Tamer Abdulbaki Alshirbaji*, Nour Aldeen Jalal and Knut Möller

Surgical Tool Classification in Laparoscopic Videos Using Convolutional Neural Network

Abstract: Laparoscopic videos are a very important source of information which is inherently available in minimally invasive surgeries. Detecting surgical tools based on that videos have gained increasing interest due to its importance in developing a context-aware system. Such system can provide guidance assistance to the surgical team and optimise the processes inside the operating room. Convolutional neural network is a robust method to learn discriminative visual features and classify objects. As it expects a uniform distribution of data over classes, it fails to identify classes which are under-presented in the training data. In this work, loss-sensitive learning approach and resampling techniques were applied to counter the negative effects of imbalanced laparoscopic data on training the CNN model. The obtained results showed improvement in the classification performance especially for detecting surgical tools which are shortly used in the procedure.

Keywords: Surgical tool detection, loss-sensitive learning, laparoscopic videos, imbalanced data.

https://doi.org/10.1515/cdbme-2018-0097

1 Introduction

The last decades have witnessed a continuous development in the equipment and systems of minimally invasive surgeries (MIS) to enhance their functioning and improve the surgical outcomes. However, the complexity of MIS has constantly increased with the technical advancement and doctors cannot make use of all available data. Thus, a pressing need for analysing the data provided by MIS' devices arises to improve the quality of MIS. In this respect, the research community seeks to develop a context-aware system to assist the surgical team and optimise processes occurring in the operating room. It is essential for that system to recognise the ongoing surgical phase. This can be achieved by detecting the presence of surgical tools [1] since each surgical phase consists of a sequence of actions, each of which is associated with the use of a specific surgical tool or a combination of two or three tools.

Endoscopic video is inherently obtainable signal and contains valuable information because it shows the surgical tools, surgical actions and tissues. Analysing endoscopic videos is a very efficient approach for detecting the surgical tools, but it is very challenging. Because of the complexity of endoscopic images, specular reflection and tool obscureness by image perturbations like smoke [9], it is difficult to design visual features for identifying the surgical tools. In [2] visual features based on different image matching techniques were used to train support vector machine (SVM) with the Bag-of-visual-Words (BoW) for segmenting the endoscopic videos according to the detected tools. This method cannot identify the surgical tool correctly when smoke blurs the scene, or some parts of the tool are occluded by an anatomical structure or blood. Another method for surgical tool recognition based on segmenting the tool's shaft and tip was proposed in [3]. The tools with a metallic tip cannot be segmented due to light reflection on the tip.

Recent publications tended to use convolutional neural networks (CNN) to learn features from the endoscopic videos. CNN was used in [6,7,4] for segmenting and tracking surgical tools. In [5] phase and tool recognition were performed using a CNN architecture called EndoNet. The CNNs showed high detection accuracy, but not for all tools. Some tools are used much more often than others, and they appear in most of the endoscopic video frames. Consequently, frames belonging to those tool classes, denoted as majority classes, outnumber frames belonging to the other tools classes, denoted as minority classes. Such imbalanced data affects the generalisation of the learning process and reduce the CNN efficiency to classify the different tools.

In this work, a method for identifying surgical tools based on CNN was presented. Loss-sensitive learning approach and resampling techniques were proposed for facing the imbalanced data problem in order to improve classification performance of minority classes.

***Corresponding author: Tamer Abdulbaki Alshirbaji:** Furtwangen University, Institute of Technical Medicine, Villingen-Schwenningen, Germany, e-mail: abd@hs-furtwangen.de
Nour Aldeen Jalal, Knut Möller: Furtwangen University, Institute of Technical Medicine, Villingen-Schwenningen, Germany.

2 Method

2.1 Dataset

Cholec80 dataset was used for training and evaluating the CNN model. The dataset consists of eighty cholecystectomy videos recorded at the University Hospital of Strasbourg (25 fps) [5]. It also contains tool presence annotation at 1 fps.

In this dataset seven tools were used which are: grasper, hook, bipolar, scissors, clipper, specimen bag and irrigator. The tool was defined to be labeled as a present tool if at least half of the tip appears in the scene. The distribution of frames over the surgical tools is shown in **Figure 1**. The first forty videos were used for training the CNN, and the other videos for validation.

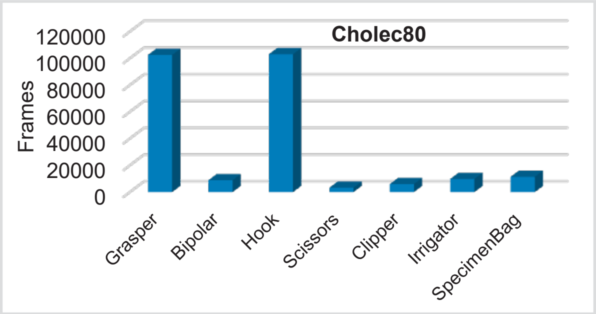

Figure 1: Frequent appearance of surgical tools in CHolec80.

2.2 Assessment Metric for Imbalanced Data

The imbalance ratio (IR) [8] was calculated to assess the imbalance level of the dataset. The imbalance ratio per class (IRPC) was also computed to assess the imbalance level of the tools (see **Equation 1**). The minority class has the highest value of IRPC.

$$IR = \frac{Class_{majority}}{Class_{minority}} \;,\; IRPC_{(i)} = \frac{Class_{majority}}{Class_{(i)}} \qquad \textbf{(1)}$$

where $Class_{majority}$ is the number of frames belonging to the majority class, $Class_{minority}$ is the number of frames belonging to the minority class, $IRPC_{(i)}$ imbalance ratio of class (i), and $Class_{(i)}$ is the number of frames belonging to class (i).

2.3 CNN Model

The pre-trained model AlexNet proposed by Krizhevsky was fine-tuned to perform surgical tool classification [5,10]. The model consists of five convolutional layers and three fully-connected layers. The deep architecture of AlexNet model results in very robust visual features. Tool classification in endoscopic videos is considered as a multi-label classification task since more than one surgical tool could appear in the scene. Therefore, the last fully-connected layer was replaced with seven fully-connected layers, each of which performs a binary classification for a specific surgical tool and its output is the confidence of tool presence.

Stochastic gradient descent (SGD) algorithm was used for training the model. The cross-entropy function was chosen to calculate the loss since detecting tool presence is a binary classification task. The mean cross-entropy loss of a tool for a batch of images was calculated according to **Equation 2**.

$$\mathrm{loss_t} = \frac{-1}{\mathrm{N}}\sum_{\mathrm{n=1}}^{\mathrm{N}}[\mathrm{p_n}\log(\vartheta(\delta_\mathrm{n})) + (1-\mathrm{p_n})\log(1-\vartheta(\delta_\mathrm{n}))] \qquad \textbf{(2)}$$

where N batch size, $\mathrm{p_n} \in \{0,1\}$, ϑ is sigmoid function, and δ_n is confidence of tool presence. The losses of the tools are summed to give the total loss (**Equation 3**). Stochastic gradient descent algorithm modifies the model weights in a manner to minimise the total loss.

$$\mathrm{Loss_{total}} = \sum_{\mathrm{t=1}}^{7}\mathrm{loss_t} \qquad \textbf{(3)}$$

2.4 Techniques to Counter Imbalanced Data Problem

Two techniques were used to reduce the impact of imbalanced data on the classification performance. The first technique deals with imbalanced data problem at the data level, while the second treats the learning process.

2.4.1 Sampling Technique

This technique attempts to obtain a dataset with a more balanced distribution by applying some sampling mechanisms on the imbalanced dataset. Under-sampling and oversampling mechanisms were implemented on our training dataset. The majority classes which have a large number of samples are under-sampled by removing some frames randomly. In contrast, the minority classes are oversampled by taking all their frames and replicating some of them. In this way, a relatively balanced data set is obtained.

2.4.2 Loss-Sensitive Learning

SGD tries to reduce the total loss without taking into consecration the data distribution, as it treats the losses of the different classes equally (**Equation 3)**. Thus CNN can learn robust features to identify all classes when the data has a uniform distribution. Loss-sensitive approach is used to increase the generalisation of the learning process in case of imbalanced data. Loss-sensitive learning is based on assigning a weighting factor for each tool's loss as shown in **Equation 4**. The weighting factors were set according to proportions of classes in the training set. The higher proportion of data a class has, the lower weighting factor it has. Therefore, the minority classes have higher weighting factors than the majority classes. The weighting factors attempt to compensate the biasing effect of imbalanced data.

$$\mathrm{Loss}_{\mathrm{total}} = \sum_{t=1}^{7} \mathrm{W}_{\mathrm{t}} * \mathrm{loss}_{\mathrm{t}} \qquad (4)$$

2.5 Implementation

AlexNet model was fine-tuned firstly without applying the previous techniques and a model, denoted as ToolMod, was resulted. ToolMod was fine-tuned using loss-sensitive learning approach after resampling the training set and a new model, denoted as BToolMod, was resulted.

The fine-tuning process was carried out using CAFFE framework on a NVIDIA GEFORCE 840M graphics card which has 8 GB of memory. The training process of ToolMod was run for 80K iterations, and the base learning rate was set 10-4.

3 Results

The first forty videos of Cholec80 dataset was used as a training set for ToolMod. This training set was resampled to obtain a relatively balanced set. The imbalance ratio of each tool in both sets is shown in **Table 1** and **Table 2**.

Table 1: Imbalance ratio per class (IRPC) of surgical tools in ToolMod training set.

Tool	Grasper	Bipolar	Hook	Scissors	Clipper	Irrigator	Specimen Bag
IRPC	1	13.83	1.17	34.97	17.65	10.55	9.86

Table 2: Imbalance ratio per class (IRPC) of surgical tools in the balanced training set.

Tool	Grasper	Bipolar	Hook	Scissors	Clipper	Irrigator	Specimen Bag
IRPC	1	2.41	1.43	2	2.85	1.81	2.69

The imbalance ratio of the ToolMod training set is equal to 35 which reflects a very high imbalance level. Applying resampling techniques reduced the IR to 3. **Figure 2** shows the co-occurrence of one and two tools which is represented by the chord diagram for the imbalanced and balanced training sets.

Table 3: Classification performance of ToolMod model.

Tool	**Accuracy**	**Recall**	**Precision**
Grasper	76.68 %	91.78 %	68.71 %
Bipolar	98.27 %	70.21 %	92.36 %
Hook	91.01 %	90.13 %	93.50 %
Scissors	98.45 %	16.13 %	63.07 %
Clipper	97.74 %	56.01 %	60.75 %
Irrigator	96.42 %	40.86 %	66.91 %
Specimen Bag	96.96 %	77.53 %	72.2 %
Average	93.65 %	63.24 %	73.93 %

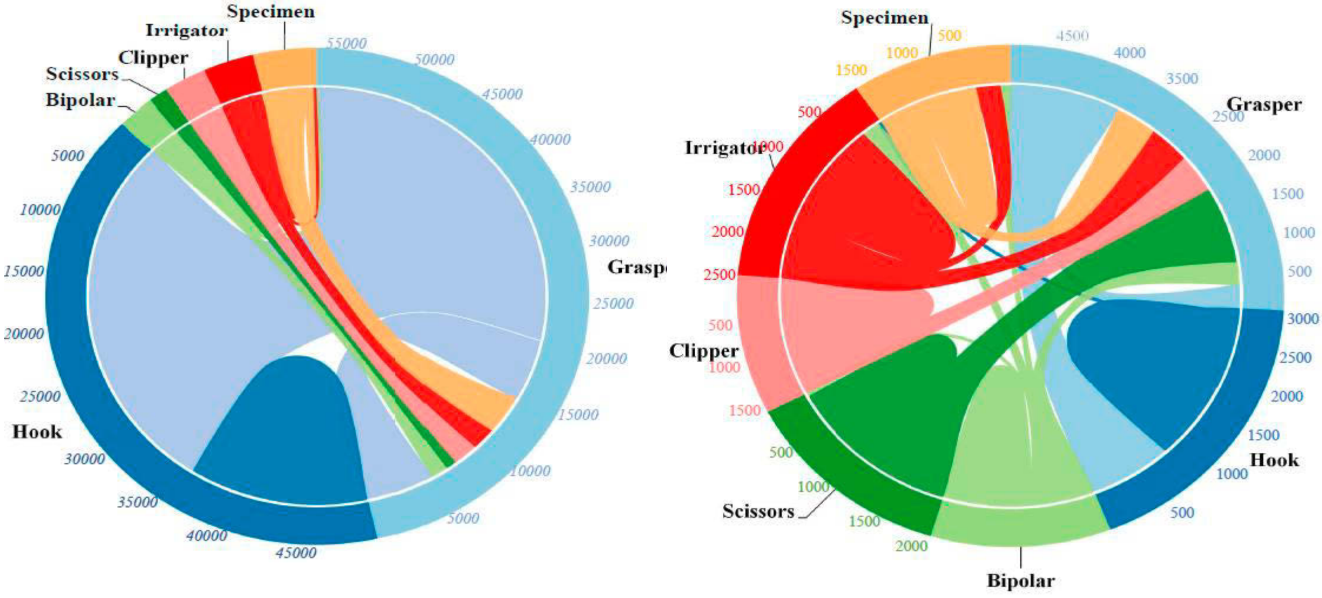

Figure 2: Chord diagram for: (left) training set of ToolMod, (right) training set of BToolMod. The co-occurrence of two tools is represented by an arc.

Table 4: Classification performance of BToolMod model.

Tool	Accuracy	Recall	Precision
Grasper	78.99 %	89.58 %	76.18 %
Bipolar	98.58 %	83.25 %	90 %
Hook	90.75 %	93.80 %	93.22 %
Scissors	98.57 %	56.39 %	68.7 %
Clipper	98.48 %	71.76 %	78.3 %
Irrigator	96.50 %	66.94 %	70.2 %
Specimen Bag	95.76 %	88.62 %	66.4 %
Average	93.95 %	78.62 %	77.57 %

Accuracy, recall and precision were used as evaluation metrics. To compare between the ToolMod and BToolMod, they were evaluated on the same testing set. The classification performance of the models is presented in **Table 3** and **Table 4**.

4 Discussion

Surgical tools are not used with the same frequency during the intervention. Grasper, for instance, is used in combination with other surgical tools in cholecystectomy and it appears in most of the endoscopic video, while some tools are used shortly in the whole procedure. Variation in the frequent appearance of the surgical tools results in imbalanced data problem. In Cholec80 dataset, Scissors and Irrigator appeared in a few frames and had the highest IRPC. Due to the high number of frames belonging to Grasper and Hook, ToolMod was biased towards those tools, and it can identify them very robustly which is reflected in their high recalls. Due to the biasing, ToolMod cannot detect Scissors and Irrigator, and they have low recalls.

During the training process, weights of CNN model are modified to reduce the total loss and obtain the best average accuracy without considering recalls or precisions of the individual tools. Therefore, ToolMod classifies most of the frames belonging to the minority class as a Grasper. Because of the tiny number of those frames (less than 3% of training data belongs to Scissors), the accuracy of the Scissors and the average accuracy are high. Moreover, Grasper has low precision due to high false positive coming from that misclassification.

Applying under-sampling and oversampling techniques results in a relatively balanced training set as shown in **Figure 2**. Oversampling technique may cause overfitting when same frames are duplicated many times. To avoid overfitting, only some of Scissors' frames were duplicated one time.

Training ToolMod on the balanced training using loss-sensitive approach reduced the CNN biasing and improved the classification performance for minority classes. Recalls of Scissors and Irrigator enhanced by more than 40% and 25% respectively. Consequently, the number of frames misclassified as Grasper decreased leading to lower false positive and higher precision.

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 has been obtained from all individuals included in this study. Ethical approval: The research related to human use complies with all the relevant national regulations, institutional policies and was performed in accordance with the tenets of the Helsinki Declaration, and has been approved by the authors' institutional review board or equivalent committee.

References

[1] Speidel, S., Benzko, J., Krappe, S., Sudra, G., Azad, P., Müller-Stich, B., Gutt, C., Dillmann, R.: Automatic classification of minimally invasive instruments based on endoscopic image sequences. In: SPIE Medical Imaging, vol. 7261 (2009).

[2] Primus MJ, Schoeffmann K, Böszörmenyi L (2015) Instrument classification in laparoscopic videos. In: International workshop on content-based multimedia indexing, Prague, pp 1–6.

[3] Primus M.J., Schoeffmann K., Böszörmenyi L., ¨ Temporal segmentation of laparoscopic videos into surgical phases. In Content-Based Multimedia Indexing (CBMI), 2016 14th International Workshop, IEEE, pp. 1–6.

[4] Laina, I., Rieke, N., Rupprecht, C., Vizcaíno, J.P., Eslami, A., Tombari, F. and Navab, N., 2017. Concurrent Segmentation and Localization for Tracking of Surgical Instruments. arXiv preprint arXiv:1703.10701.

[5] 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.

[6] García-Peraza-Herrera, L.C., Li, W., Gruijthuijsen, C., Devreker, A., Attilakos, G., Deprest, J., Vander Poorten, E., Stoyanov, D., Vercauteren, T. and Ourselin, S., 2016, October. Real-Time Segmentation of Non-rigid Surgical Tools Based on Deep Learning and Tracking. In International Workshop on Computer-Assisted and Robotic Endoscopy (pp. 84-95). Springer, Cham.

[7] Choi, B., Jo, K., Choi, S. and Choi, J., 2017, July. Surgical-tools detection based on Convolutional Neural Network in laparoscopic robot-assisted surgery. In Engineering in Medicine and Biology Society (EMBC), 2017 39th Annual International Conference of the IEEE (pp. 1756-1759). IEEE.

[8] Charte F, Rivera AJ, del Jesus MJ, Herrera F. Addressing imbalance in multilabel classification: Measures and random resampling algorithms. Neurocomputing. 2015 Sep 2;163:3-16.

[9] Alshirbaji TA, Jalal NA, Mündermann L, Möller K. Classifying smoke in laparoscopic videos using SVM. Current Directions in Biomedical Engineering.;3(2):191-4.

[10] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. InAdvances in neural information processing systems 2012 (pp. 1097-1105).