Proceedings on Automation in Medical Engineering
Proc AUTOMED, Vol 2, No 1 (2023): Article Ref 749
www.journals.infinite-science.de/automed/article/view/749

A deep learning based instrument detection approach for automated surgical systems

F. Bajraktari[1*], K. Fleissner[1], and P. P. Pott[1]

[1] *Institute of Medical Device Technology, University of Stuttgart, Stuttgart, Germany*
[*] *Corresponding author, email: flake.bajraktari@imt.uni-stuttgart.de*

Abstract: The demand of automated systems based on artificial intelligence in healthcare has remarkably increased in the last few years. Due to the growing shortage of skilled workers and the associated error potentials, the reduction of the workload is essential for the care of patients. Surgery assisting tasks could be automated in order to overcame these negative effects. This study presents the development and evaluation of a deep learning system for the recognition of surgical instruments. Already implemented algorithms based on Convolutional Neural Networks (CNN) were used. The object detection was carried out with YOLOv5. Altogether 18 models have been trained on a self-generated dataset of around 800 images. A mean average precision (mAP) of 0.978 for the recognition of three classes, and an mAP of 0.874 for the recognition of six classes was achieved.

I. Introduction

In the last few years robotic scrub nurse (RSN) systems have increasingly become the focus of research in the field of surgical assistance systems. Due to the growing shortage of operating room (OR) technicians ("scrub nurses"), the demand for automated systems in the OR has increased rapidly. The lack may lead to delays in surgeries, which can cause longer waiting times for patients. Moreover, an increased workload for the existing staff can result in a higher amount of errors and lower quality in care [1]. Under-staffing of nurses makes the deployment of automated systems in the OR crucial and inevitable in the future.

An RSN performs OR technician tasks, such as the handling and preparation of instruments and documentation of the surgery [2]. Assistance systems are being developed for instrumentation tasks, where recognition and instrument selection is based on artificial intelligence (AI)-powered computer vision [3]. This, furthermore, enables an efficient way of documenting the surgery and automated generation of a report [4].

This work focusses on the sensing part of an RSN, detecting the surgical instruments. Several studies conducted in the last decade include the development of tool detection systems as an auxiliary system to surgical phase detection, since tool presence tasks and phase recognition strongly correlate [5]. Most commonly, convolutional neural networks (CNN) are used in order to detect both surgical phases and surgical tools since they show outstanding performance in detecting specific objects [6, 7]. Recent advances in surgical tool detection and tracking, such as EndoNet [5], have enabled near real-time detection methods. *Wang et al.* [8] proposed a real-time method based on YOLOv4 [9] for the recognition of instruments during a laparoscopic intervention. However, since nursing environments display different features and conditions as compared to laparoscopic images, it is necessary to develop specified recognition systems for the nursing application.

In this paper, we propose a real-time deep learning method based on YOLOv5 to detect surgical instruments on the instrument tray for the assistance of surgeons and staff.

II. Material and methods

Object recognition in general shows to be very accurate when using CNNs. Besides region-based methods, it has been shown that the real-time capable YOLO algorithm is suitable under the given conditions and requirements. YOLOv5 was used to develop the instrument detection system with bounding boxes in order to provide real-time assistance during surgeries. The network was trained on images of the basic surgical instruments: scissors, forceps and scalpel. The images were taken with a Canon EOS 200D (Canon Inc., Tokio, Japan) and an Apple iPhone 11 (Apple Inc., Cupertino, USA). The ambient illumination of the photos was oriented to DIN EN 12464-1, which determines the illumination levels of the surgical field. In addition to the instruments to be recognized, a number of interfering factors – such as gloves, sterile compresses and cannulas – were placed in the image in order to simulate real-world scenarios. Moreover, images were taken without objects to show examples of true negatives. A total of 816 photos was taken for the training dataset and the validation dataset, of which 153 contained no objects. For the test dataset, an additional 84 photos were provided. The data were divided into three classes (scissors, forceps, scalpel) and six classes with more subtle differences between the instruments (pointed and blunt scissors, anatomical and surgical forceps, scalpel with a fixed and replaceable blade).

The distribution of the images of the instruments into the classes can be seen in Table 1.

Table 1: Distribution of the dataset images.

Number of classes	Instrument	Number of images
3	Scissors	722
	Forceps	716
	Scalpel	726
6	Pointed scissors	367
	Blunt scissors	351
	Anatomical forceps	369
	Surgical forceps	349
	Scalpel with a fixed blade	373
	Scalpel with a replaceable blade	352

Additionally, the training dataset was augmented through rotating or changing image parameters, such as hue or saturation – via the integrated augmentation of YOLOv5 and partially via external augmentation. After having labeled the instruments in the images, the virtual machine environment GoogleColab was used to train the network models with around 80 % of the dataset. The models were trained from scratch and by means of transfer learning. For the latter, YOLOv5 provides pretrained models trained on 80 classes of the COCO dataset.

For the evaluation and comparison of the different approaches, the performance metrics mean average precision (mAP) and F1-Score were used. The mAP is computed at an intersection over union (IoU) of 0.5. The F1-Score displays the harmonic mean of precision and recall. The calculation of the same metrics on disjoint test data allows the verification of the generalization ability of the models.

III. Results

We implemented a surgical detection system that is able to detect three or six classes as defined in section II. Figure 1 presents an exemplar display of the results generated by a detection model that comprises six distinct classes.

Figure 1: Example image of the instrument tray with six detected classes with bounding boxes and the probability of each class.

Overall, nine models were trained for each number of classes, i.e. 18 models were trained. The models differed among others in weights size, network depth, use of transfer learning and external data augmentation.

The results of the models with highest performance are shown in Table 2. They were obtained using the available pretrained weights *yolov5x*.

Table 2: Results of the models with highest mAP on the validation dataset evaluated with the validation and test dataset for 3 and 6 classes.

Number of classes	Dataset	mAP (0.5)	F1-Score
3	Validation	0.985	0.977
	Test	0.978	0.944
6	Validation	0.987	0.981
	Test	0.874	0.830

When detecting six classes, the confusion matrix indicated that the anatomical and surgical forceps were frequently misclassified. With respect to detecting three classes, the misclassifications between the instruments remained below 1 % and were primarily concentrated in the detection of the forceps as well. Instrument confusion was minimal for the selected models in Table 2. Loss functions suggest that only models trained from scratch, with external augmentation, or using smaller weight models may exhibit poor generalization ability or overfitting.

IV. Discussion and Conclusions

Overall, models for the detection of three and six classes of surgical instruments were implemented with high accuracy results. Models with larger pretrained weights generally achieved higher performance, as more parameters were included for the detection. Models utilizing additional synthetically augmented data produced suboptimal accuracy results, contrary to expectations for an augmented dataset. This could be attributed to the fact that the combination of augmentations from YOLOv5 and additional external augmentations led to excessively altered images. The loss functions of the models with the highest performance did not provide conclusive evidence of overfitting; true generalization capability was established through verification on the test data. While the findings in this study offer valuable insights, it is important to acknowledge that their generalizability may be limited.. Future work should involve training the models on larger datasets that encompass every possible scenario in the OR and testing them on data from real-world applications.

Author's Statement

The author states no funding involved. Authors state no conflict of interest.

References

[1] D. Muralidhar *et al.*, "Collaborative Robot as Scrub Nurse," *Current Directions in Biomedical Engineering*, vol. 7, no. 1, pp. 162–165, 2021.
[2] *OP-Handbuch: Grundlagen, Instrumentarium, OP-Ablauf*, 7th ed. Berlin: Springer, 2021.
[3] C. Perez-Vidal *et al.*, "Steps in the development of a robotic scrub nurse," *Robotics and Autonomous Systems*, vol. 60, no. 6, pp. 901–911, 2012.
[4] D. Wilhelm *et al.*, "Künstliche Intelligenz in der Allgemein- und Viszeralchirurgie," *Der Chirurg; Zeitschrift fur alle Gebiete der operativen Medizen*, vol. 91, no. 3, pp. 181–189, 2020.
[5] A. P. Twinanda *et al.*, "EndoNet: A Deep Architecture for Recognition Tasks on Laparoscopic Videos," *IEEE transactions on medical imaging*, early access.
[6] K. C. Demir *et al.*, "Surgical Phase Recognition: A Review and Evaluation of Current Approaches," *TechRxiv. Preprint.*, 2022.
[7] I. Arel *et al.*, "Deep Machine Learning - A New Frontier in Artificial Intelligence Research [Research Frontier]," *IEEE Comput. Intell. Mag.*, vol. 5, no. 4, pp. 13–18, 2010.
[8] Y. Wang *et al.*, "Object Detection of Surgical Instruments Based on YOLOv4," in *2021 6th IEEE International Conference on Advanced Robotics and Mechatronics (ICARM)*, Chongqing, China, 2021, pp. 578–581.
[9] A. Bochkovskiy *et al.*, "YOLOv4: Optimal Speed and Accuracy of Object Detection," *arXiv. Preprint.*, 2020.
[10] T.-Y. Lin *et al.*, "Microsoft coco: Common objects in context," in *European conference on computer vision*, 2014, pp. 740–755.