets need to be assessed. Abdulbaki Alshirbaji et al. performed cross-data evaluation of CNN models for surgical tool classification [7,8]. The datasets were recorded at different clinics. The results showed a performance drop on data from unseen sources [7].

This work aims to evaluate generalisation capability of a base-CNN model and propose a method to improve CNN generalisability to new data source. Following the concepts presented in [9], features learnt by CNN at different stages (shallow and deep layers) were combined with the model features (at the last layer) to boost model generalisability. In this study, the architecture of ResNet-50 model was modified accordingly and denoted as ResNet50-Fused. The generalisation performance of the base and the modified model was evaluated on three datasets.

II. Material and methods

II.I. Datasets

Three datasets were used in this work: Cholec80 [4], EndovisChole [10] and Gyna16. Cholec80 contains 80 cholecystectomy videos acquired from one clinic in France. Seven surgical tools were used in Cholec80 dataset. EndovisChole dataset has 33 videos of cholecystectomy procedures recoded at three clinics in Germany. There are 21 surgical tools used in these procedures. Gyna16 dataset was recorded in another clinic and was labelled for surgical tools. The procedures of Gyna16 are of type gynaecology and includes 26 surgical tools.

II.I. CNN model

The architecture of ResNet-50 model was modulated. The features of the third, seventh and thirteenth layer were fused with the features of the last layer after passing through rectified linear unit (ReLU) and a global average pooling (GAP) layers (see Figure 1). Figure 1 shows the modified architecture (ResNet50-Fused).

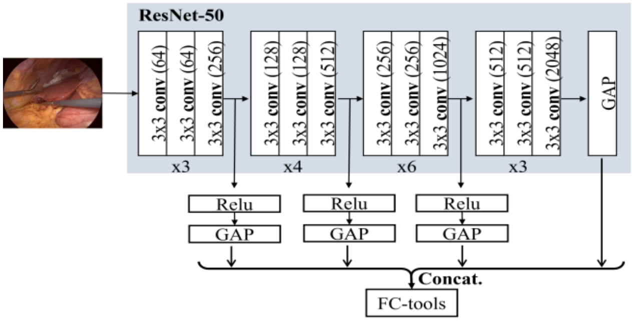

Figure 1: Proposed multi-layer fused feature (ResNet50-Fused).

ResNet50 and ResNet50-Fused were trained using 40 videos of Cholec80 dataset. The remaining data were used for evaluation. Additionally, EndovisChole and Gyna16 were used for evaluating generalisation capability of the CNN model. The surgical tools which have similar visual appearance across the datasets were only considered in this study. The binary cross-entropy function was used for computing the tool loss. The tool losses were weighted according to the distribution of tool samples in the training data to reduce the effect of imbalanced data on model training.

III. Results and discussion

This work presents an approach to improve CNN generalisability by combining features of intermediate layers with the last layer features. The performance of ResNet-50 and ResNet50-Fused was evaluated on three different datasets for surgical tool classification.

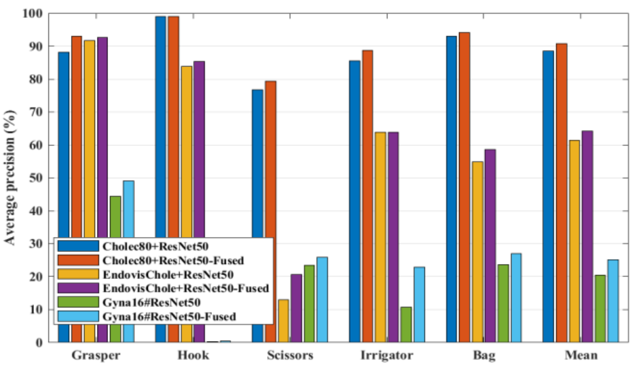

Figure 2: Average precision of ResNet-50 and ResNet-50-Fused on Cholec80, EndovisChole and Gyna16 datasets.

ResNet-50 achieved a mean average precision (mAP) of 88.54% on data from Cholec80 dataset (see Figure 2). However, experimental results revealed a limited generalisation capability of ResNet-50 to other datasets (EndovisChole and Gyna16). The mAP dropped by 27.1% on EndovisChole dataset. Testing the model on a different type of procedures yielded lower performance. By fusing CNN features at multiple stages, the performance enhanced by 2.79% and 4.56% on EndovisChole and Gyna16, respectively (see Figure 2).

Surgical tools rarely used like scissors and bag have lower generalisation performance than other dominant tools like grasper and hook. ResNet50-Fused improved the performance particularly for scissors, irrigator and bag. However, ResNet50 and ResNet50-Fused failed to detect hook in unseen datasets (EndovisChole and Gyna16).

IV. Conclusions

This work evaluates the feasibility of fusing CNN features at multiple layers to improve generalisability to data from unseen sources. The proposed method showed improvement particularly for detecting under-presented surgical tools.

Author's statement

Research funding: This work was supported by the German Federal Ministry of Research and Education (BMBF under grant CoHMed/IntelliMed 13FH5I05IA and grant CoHMed/PersonaMed B-3FH5I09IA). 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 study was conducted in accordance with the Declaration of Helsinki, and approved by the Ethics Committee of Furtwangen University (application Nr. 19-0306LEKHFU, 1 July 2019).

References

[1] Miyawaki F, Tsunoi T, Namiki H, Yaginuma T, et al. Development of automatic acquisition system of surgical-instrument informantion in endoscopic and laparoscopic surgey. In: 2009 4th IEEE conference on industrial electronics and applications. IEEE; 2009. p. 3058–63.

[2] Bouget D, Benenson R, Omran M, Riffaud L, Schiele B, Jannin P. Detecting surgical tools by modelling local appearance and global shape. IEEE transactions on medical imaging. 2015;34(12):2603–17.

[3] Haase S, Wasza J, Kilgus T, Hornegger J. Laparoscopic instrument localization using a 3-d time-of-flight/rgb endoscope. In: 2013 IEEE Workshop on Applications of Computer Vision (WACV). IEEE; 2013. p. 449–54.

[4] Twinanda AP, Shehata S, Mutter D, Marescaux J, De Mathelin M, Padoy N. EndoNet: a deep architecture for recognition tasks on laparoscopic videos. IEEE transactions on medical imaging. 2016;36(1):86–97.

[5] Alshirbaji TA, Jalal NA, Möller K. Surgical tool classification in laparoscopic videos using convolutional neural network. Current Directions in Biomedical Engineering. 2018;4(1):407–10.

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

[7] Abdulbaki Alshirbaji T, Jalal NA, Docherty PD, Neumuth T, Möller K. Robustness of Convolutional Neural Networks for Surgical Tool Classification in Laparoscopic Videos from Multiple Sources and of Multiple Types: A Systematic Evaluation. electronics. 2022;11(18):2849.

[8] Abdulbaki Alshirbaji T, Jalal NA, Docherty PD, Neumuth T, Möller K. Cross-dataset evaluation of a CNN-based approach for surgical tool detection. In: AUTOMED 2021-Automatisierung in der Medizintechnik, 15 Interdisziplinäres Symposium, 08-09 Juni 2021, Basel, Switzerland. 2021.

[9] Alshirbaji TA, Jalal NA, Docherty PD, Neumuth PT, Möller K. Improving the Generalisability of Deep CNNs by Combining Multi-stage Features for Surgical Tool Classification. In: 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE; 2022. p. 533–6.

[10] Wagner M, Müller-Stich BP, Kisilenko A, Tran D, et al. Comparative validation of machine learning algorithms for surgical workflow and skill analysis with the HeiChole benchmark. arXiv preprint arXiv:210914956. 2021;