

Surgical Instrument Detection Using Deep Neural Networks and Image Preprocessing with Image Processing Filters in Laparoscopic Images

Abstract:

Background and Objective: The complexity of laparoscopy requires specialized training and evaluation. Analyzing streamed videos during surgical procedures may enhance surgical training. Fatigue and costs associated with such analysis can be significantly reduced using an automated detection system, thereby reducing potential medical errors. Identification of surgical instruments is a fundamental task for analyzing and evaluating surgical videos. However, the focus has mainly been on minimally invasive surgery (MIS) and cataract surgery. Therefore, identifying surgical instruments can be a crucial and supportive step in reducing evaluation errors, training human resources, as well as analyzing and summarizing the surgical process. Various tools, including deep learning for surgical instrument detection, can be employed, but appropriate image preprocessing is necessary to improve the results. The aim of this paper is to present a method to enhance the results of deep learning models based on CNN and LSTM by preprocessing and edge detection of input images.

Method: Our proposed method consists of three steps. In the first step, images are preprocessed using edge detection filters like Sharp, Sobel, Canny, Roberts, and Prewitt. Then, a hierarchical CNN and LSTM-based method is used to extract suitable features from the image and perform classification. A CNN model was used to learn spatial features from laparoscopic images. An LSTM network named LSTM-clip was utilized to learn temporal dependencies from intermediate video clips, and then temporal dependencies in full surgical videos were modeled using another LSTM named LSTM-video.

Results: The proposed method was implemented on the large publicly available Cholec80 dataset. Results were evaluated based on Precision, Recall, Accuracy, and AUC metrics. Two types of experiments were conducted in this research; in the first experiment, the presence or absence of surgical instruments in the image was examined. Based on this experiment and the results obtained, the deep learning model using the Prewitt filter showed better performance. In the second experiment, the detection of surgical instrument types was investigated. In this experiment, the deep learning model using the Sharp filter outperformed other filters.

Key Words:

Surgical instrument detection, edge detection, deep learning, image preprocessing

1.Introduction

Machine vision and image processing have widespread applications in various fields and industries today. The goal of machine vision is to analyze and extract information from images for better understanding. Different processing tools are used in this field depending on the type of information that needs to be extracted from the image. The advancement of machine learning algorithms such as deep learning in recent years has led to increased attention to machine vision. One of the areas where machine learning, especially machine vision, is highly applicable today is in the field of medicine. Applications such as cancer diagnosis using medical images [1], detection of medical instruments [2], estimating age from dental images [3], and more are examples of machine vision applications in medicine and healthcare.

The detection of surgical instruments from a self-centric perspective in the operating room is a fundamental task for the development of intelligent systems that can assist surgeons in real-time. For example, detecting a tool can inform surgeons and prevent incidents such as leaving gauze inside the body. Recently, various approaches for detecting surgical instruments, especially in minimally invasive surgeries (MIS), have been proposed [4,5]. The numerous advantages of minimally invasive surgery, such as shorter recovery time, less pain and blood loss, and better cosmetic outcomes, have made it a preferred choice over traditional open surgeries [6]. In laparoscopy, surgical instruments are inserted through small incisions in the abdominal wall and monitored using a laparoscope. The specific manipulation of surgical instruments and indirect observation of the surgical scene pose additional challenges in performing laparoscopic procedures [7]. The complexity of laparoscopy requires specialized training and evaluation for surgical assistants to acquire the necessary skills. For such training and evaluation, videos of previous procedures performed by experienced surgeons can be utilized. Fatigue and the cost of training and evaluation can be significantly reduced, among other things, by using an automated tool detection system.

In robotic-assisted interventions, surgical instruments are controlled by a robot specifically designed to assist a surgeon who needs real-time understanding of the current task. This robot aids the surgeon in controlling surgical instruments and provides real-time task awareness. In this approach, identifying the presence, position, and placement of surgical instruments in robotic surgeries can also be beneficial. On the other hand, the actual location and movement of the instruments can significantly contribute to the evaluation of surgeries and the generation of useful operation summaries [8, 9]. In the future, operating rooms will utilize machine assistance for knowledge-based analysis of available data to enable new decision support systems or context-aware systems (CAS). The goal of CAS is to optimize surgical treatment by analyzing and correlating data from various medical fields, improving surgical performance, and providing relevant information to human operators during surgical interventions. Analyzing the use of surgical instruments is a fundamental objective in CAS as it enables the recognition of surgical phases. Additionally, automatic detection of surgical instruments may be utilized for automated reporting during the surgical process [10, 11, 12].

Various methods have been proposed for the detection of surgical instruments, with machine learning-based approaches and algorithms related to video analysis having wide applications in

this field [13]. Automatic analysis of surgical videos for gathering information about surgeries has been crucial. By combining machine vision and machine learning, performance can be improved to enhance the safety and efficiency of surgeries. Significant attention to the analysis of surgical videos has sparked active research in the medical machine vision community [14, 15]. Tool detection based on video is a desirable method in minimally invasive surgery, as endoscopic videos provide an easy source for recording information. However, tool detection based on video remains a challenging task. In particular, detecting surgical instruments involves multi-object detection tasks with various combinations of tools. Laparoscopic images may face limitations and challenges due to the rapid movement of the laparoscope camera, smoke, tissue deformations, and blood covering the instruments [16]. Figure 1 illustrates some of these challenges. Therefore, much research focus lies on tool detection in such obstructed scenarios.

Traditional methods for counting surgical instruments usually involve medical staff manually checking the name, quantity, and model of each instrument on a list to ensure compatibility with the surgical plan, as well as inspecting them for damages. However, due to the wide variety of surgical instruments and the possibility of confusion between instruments with similar appearances, as well as factors such as patient cases, individuals, and environments, the efficiency and accuracy of manual tool counting are limited. This method also requires significant human resources and is prone to errors and mistakes, leading to difficulties in meeting final counting needs. Despite proposed methods such as optimizing tool placement [17], improving inventory systems [18], and introducing equipment maps [19], the inadequacy of manual counting remains unsolved. To enhance the method of counting surgical instruments, radio frequency identification (RFID) technology has been introduced [20]. RFID uses radio waves to identify specific targets and read/write relevant data. After installing RFID tags on surgical instruments, an RFID reader can identify and locate them. This technology offers advantages such as real-time high performance and ease of operation, but some specialized tools may be difficult to tag, and the cost of equipment and tags is relatively high.

With advancements in artificial intelligence, it is now possible to classify and count surgical instruments using computer vision technology. Li et al. [21] conducted experiments on a dataset for surgical instrument identification, comparing and analyzing the performance differences of three common object detection algorithms, Faster R-CNN, Mask R-CNN, and SSD, and evaluating their usability in various scenarios. Object detection algorithms based on deep learning can assist medical staff in quickly and accurately counting surgical instruments by identifying and processing images of the tools, thereby improving operating room efficiency and reducing medical incidents.

This article presents a method to improve the detection of tools in previous studies. The proposed method in this research includes image preprocessing based on edge detection algorithms and the use of the LSTM-CNN model for tool detection and classification in surgical procedures. Accordingly, this article consists of 5 sections. The first section includes the introduction and important aspects related to the subject of the article. The second section reviews some related works to the proposed method. The third section describes the proposed method in detail. The

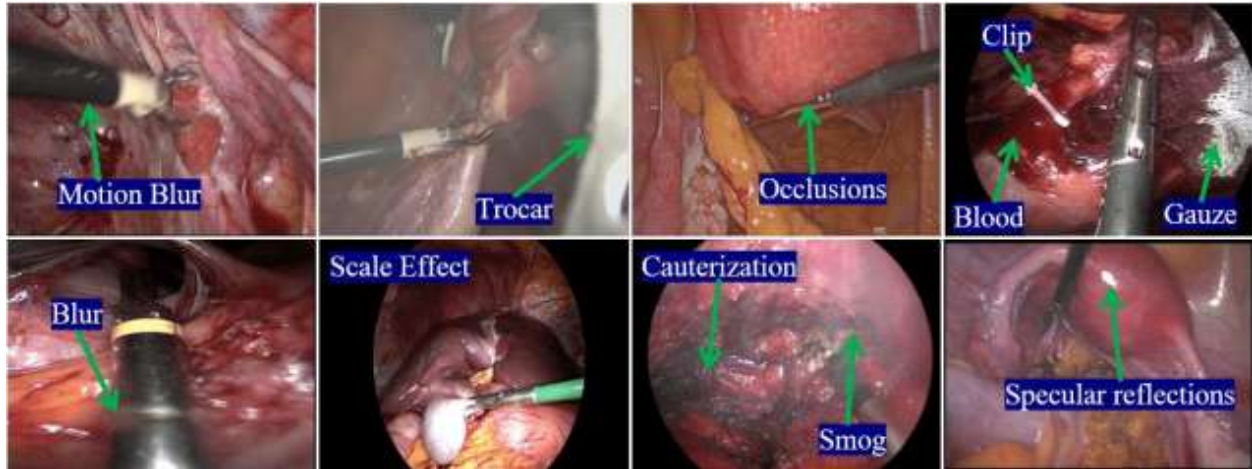


Figure (1): Examples of challenges in medical image analysis for surgical tool detection

fourth section presents the results of the proposed method. Finally, the last section provides a summary of the proposed method and suggestions.

2.Related works

Currently, deep learning-based machine vision methods are widely used in surgical tool detection. S. Wang et al. [22] proposed a multi-label classification method based on deep learning for surgical tool identification in laparoscopic surgery videos. This method combines VGGNet and GoogLeNet and combines the results of the models through cumulative learning to obtain the final result. Y. Wang et al. [23] and Y. Zhou et al. [24] proposed real-time surgical tool detection based on YOLOv4 and YOLOv5 models respectively, to assist surgeons in minimally invasive surgery. Kaidi Liu et al. [25] proposed an enhanced feature fusion network (EFFNet) for real-time surgical tool detection during surgery, which helps surgeons obtain more comprehensive visual information. A. Jin et al. [26] proposed using region-based convolutional neural networks for tracking and analyzing surgical tools in surgical videos and automatically evaluating the performance of surgeons. Kurmann et al. [27] proposed a method for identifying and estimating the status of surgical tools in minimally invasive surgery scenarios. This method first identifies the surgical tools based on RetinaNet and then estimates the 3D position of the surgical tools based on MASK R-CNN. These two network models use shared layers of convolutional neural networks, which may improve the accuracy of position estimation while maintaining the accuracy of tool detection.

To address the challenges mentioned in Figure 1, several studies have utilized the relationships between surgical tools and surgical phases and employed multi-task convolutional neural network architectures for simultaneous tool and phase identification [28,29]. Twinanda et al. introduced a baseline model called EndoNet [28]. EndoNet performs phase identification and surgical tool presence detection in laparoscopic cholecystectomy videos. Similarly, Jin et al. developed a multi-task deep learning framework but also defined a new correlation loss to leverage the tool-phase relationship. Abdalbaki Alshirbaji et al. utilized techniques such as re-sampling and weighted loss to cope with this issue [30]. Additionally, since surgical videos provide sequential data, recurrent neural networks (RCNNs) have been used to consider the temporal aspect of the surgical process

and improve the classification of CNNs [31]. Mishra et al. proposed the use of Long Short-Term Memory (LSTM) networks to learn temporal dependencies in neighboring frames [32]. Similarly, Chen et al. examined the use of 3D convolutional neural networks to learn spatiotemporal features from short video clips [33]. Elhajj et al. employed a CNN-RNN pipeline for tool identification in surgical videos. Instead of training both networks end-to-end, they introduced a reinforcement strategy that uses weakly supervised classes to guide the training of CNNs to align with the output of RNNs [34]. Nwoye et al. proposed a deep learning framework trained using binary annotations to perform tool presence and tool tracking [31]. They used ConvLSTM (Convolutional LSTM) to learn spatiotemporal features in surgical videos. Recently, Wang et al. demonstrated the capability of using Graph Convolutional Networks (GCNs) to learn temporal relationships across consecutive frames [35]. Their method involved using video sequences of labeled frames and unlabeled close processes, and achieved significant improvements compared to reference methods.

In recent years, deep learning tools have been widely used for surgical instrument detection. Among these, Bajraktari et al. [36] proposed a method based on CNN neural network for detecting surgical instruments in limited surgeries. In another study, Ran et al. [37] suggested a method based on the YOLOv7x model for detecting surgical instruments in open and complex surgeries. Zheng [38] utilized the YOLOv7 model for detecting surgical instruments in limited and small surgeries. Bai [39] employed tracking and deep learning-based detection techniques for surgical instrument detection.

In the reviewed methods, it has been observed that few studies focus on image preprocessing and simplification for surgical tool detection. In this method, while utilizing the approach presented in [16], which is based on LSTM-CNN, edge detection has been used as a preprocessing step to improve the detection of surgical tools. The goal of this step is to eliminate noise and non-tool tissues and simplify the image data. Thus, edge information can enhance the phase transitions and movements in the LSTM-CNN model. The following section provides a detailed explanation of this process.

The main challenges of previous articles include not considering specific features and appropriate preprocessing for image analysis. To improve processing speed and meet the constraints of real-time processing, it seems necessary to reduce the volume of irrelevant information from the image and have image processing models focus only on regions where changes occur. Therefore, presenting methods for representing these changes is essential. In this article, the use of edge detection tools is proposed to address this issue. The reason for using edge information in the image is to eliminate irrelevant information and emphasize useful information.

3. Proposed Method

This section describes the tools used and the proposed method in the article.

3.1 Dataset

The Cholec80 dataset, created by A. P. Twinanda et al. [28], was used in this work. This dataset consists of endoscopic videos of 80 cholecystectomy surgeries performed by 13 surgeons at the University Hospital of Strasbourg. The videos were recorded at a frame rate of 25 frames per

second (fps), and the surgical phases were continuously labeled, while the surgical tools were labeled at a rate of 1 Hz. Therefore, each frame with a tool label is surrounded by 48 unlabeled frames. Surgeons used seven tools for the laparoscopic intervention, including grasper, bipolar, hook, clipper, scissors, irrigator, and specimen bag. The length of the videos had an average of 2095 seconds (minimum: 739, maximum: 5993, first quartile: 1641, third quartile: 2882).

3.2 Edge Detection

Edge detection is one of the crucial stages in image processing and machine vision that deals with detecting and distinguishing edges in images. Edges are points in the image that indicate sharp changes in color or intensity of light and define the boundaries of different objects. The main goal of edge detection in machine vision is to detect and extract edges present in the image for use in other tasks such as pattern recognition, object detection, and image processing for visual analysis. Various algorithms exist for edge detection that utilize different filters, methods for detecting changes in color intensity, and image processing techniques to identify edges. Through edge detection, machines will be able to obtain useful information from the image and perform more precise and intelligent actions. This process enhances the performance of artificial intelligence systems and machine vision in pattern recognition, object detection, and image analysis. In this article, the following algorithms are independently used for image edge detection.

- Sharp Algorithm: The Sharp algorithm is a simple method for edge detection that highlights differences between neighboring pixels by filtering the image with sharpening filters. This algorithm usually enhances and emphasizes edges in the image.
- Sobel Algorithm: The Sobel algorithm is a widely used method for edge detection that calculates the gradient of the image using Sobel filters in perpendicular directions. This algorithm effectively identifies horizontal and vertical edges in the image.
- Canny Algorithm: The Canny algorithm is an advanced and precise method for edge detection that involves several different stages such as Gaussian filtering, gradient calculation, non-maximum suppression, and edge detection and merging. This algorithm is well-suited for identifying thin and precise edges in images.
- Roberts Algorithm: The Roberts algorithm is a simple method for edge detection that calculates differences between neighboring pixels using two 2x2 filters. This algorithm is useful for identifying thin and angled edges in images.
- Prewitt Algorithm: The Prewitt algorithm is relatively similar to the Sobel algorithm and calculates the image gradient using Prewitt filters in horizontal and vertical directions. This algorithm is relatively fast and efficient for edge detection in images.

Overall, different edge detection algorithms utilize filtering and gradient calculations to identify and emphasize edges in images, each with its own features and capabilities. In this article, edge extraction is used to remove incorrect information from the image and extract useful information from it. The goal of edge extraction is to retain useful information and eliminate unnecessary details from the image. According to Figure 2, in this research, the following edge detection algorithms were independently used. That is, only one of the edge detection algorithms was used in each program execution, and the results were examined. Two objectives are pursued by this

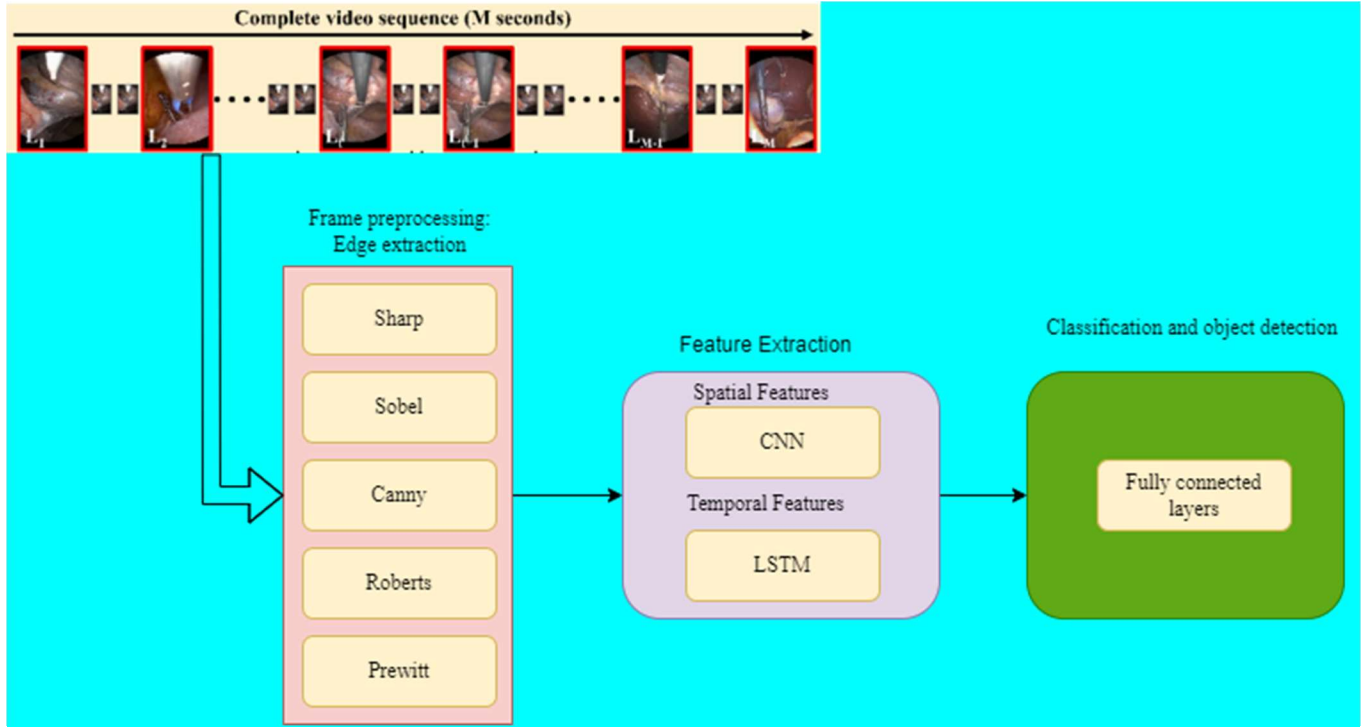


Figure (2): General Process of the Proposed Method

action. The first objective is to investigate the impact of using edges in surgical tool detection, and the second objective is to examine each of the edge detection algorithms.

3.4 Feature Extraction of Spatial and Temporal Features from Images

The proposed method in this article consists of three main steps. In the first step, each frame of the video is preprocessed, and the edges of the frames are extracted. Then, a CNN model is applied to the resulting image for spatial feature extraction, and two LSTM models are applied to extract temporal features. LSTMs are RNN-based neural network models that can preserve the sequence of events. The proposed method in this research is an improved version of the model presented in citation. The overall process of the proposed method is illustrated in Figure 2. According to Figure 2, this project focuses on investigating the application of edge detection methods to improve the detection of surgical tools in laparoscopic images.

As mentioned in previous sections, this article is based on reference [16]. Therefore, the innovation of our article lies in using edge detection to improve the performance of the model presented in reference [16]. Based on the process of edge detection, feature extraction, and object recognition outlined in this article, the feature extraction and object recognition section is based on the model presented in reference [16]. Therefore, the extraction process includes two steps: extracting temporal features based on the LSTM algorithm and extracting spatial features based on the CNN algorithm. In this study, two advanced CNN models named VGG-16 and ResNet-50 were used on a dataset to train for the identification of surgical tools and phases. The VGG-16 model consists of five convolutional blocks and three fully connected layers that were replaced with new layers for tool and surgical phase detection. The ResNet-50 model underwent similar modifications. For

image representation, 4096-dimensional or 2048-dimensional feature vectors from the VGG-16 and ResNet-50 models were used as feature extractors and then forwarded to the LSTM model. Endoscopic videos contain both static and sequential information. Two LSTM units named LSTM-clip and LSTM-video were introduced to capture the presence of tools in sequential images. LSTM-clip predicts the presence of tools in the target frame based on a sequence of unlabeled frames. Training videos are divided into short clips, feature vectors are extracted using CNN for each clip, and then sent to LSTM-clip. Subsequently, VLSTM-clip with dimension C is generated for tool classification. LSTM-video is utilized to incorporate sequential information across consecutive clips of a video. The feature vector sequences produced by LSTM-clip are input into the LSTM-video model, which has a fully connected layer for tool classification.

3.5 Evaluation Metrics

The evaluation metrics used in this article include:

The Average Precision (AP) metric was used as an evaluation metric for the tool classification task. This metric is defined as the area under the precision-recall curve and can be calculated based on formula 1 [16]:

$$Ap_t = \int_0^t P_t \cdot d[R_t] \quad (1)$$

where P_t is the precision and R_t is the recall of tool t as a function of confidence threshold. Other metrics such as accuracy, precision, recall, and ROC curve are also provided.

4. Results

This section presents the experimental results. First, the evaluation process is described, followed by the results. In this article, the results are presented based on multi-class scenarios, detection of surgical tool types based on different edge detection methods. For the multi-class scenario (4 classes), the detection of grasper, hook, scissor, and clipper tools is considered. In this article, to evaluate the proposed method based on the process of evaluating machine learning models, the data was divided into two categories: training and testing. The holdout method was used to split the data into training and testing sets. Therefore, 70% of the data was considered for training and 30% for testing.

Table 2 shows the results of the algorithms used based on evaluation metrics. According to Table 2, the Prewitt filter outperformed other methods in terms of performance. This indicates that the Prewitt filter extracts more useful information from the image and improves the model's performance by eliminating less useful information. In Figure 3, the ROC curve for the overall performance of the filters in tool detection is presented. It is observed that a higher area under the curve (AUC) in the ROC curve indicates better model performance. In this case, based on Table 2 and Figure 3, the Prewitt method had better performance in detecting the presence or absence of surgical tools in the image compared to other methods. The presence of a surgical tool refers to whether a specific surgical tool is present in the current frame or not, indicating a binary classification by the model. The last row in the table ranks each model, with lower numbers indicating better performance and ranking.

Table 2: Results of each edge detection method in detecting the presence or absence of objects

Metric\filter	Sharp	Sobel	Canny	Roberts	Prewitt
Precision	89.01	90.05	87.15	90.15	90.28
Recall	88.09	89.92	86.30	89.84	90.28
Accuracy	88.05	89.90	86.27	89.8	90.15
ROC AUC	0.9936	0.9954	0.9918	0.9952	0.9956
Rank	5	3	4	2	1

Table 3: Average results in detecting the types of surgical tools present in the image

Metric\filter	Sharp	Sobel	Canny	Roberts	Prewitt
Precision	54.30	43.94	48.43	43.78	46.26
Recall	59.76	44.07	38.06	30.55	47.57
Accuracy	59.76	44.07	38.06	30.55	47.57
Rank	1	4	2	5	3

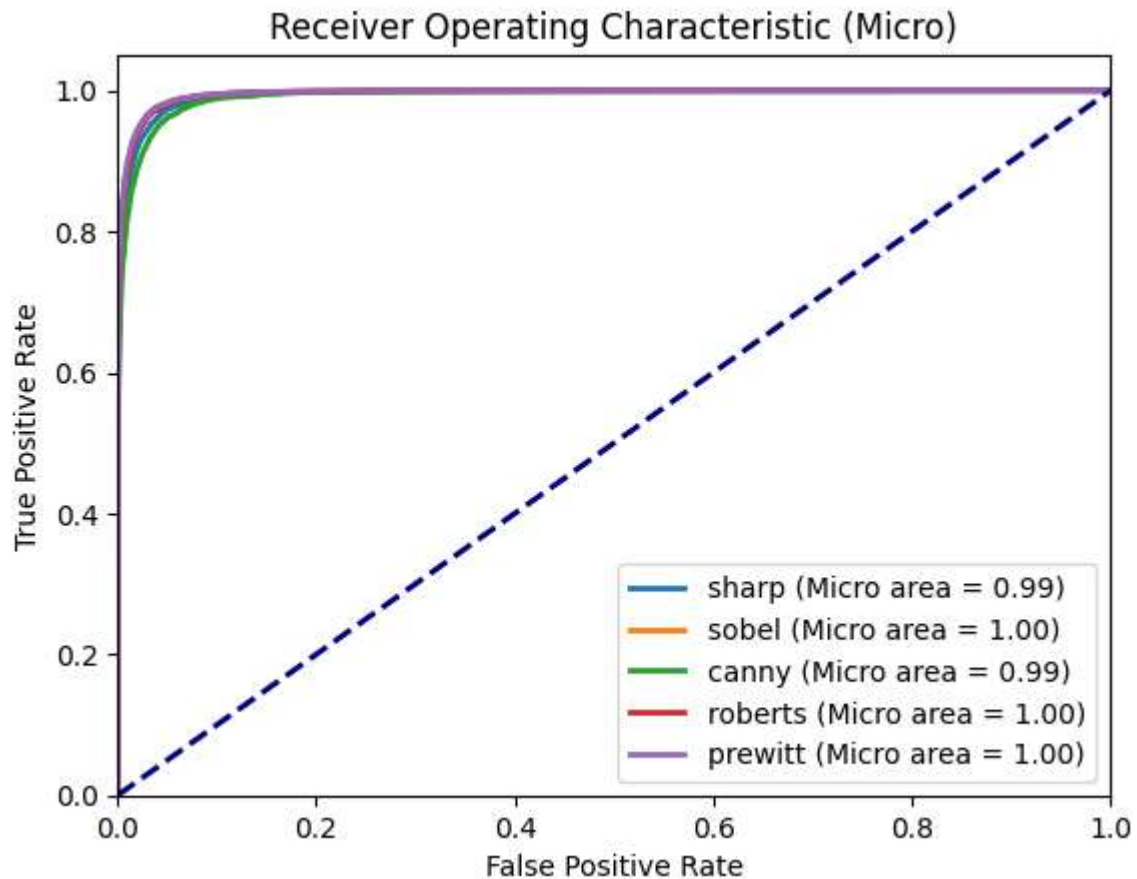


Figure 3: ROC curve for detecting surgical tools using different edge detection filters

The following results of the proposed method are presented based on the detection of surgical tool types. The results in this section are based on multi-class classification. Table 3 presents the average results for detecting surgical tools for each of the methods. In this case, the Sharp method showed better performance compared to other methods. Figures 4 to 8 show ROC curves for the Sharp, Sobel, Canny, Roberts, and Prewitt filters, respectively. A higher area under the ROC curve indicates better performance.

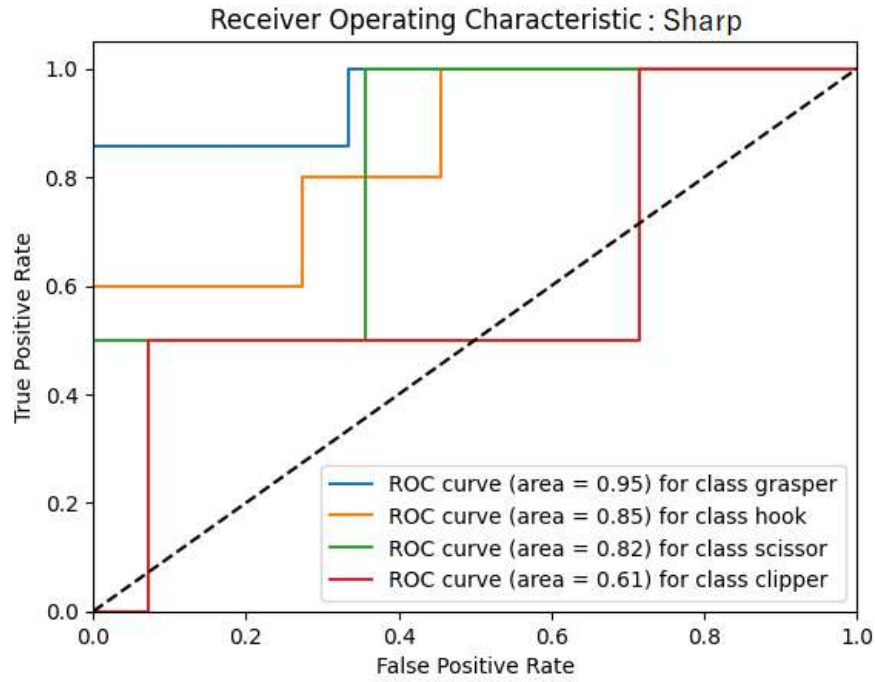


Figure 4: ROC curve for detecting the type of surgical instrument using the Sharp filter

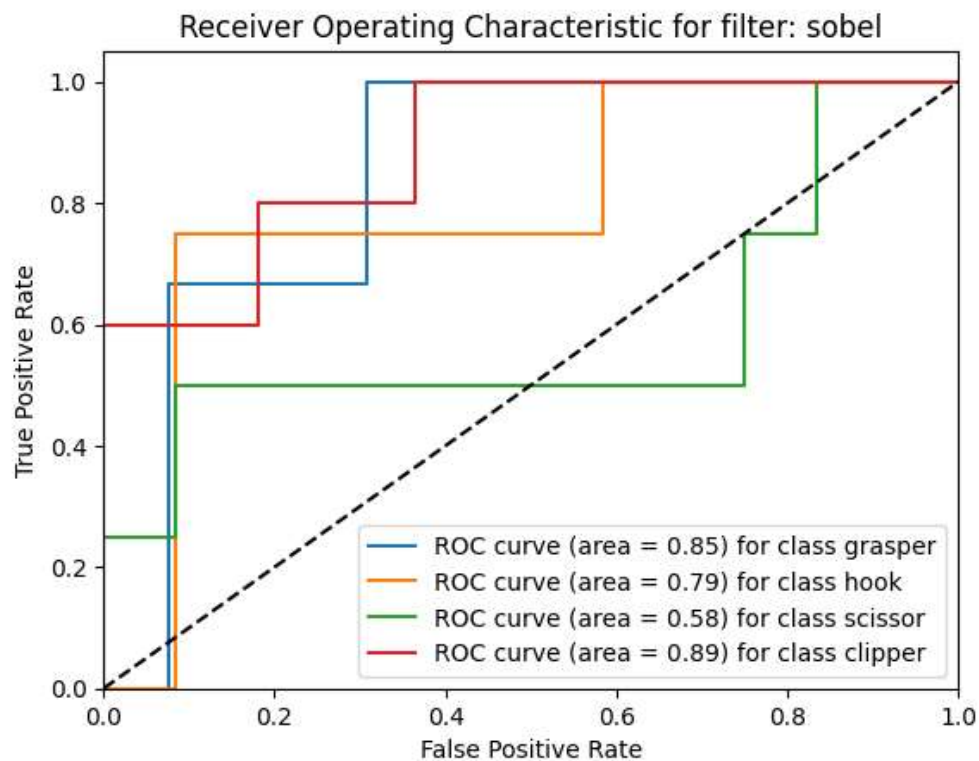


Figure 5: ROC curve for detecting the type of surgical instrument using the Sobel filter

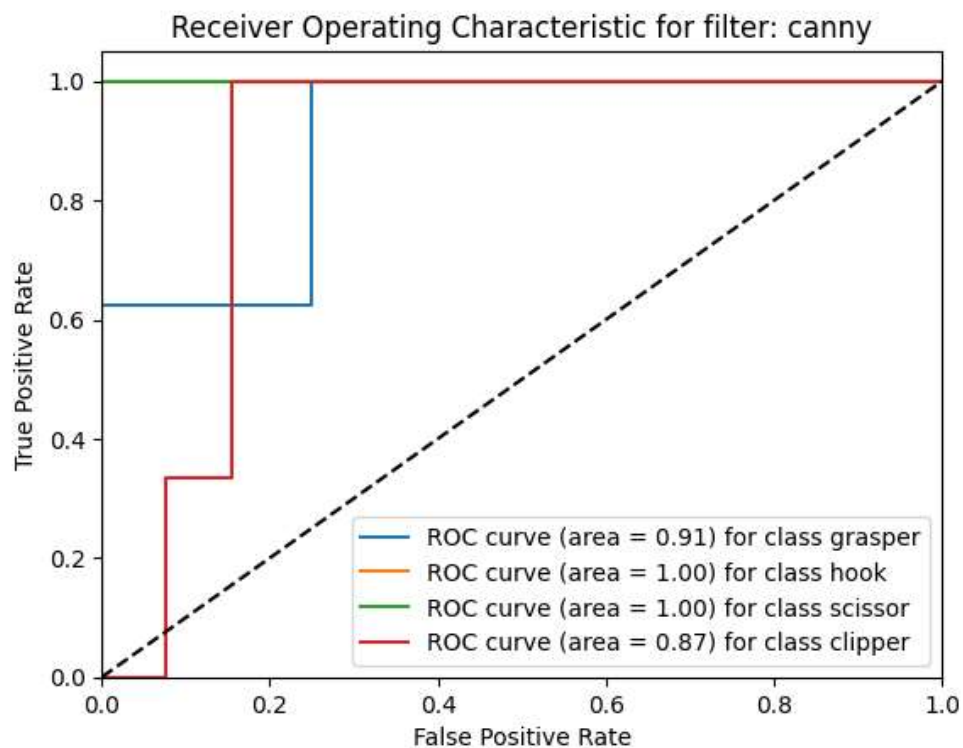


Figure 6: ROC curve for detecting the type of surgical instrument using the Canny filter

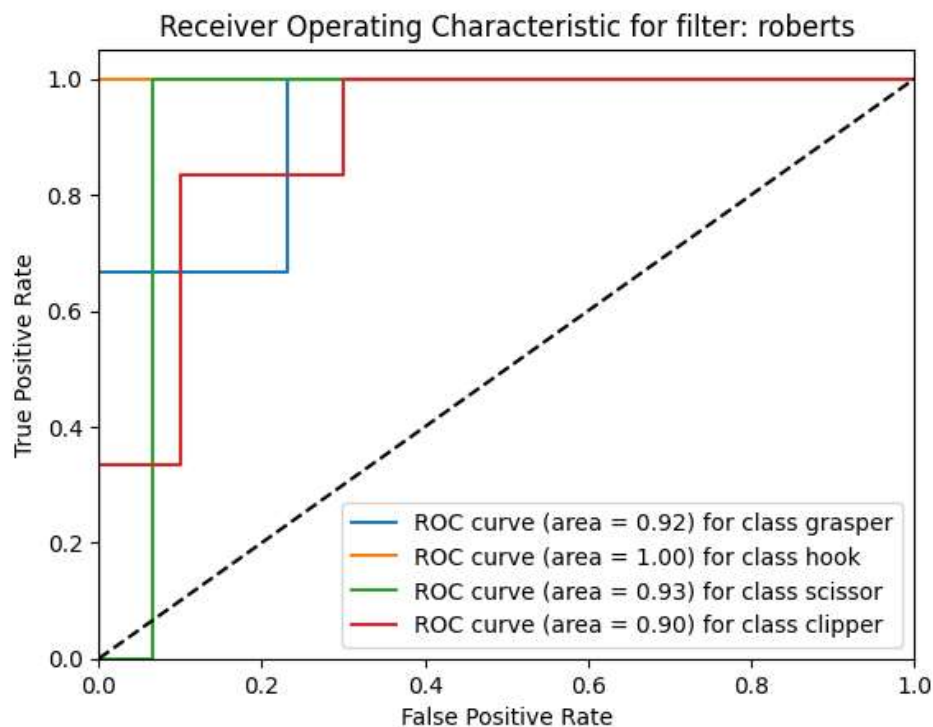


Figure 7: ROC curve for detecting the type of surgical instrument using the Roberts filter

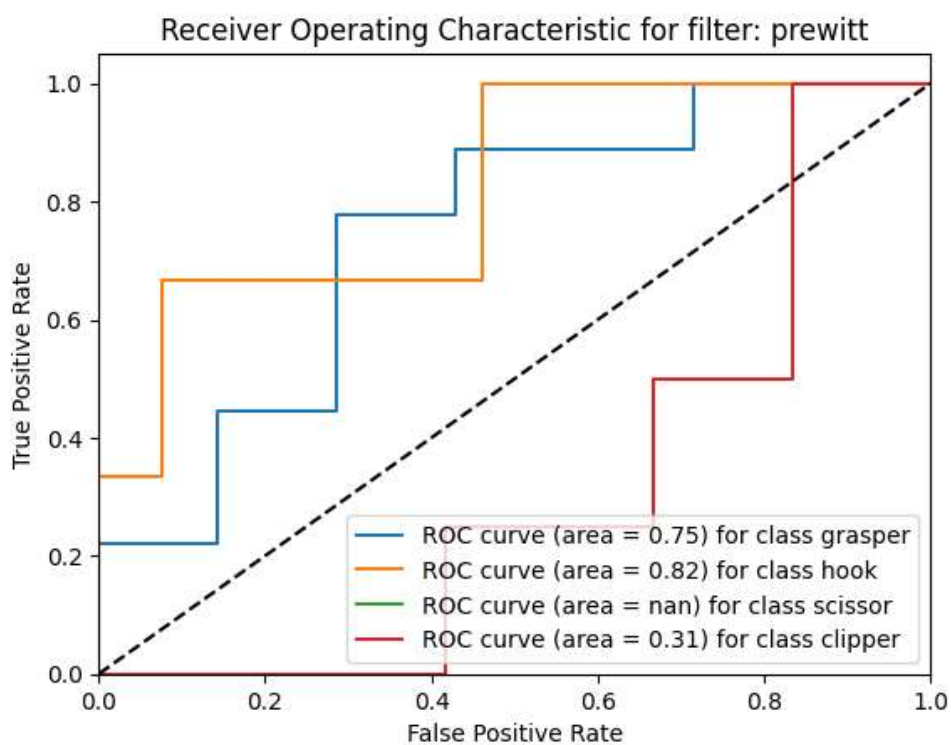


Figure 8: ROC curve for detecting the type of surgical instrument using the Prewitt filter

For a better understanding of the AUC value for each class of images and the filter used in Table 4, the results related to the AUC of filters for each class are provided. In the row and column labeled "average" in Table 4, the average results for each class (average by column) and the average results for each filter (average by row) are presented.

5. Results and Discussion

This paper focuses on the detection of surgical tools and their types using various edge detection filters combined with the model proposed in [16]. Therefore, the idea and innovation of the proposed method lie in utilizing edge detection filters to enhance important features and eliminate irrelevant information from the image. This helps the proposed method to detect tools more efficiently. Based on Table 2 and Figure 3, the proposed method using different edge detection filters was able to accurately detect the presence or absence of surgical tools in the image. Among the edge detection filters, the Prewitt filter showed better performance in detecting the presence of surgical tools in the image compared to other filters. Following the Prewitt filter, the Roberts, Sobel, Canny, and Sharp filters were ranked accordingly.

According to Table 3 and Figures 4 to 8, which represent the results for detecting the type of surgical tools, the results show relatively good performance. Among them, the Sharp method demonstrated better performance in detecting the type of surgical instrument compared to other filters. Following the Sharp filters, the Canny, Prewitt, Sobel, and Roberts filters were placed in subsequent positions. An interesting point is that in detecting the presence or absence of tools in the image, the Sharp method had the worst performance, and the Prewitt method, which initially

Table 4: AUC value for each tool based on the filter used

Filter\Class	Grasper	Hook	Scissor	Clipper	Average	Rank
Sharp	0.95	0.85	0.82	0.61	0.807	3
Sobel	0.85	0.79	0.58	0.89	0.775	4
Canny	0.91	1	1	0.87	0.945	1
Robert	0.92	1	0.93	0.90	0.937	2
Prewitt	0.75	0.82	Nan (=0)	0.31	0.465	5
Average	0.876	0.892	0.667	0.726		
Rank	2	1	4	3		

detects the presence of tools in the image, is in third place in this case. In general, the results indicate that the use of different filters, depending on the type of information they capture, directly affects performance.

Based on the overall results of the two general experiments conducted, it can be said that the Prewitt filter preserves general image information, which can be suitable for image classification, while the Sharp filter provides more detailed information about image details, which can lead to better performance in detecting the type of objects present in the image. This conclusion can be extended to other filters as well.

Source

- [1] Rana, M., & Bhushan, M. (2023). Machine learning and deep learning approach for medical image analysis: diagnosis to detection. *Multimedia Tools and Applications*, 82(17), 26731-26769.
- [2] Fujii, R., Hachiuma, R., Kajita, H., & Saito, H. (2022). Surgical tool detection in open surgery videos. *Applied Sciences*, 12(20), 10473.
- [3] Vila-Blanco, N., Carreira, M. J., Varas-Quintana, P., Balsa-Castro, C., & Tomas, I. (2020). Deep neural networks for chronological age estimation from OPG images. *IEEE transactions on medical imaging*, 39(7), 2374-2384.
- [4] Mansoor Ali, Gilberto Ochoa-Ruiz, and Sharib Ali. A semisupervised Teacher-Student framework for surgical tool detection and localization. *CMBBE*, 2022.
- [5] Ryo Fujii, Ryo Hachiuma, and Hideo Saito. Weakly SemiSupervised Tool Detection in Minimally Invasive Surgery Videos. In *ICASSP*, 2024.
- [6] V Velanovich. Laparoscopic vs open surgery. *Surgical Endoscopy*, 14(1):16–21, 1 2000. ISSN 0930-2794. doi: 10.1007/s004649900003.
- [7] Garth H Ballantyne. The pitfalls of laparoscopic surgery: challenges for robotics and telerobotic surgery. *Surgical laparoscopy, endoscopy & percutaneous techniques*, 12(1):1–5, 2 2002. ISSN 1530-4515.
- [8] Maria Antico, Fumio Sasazawa, Liao Wu, Anjali Jaiprakash, Jonathan Roberts, Ross Crawford, Ajay K. Pandey, and Davide Fontanarosa. Ultrasound guidance in minimally invasive robotic procedures. *Medical Image Analysis*, 54:149–167, 5 2019. ISSN 1361-8415. doi: 10.1016/J.MEDIA.2019.01.002.
- [9] M. Allan, S. Ourselin, D. J. Hawkes, J. D. Kelly, and D. Stoyanov. 3-D Pose Estimation of Articulated Instruments in Robotic Minimally Invasive Surgery. *IEEE Transactions on Medical Imaging*, 37 (5):1204–1213, 5 2018. ISSN 0278-0062. doi: 10.1109/TMI.2018.2794439.
- [10] N.A. Jalal, T.A. Alshirbaji, K. Moller, " Predicting surgical phases using CNN-NARX neural network, *Curr. Dir. Biomed. Eng.* 5 (1) (2019) 405–407.
- [11] N. Padoy, T. Blum, S.A. Ahmadi, H. Feussner, M.O. Berger, N. Navab, Statistical modeling and recognition of surgical workflow, *Med. Image Anal.* 16 (3) (2012) 632–641.
- [12] O. Dergachyova, D. Bouget, A. Hualme, X. Morandi, P. Jannin, Automatic datadriven real-time segmentation and recognition of surgical workflow, *Int. J. Comput. Assist. Radiol. Surg.* 11 (6) (2016) 1081–1089.
- [13] Attia, M., Hossny, M., Nahavandi, S., & Asadi, H. (2017, October). Surgical tool segmentation using a hybrid deep CNN-RNN auto encoder-decoder. In *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (pp. 3373-3378). IEEE.

- [14] Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(12), 2481-2495.
- [15] Hasan, M. K., Calvet, L., Rabbani, N., & Bartoli, A. (2021). Detection, segmentation, and 3D pose estimation of surgical tools using convolutional neural networks and algebraic geometry. *Medical Image Analysis*, 70, 101994.
- [16] Alshirbaji, T. A., Jalal, N. A., Docherty, P. D., Neumuth, T., & Möller, K. (2021). A deep learning spatial-temporal framework for detecting surgical tools in laparoscopic videos. *Biomedical Signal Processing and Control*, 68, 102801.
- [17] Ishii, L. E., Tollefson, T. T., Basura, G. J., Rosenfeld, R. M., Abramson, P. J., Chalet, S. R., ... & Nnacheta, L. C. (2017). Clinical practice guideline: improving nasal form and function after rhinoplasty. *Otolaryngology–Head and Neck Surgery*, 156, S1-S30.
- [18] Yi, H., Ke, Y., Zhu, H., Lu, S., Cheng, L., & Liu, M. (2021). Hidden Dangers and Preventive Measures in the Inventory of Articles in Operating Room. *Open Journal of Preventive Medicine*, 11(5), 192-198.
- [19] Sanchez, J. A., Barach, P., Johnson, J. K., & Jacobs, J. P. (2017). Surgical patient care. *Cham: Springer International Publishing*, 683-694.
- [20] Profetto, L., Gherardelli, M., & Iadanza, E. (2022). Radio Frequency Identification (RFID) in health care: where are we? A scoping review. *Health and technology*, 12(5), 879-891.
- [21] Lee, J. D., Chien, J. C., Hsu, Y. T., & Wu, C. T. (2021). Automatic surgical instrument recognition—a case of comparison study between the faster R-CNN, mask R-CNN, and single-shot multi-box detectors. *Applied Sciences*, 11(17), 8097.
- [22] Wang, S., Raju, A., & Huang, J. (2017, April). Deep learning based multi-label classification for surgical tool presence detection in laparoscopic videos. In *2017 IEEE 14th international symposium on biomedical imaging (ISBI 2017)* (pp. 620-623). IEEE.
- [23] Wang, Y., Sun, Q., Sun, G., Gu, L., & Liu, Z. (2021, July). Object detection of surgical instruments based on Yolov4. In *2021 6th IEEE International Conference on Advanced Robotics and Mechatronics (ICARM)* (pp. 578-581). IEEE.
- [24] Zhou, Y., & Liu, Z. (2022, August). Detection of surgical instruments based on YOLOv5. In *2022 IEEE International Conference on Manipulation, Manufacturing and Measurement on the Nanoscale (3M-NANO)* (pp. 470-473). IEEE.
- [25] Liu, K., Zhao, Z., Shi, P., Li, F., & Song, H. (2022). Real-time surgical tool detection in computer-aided surgery based on enhanced feature-fusion convolutional neural network. *Journal of Computational Design and Engineering*, 9(3), 1123-1134.
- [26] Jin, A., Yeung, S., Jopling, J., Krause, J., Azagury, D., Milstein, A., & Fei-Fei, L. (2018, March). Tool detection and operative skill assessment in surgical videos using region-based

convolutional neural networks. In *2018 IEEE winter conference on applications of computer vision (WACV)* (pp. 691-699). IEEE.

[27] Kurmann, T., Marquez Neila, P., Du, X., Fua, P., Stoyanov, D., Wolf, S., & Sznitman, R. (2017). Simultaneous recognition and pose estimation of instruments in minimally invasive surgery. In *Medical Image Computing and Computer-Assisted Intervention– MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part II 20* (pp. 505-513). Springer International Publishing.

[28] Twinanda, A. P., Shehata, S., Mutter, D., Marescaux, J., De Mathelin, M., & Padoy, N. (2016). Endonet: a deep architecture for recognition tasks on laparoscopic videos. *IEEE transactions on medical imaging*, 36(1), 86-97.

[29] Jin, A., Yeung, S., Jopling, J., Krause, J., Azagury, D., Milstein, A., & Fei-Fei, L. (2018, March). Tool detection and operative skill assessment in surgical videos using region-based convolutional neural networks. In *2018 IEEE winter conference on applications of computer vision (WACV)* (pp. 691-699). IEEE.

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

[31] Nwoye, C. I. (2021). *Deep learning methods for the detection and recognition of surgical tools and activities in laparoscopic videos* (Doctoral dissertation, Université de Strasbourg).

[32] Bhunia, A. K., Konwer, A., Bhunia, A. K., Bhowmick, A., Roy, P. P., & Pal, U. (2019). Script identification in natural scene image and video frames using an attention based convolutional-LSTM network. *Pattern Recognition*, 85, 172-184.

[33] Jin, Y., Li, H., Dou, Q., Chen, H., Qin, J., Fu, C. W., & Heng, P. A. (2020). Multi-task recurrent convolutional network with correlation loss for surgical video analysis. *Medical image analysis*, 59, 101572.

[34] Al Hajj, H. (2018). *Video analysis for augmented cataract surgery* (Doctoral dissertation, Université de Bretagne occidentale-Brest).

[35] Wang, X., Zhu, M., Bo, D., Cui, P., Shi, C., & Pei, J. (2020, August). Am-gen: Adaptive multi-channel graph convolutional networks. In *Proceedings of the 26th ACM SIGKDD International conference on knowledge discovery & data mining* (pp. 1243-1253).

[36] Bajraktari, F., Fleissner, K., & Pott, P. (2023). A deep learning based instrument detection approach for automated surgical systems. *Proceedings on Automation in Medical Engineering*, 2(1), 749-749.

[37] Ran, B., Huang, B., Liang, S., & Hou, Y. (2023). Surgical instrument detection algorithm based on improved YOLOv7x. *Sensors*, 23(11), 5037.

[38] Zheng, L., & Liu, Z. (2022, December). Real Time Surgical Instrument Object Detection Using YOLOv7. In *International Conference on Robotics and Networks* (pp. 81-90). Cham: Springer International Publishing.

[39] Bai, J. (2023). Deep learning-based intraoperative video analysis for supporting surgery. *Concurrency and Computation: Practice and Experience*, 35(26), e7837.