Tamer Abdulbaki Alshirbaji*, Nour Aldeen Jalal, Paul D. Docherty, Thomas Neumuth and Knut Moeller

Neural Network Classification of Surgical Tools in Gynecological Videos

https://doi.org/10.1515/cdbme-2022-1164

Abstract: Automated surgical tool classification will improve the workflow of surgery. Previous research tackled this task mainly in cholecystectomy procedures due to availability of a relatively large and labelled set (Cholec80 dataset). However, the complexity of the procedure type has an impact on the robustness of the deep learning approaches. Therefore, the classification capability of CNNs on data of more complex procedures with many surgical tools was investigated. In this work, laparoscopic videos of 14 gynaecological procedures were recorded and labelled for surgical tool presence. Then, the DenseNet-121 model was trained to identify surgical tools according to functionality. Experimental results imply high classification performance for some surgical tools. The mean average precision over all the tools was 67%. This study is an initial benchmark for detecting surgical tools in realistic settings.

Keywords: Convolutional neural network (CNN), surgical tool classification, laparoscopic videos, gynaecology.

***Corresponding author: Tamer Abdulbaki Alshirbaji:** Institute of Technical Medicine (ITeM), Furtwangen University, Jakob-Kienzle-Strasse 17, 78054 Villingen-Schwenningen, Germany, and Innovation Centre Computer Assisted Surgery (ICCAS), University of Leipzig, Leipzig, Germany, e-mail: abd@hs-furtwangen.de
Nour Aldeen Jalal: Institute of Technical Medicine (ITeM), Furtwangen University, Villingen-Schwenningen, Germany, and Innovation Centre Computer Assisted Surgery (ICCAS), University of Leipzig, Leipzig, Germany.
Paul D. Docherty: Department of Mechanical Engineering, University of Canterbury, Christchurch, New Zealand, and Institute of Technical Medicine (ITeM), Furtwangen University, Villingen-Schwenningen, Germany.
Thomas Neumuth: Innovation Centre Computer Assisted Surgery (ICCAS), University of Leipzig, Leipzig, Germany.
Knut Moeller: Institute of Technical Medicine (ITeM), Furtwangen University, Villingen-Schwenningen, Germany.

1 Introduction

Surgical data science (SDS) involves recording, organising and analysing data inside the operating theatre to develop models that describe the status of the surgery [1]. The main motivation of SDS is improving patient safety and outcome [1]. However, accessing data from medical devices inside the operating room (OR) remains a challenging task. This difficulty has led to a lack of available labelled datasets. In this respect, laparoscopic videos represent a rich source of data. Therefore, analysing laparoscopic videos is necessary to develop context-aware systems (CASs) that are based on detecting surgical phases [2] and/or the surgical tools used.

Surgical tool detection in laparoscopic videos has been extensively studied during last years. Recently, convolutional neural network (CNNs) approaches have become dominant in the surgical tool recognition task by providing superior performance compared to traditional machine learning approaches. However, purely CNN-based approaches have shown some limitations. For instance, training CNN models with imbalanced data, a ubiquitous characteristic of laparoscopic video datasets, will exhibit bias towards precise classification of the more frequently observed classes [3]. Additionally, smoke, emerged due to electrocauterization [4,5], and blood can cover some parts of surgical tools. Thus, detect surgical tools in such images is challenging for CNN as it learns only spatial features without considering temporal dependencies along the laparoscopic video [6]. To overcome the previous obstacles, several approaches have been introduced. Abdulbaki Alshirbaji et al. employed loss-sensitive learning and resampling techniques to reduce the impact of imbalanced distribution of data on training process [3]. The temporal information encoded in the laparoscopic video was exploited using different modalities such as hidden Markov model [7], recurrent neural networks [8,9] and transformer module [10].

The Cholec80 dataset is one of only a few labelled datasets that are publicly available [7]. Therefore, methods proposed to date have generally been evaluated using the same

dataset for training and testing [11,12]. Additionally, the complexity (i.e. number of surgical phases and tools) of the procedure type has an impact on the robustness of the deep learning approaches. Jalal et al. showed a large drop in performance of a CNN-based method, that achieved high performance on cholecystectomy data, when evaluated on sigmoid resection surgeries [13].

Similar to [13], the performance of a CNN base model was evaluated on performing surgical tool classification in laparoscopic gynaecology procedures. To this end, videos of 14 gynaecology procedures were recorded and prepared to carry out this study. The performance of DenseNet-121 model was evaluated on the gynaecological data after modifying the model architecture for a tool classification task.

2 Method

2.1 Dataset

Data acquisition framework was implemented to record data of laparoscopic procedures in an integrated operating room (OR1, KARL STORZ) at the Schwarzwald-Baar clinic in Villingen-Schwenningen, Germany [14]. Fourteen gynaecological videos were recorded at 25 Hz and a resolution of 1920×1080. The procedures had different execution lengths, with median duration of 81.5 minutes (min: 23.0, max 150.9).

The videos were labelled for surgical tool presence at 1 Hz. Surgical tools were used in different sequences, and at different rates across the recorded procedures. In total, 28 fine surgical tool-classes were observed in the videos. Images captured when the camera was outside the endoscope were replaced by white images for anonymisation. The white images were excluded (~5% of data). Additionally, images containing surgical tools which were used in a single procedure or for a very short period (≤ 3 minutes) were excluded (3.7% of data).

The 28 surgical tools observed were grouped into 9 main classes according to functionality. These groups were grasper, bipolar, irrigator, scissors, sling, rotary blade, needle, mesh and bag. Figure 1 presents the main classes of surgical tools and quantity of images in each class.

2.2 Model architecture

In this work, the DenseNet-121 model [15] was employed to identify surgical tools in gynaecological images. DenseNet-121 was selected due to the high performance achieved on cholecystectomy data for the same task [8]. The model is composed of four convolutional blocks. The model exhibits dense connections to enhance information flow through the layers [15]. The last layer of the model was replaced with another fully-connected layer (Fc-tool). Fc-tool has 9 nodes compatible with the defined main tool-classes. A drop-out layer was added before the Fc-tool layer with a drop rate of 30%.

2.3 Training setup

The training set was composed of the labelled images of 10 videos. The remaining 4 videos were used to evaluate the model performance. The distribution frames containing each tool in the training and testing sets, respectively, is presented in Figure 1.

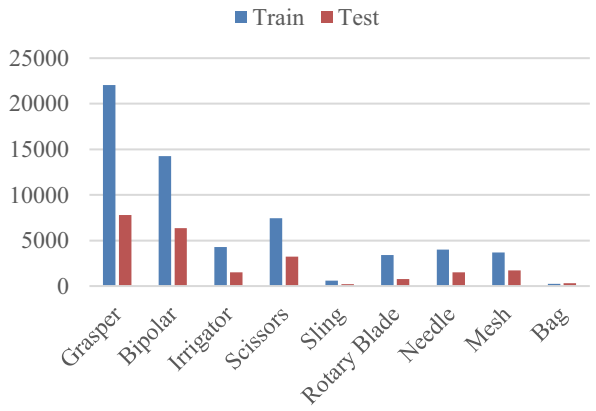

Figure 1: Data distribution for main classes of surgical tools.

ImageNet pretraining weights were used for model initialisation. The activation function for the Fc-tool layer was a sigmoid function as the layer performs multi-object binary classification. The model was trained with an initial learning rate of 2×10^{-4} and a decay of 27×10^{-4}. Training was conducted using a batch size of 40 images and the Adam optimiser.

Each surgical tool had a different number of images and therefore, the distribution of the data was imbalanced. When this problem is not addressed, it leads to bias towards majority classes. To alleviate this effect, the losses of the surgical tools were weighted according to the inverse of the number of frames the corresponding tool appears in the training set [3]. The binary cross-entropy function was employed to compute the losses [8].

This work was conducted using the Keras framework and a graphics processing unit (GPU) on a PC with Intel Xeon 2.20 GHz CPU. The GPU was NVIDIA GeForce RTX 2080Ti. The

training time was approximately 84 minutes per epoch. The inference time for complete testing set took ~30.6 minutes.

3 Results

The average precision (AP) metric was used to evaluate the classification performance of the model. AP represents the

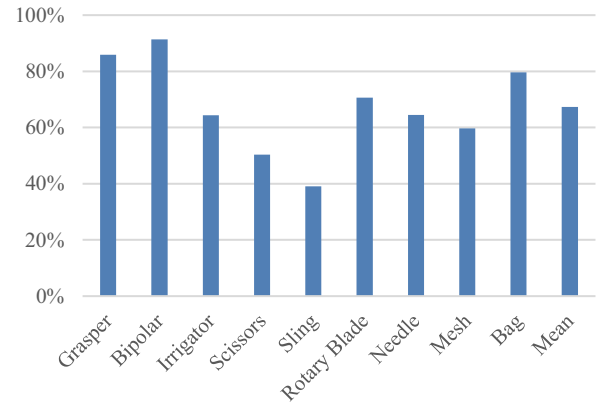

Figure 2: Average precision of classification results for the surgical tools.

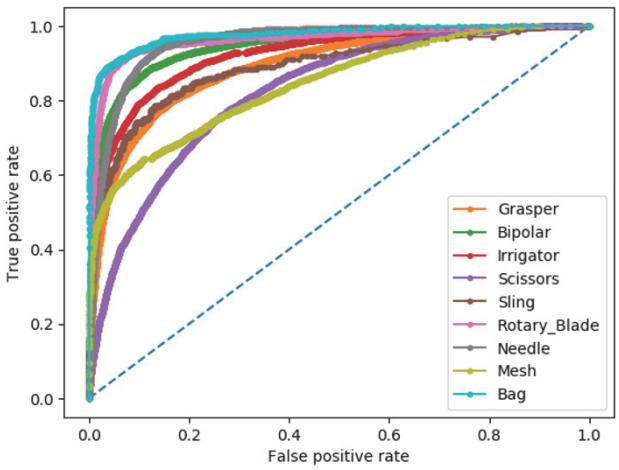

Figure 3: ROC curve for the main classes of surgical tools.

area under the precision-recall curve. Figure 2 shows AP for the defined surgical tool-classes and the mean average precision (mAP) for the surgical tools. Figure 3 presents receiver operating characteristics (ROC) curves that illustrate the performance at different classification thresholds.

To analyse classification results, misclassified instances were examined. Figure 4 shows two images from the testing set with visualisation of gradient weighted class activation (Grad-CAM) [16] and the prediction probability for (a) class grasper and (b) class scissors.

4 Discussion

In this work, we addressed the problem of detecting the presence of surgical tools in gynaecological images. Initially, laparoscopic videos were recorded and labelled for surgical tool presence. The DenseNet-121 model was adapted to perform surgical tool classification.

The CNN model achieved moderate classification performance for the surgical tools, except for sling with an AP of 39% (see Figure 2). The grasper and bipolar had the highest AP of 86% and 91%, respectively. While there were a low number of bag samples in the training set (about 1% of training data), the model was able to distinguish bag with AP of ~77%. Weighting losses of the surgical tools helped to compensate biasing effect of imbalanced training data. However, classification performance for other low-represented tools like sling and mesh were less improved.

The DenseNet-121 model had lower classification performance on gynaecological data compared with cholecystectomy data. The mAP over the main tool classes is 67% (see Figure 2), while DenseNet-121 model reached a mAP of 92% over seven surgical tools on cholecystectomy images from Cholec80 dataset, as reported in [8]. In the Cholec80 dataset, the same surgical tools were utilised in all

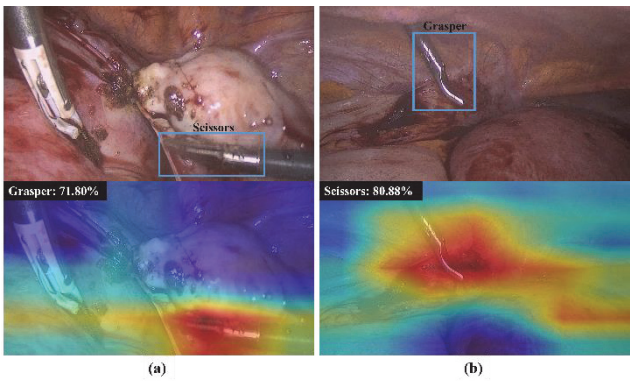

Figure 4: Two examples of model misclassification. (a) test image containing scissors and activation map with prediction probability of grasper class. (b)test image containing grasper and activation map with prediction probability of the scissors.

procedures. In contrast, various types of surgical tools were used in the gynaecological procedure data. For instance, seven types of graspers appeared in the recorded videos, and only one of them was used in all procedures. Although, the various types of graspers shared similar functionality, they did not

necessarily have the same shape and visual appearance. Considering different makes of a particular surgical tool as a single class helps the model to learn general features, but not the optimal discriminating features for every tool type.

Scissors have more training samples (17% of training data) compared to irrigator, rotary blad, needle, mesh and bag. However, scissors have a lower classification result than these tools. In fact, scissors are used mostly in conjunction with grasper. In 54% of the frames that contained scissors in the training data, graspers were also present. Thus, the model had difficulties to learn accurate classification boundaries between those classes. Figure 4 (a) shows an image containing scissors, but the model classified it as grasper. The activation maps illustrate that the classification decision was based on the scissors' region. On the other hand, the grasper presented in Figure 4 (b) was wrongly classified as scissors with a probability of 80%. Those examples demonstrate the lower sensitivity of the scissors compared to almost all other tools, as shown in ROC curve (see Figure 3).

5 Conclusion

The complexity of surgical procedures affects CNN model training and classification performance. This study demonstrates classification capability of DenseNet-121 model on gynaecology data. The presence of various makes of some tools affected the model performance. Thus, more investigations are required to improve classification robustness for tools that have various types. Moreover, in future work, temporal information could be modelled and combined with spatial features.

Author Statement

Research funding: This work was supported by the German Federal Ministry of Research and Education (BMBF under project CoHMed/DigiMedOP grant no. 13FH5I05IA).

Ethical approval: This study received an ethical approval from the ethics commission of the Furtwangen University (application Nr. 19 -0306LEKHFU).

References

[1] Maier-Hein L, Vedula SS, Speidel S, Navab N, et al. Surgical data science for next-generation interventions. Nature Biomedical Engineering. 2017;1(9):691–6.

[2] Jalal NA, Alshirbaji TA, Möller K. Predicting surgical phases using CNN-NARX neural network. Current Directions in Biomedical Engineering. 2019;5(1):405–7.

[3] Alshirbaji TA, Jalal NA, Möller K. Surgical Tool Classification in Laparoscopic Videos Using Convolutional Neural Network. Current Directions in Biomedical Engineering. 2018 Sep 1;4(1):407–10.

[4] Abdulbaki Alshirbaji T, Jalal NA, Mündermann L, Möller K. Classifying smoke in laparoscopic videos using SVM.

[5] Jalal NA, Alshirbaji TA, Mündermann L, Möller K. Features for detecting smoke in laparoscopic videos. Current Directions in Biomedical Engineering. 2017;3(2):521–4.

[6] Jalal NA, Abdulbaki Alshirbaji T, Docherty PD, Neumuth T, Möller K. Surgical Tool Detection in Laparoscopic Videos by Modeling Temporal Dependencies Between Adjacent Frames. In: European Medical and Biological Engineering Conference. Springer; 2020. p. 1045–52.

[7] 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. 2017 Jan;36(1):86–97.

[8] Abdulbaki Alshirbaji T, Jalal NA, Docherty PD, Neumuth T, Möller K. A deep learning spatial-temporal framework for detecting surgical tools in laparoscopic videos. Biomedical Signal Processing and Control. 2021 Jul 1;68:102801.

[9] 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 [Internet]. 2020 May 1 [cited 2022 Feb 24];6(1). Available from: https://www.degruyter.com/document/doi/10.1515/cdbme-2020-0002/html

[10] Kondo S. Lapformer: surgical tool detection in laparoscopic surgical video using transformer architecture. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization. 2021;9(3):302–7.

[11] Abdulbaki Alshirbaji T, Jalal NA, Docherty PD, Neumuth T, Möller K. Cross-dataset evaluation of a CNN-based approach for surgical tool detection [Internet]. Zenodo; 2021 Jun 17 [cited 2022 Mar 1]. Available from: https://zenodo.org/record/4922977

[12] Alshirbaji TA, Jalal NA, Docherty PD, Neumuth T, Moeller K. Assessing Generalisation Capabilities of CNN Models for Surgical Tool Classification. Current Directions in Biomedical Engineering. 2021 Oct 1;7(2):476–9.

[13] Jalal NA, Alshirbaji TA, Möller K. Evaluating convolutional neural network and hidden markov model for recognising surgical phases in sigmoid resection. Current Directions in Biomedical Engineering. 2018;4(1):415–8.

[14] Alshirbaji TA, Jalal NA, Möller K. Data Recording Framework for Physiological and Surgical Data in Operating Theatres. Current Directions in Biomedical Engineering. 2020 Sep 1;6(3):364–7.

[15] Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely Connected Convolutional Networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) [Internet]. Honolulu, HI: IEEE; 2017 [cited 2022 Mar 1]. p. 2261–9. Available from: https://ieeexplore.ieee.org/document/8099726/

[16] Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-cam: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE international conference on computer vision. 2017. p. 618–26.