Towards Real-Time Multi-Class Object Detection and Tracking for the FLS Pattern Cutting Task

Koloud N. Alkhamaiseh[*,1], Janos L. Grantner[2], Ikhlas Abdel-Qader[2], Saad Shebrain[3]

[1]Department of Computer Science, Michigan Technological University, Houghton, 49931, MI, USA

[2]Department of Electrical and Computer Engineering, Western Michigan University, Kalamazoo, 49008, MI, USA

[3]Western Michigan University Homer Stryker MD School of Medicine, Kalamazoo, 49008, MI, USA

A R T I C L E   I N F O

A B S T R A C T

The advent of laparoscopic surgery has increased the need to incorporate simulator-based training into traditional training programs to improve resident training and feedback. However, current training methods rely on expert surgeons to evaluate the dexterity of trainees, a time-consuming and subjective process. Through this research, we aim to extend the use of object detection in laparoscopic training by detecting and tracking surgical tools and objects. In this project, we trained YOLOv7 object detection neural networks on Fundamentals of Laparoscopic Surgery pattern-cutting exercise videos using a trainable bag of freebies. Experiments show that YOLOv7 has a mAP score of 95.2, 95.3 precision, 94.1 Recall, and 78 accuracy for bounding boxes on a limited-size training dataset. This research clearly demonstrates the potential of using YOLOv7 as a single-stage real-time object detector in automated tool motion analysis for the assessment of the resident's performance during training.

1. Introduction

This paper is an extension of work originally presented in DICTA 2021 [1]. In this project, more data is collected and prepared to train YOLOv7 [2] as a real-time object detector of laparoscopic tools and objects in the Fundamentals of Laparoscopic Surgery (FLS) pattern-cutting exercise using a box trainer [3]. Laparoscopic procedures have become increasingly popular in operating rooms worldwide due to their numerous benefits, leading medical schools to incorporate this technique into their surgery curricula [4]. However, the one-on-one apprenticeship model is subjective and time-consuming. To address this issue, laparoscopic trainers and simulators have become well-accepted alternatives that allow for safe and harm-free training [5]. Although simulation systems offer objective measurements and remote training, expert surgeons are still required to assess surgical skills proficiency. Virtual Reality (VR) training provides a completely virtual environment with haptic feedback and complex software, but it is expensive and requires highly sophisticated mechanical design [6]. By improving the real-time object detection of laparoscopic tools and objects, this project aims to enhance the effectiveness and accessibility of laparoscopic training.

Box trainers and physical trainers provide a practical environment for using real laparoscopic instruments to improve basic skills such as knotting, handling objects, and cutting tissues. While time is currently the primary metric for evaluating a surgeon's performance using statistical tools, studies have shown that box trainers enhance trainee confidence and dexterity [7], [8]. However, objective assessments of laparoscopic skills still require experienced surgeon evaluations. To address this issue, hybrid trainers combine the benefits of simulators and physical trainers to recreate real-world conditions and provide objective assessments through integrated software. Hybrid trainers provide a comprehensive approach to evaluating real-world situations by merging both simulators and physical trainers.

The emergence of deep learning [9] has proven to be a highly effective machine learning approach for detecting and classifying objects from raw data by learning representations from the data. Its superior feature extraction and expression capability has surpassed other machine learning methods in many areas, especially when dealing with large data sets. Therefore, deep learning appears to be a very promising method for detecting tool presence [9].

*Corresponding Author: Koloud N Alkhamaiseh, +12695449311 & kalkhama@mtu.edu

The main contributions of this research are as follows:

- Contribution to the creation of the first laparoscopic box trainer custom dataset, i.e., the WMU's Laparoscopic Box-Trainer Dataset [10]. This custom dataset was developed through a research collaboration between Western Michigan University's Department of Electrical and Computer Engineering and the Department of General Surgery at Homer Stryker M.D. School of Medicine. Researchers are free to download the dataset at their convenience. This dataset was created specifically to aid research in the field of Laparoscopic Surgery Skill Assessment. It consists of videos showcasing four different tasks on the Laparoscopic Box-Trainer - two precision cutting tests, intracorporeal suturing, and peg transfer. These videos were recorded by surgeons, surgical residents, and OB/GYN residents at the Intelligent Fuzzy Controllers Laboratory at WMU. You can access the dataset at https://drive.google.com/drive/folders/1F97CvN3GnLj-rqg1tk2rHu8x0J740DpC. The dataset also contains labeled images with related labels for all tasks, and more files will be added as the research progresses.

- Proposing a robust real-time multi-class object detection and tracking module based on YOLOv7 as a single-stage real-time object detection neural network for surgical tools and objects in FLS pattern cutting test.

The adoption of YOLOv7 is a wise choice because of its superior network architecture, precise object detection, efficient label assignment, and resilient loss function and model training. Moreover, YOLOv7 is more cost-effective than other deep learning models [2] and is highly proficient in detecting and tracking surgical tools and objects within spatial boundaries.

In order to fully explain our proposed system, this paper is organized as follows: Section 2 gives a brief introduction to the methods used for evaluating the performance of laparoscopic surgery training. In Section 3, we present our methodology. Section 4 contains a summary of our experimental findings, and in Section 5, we outline our plans for future work.

2. Background

FLS tool is widely used for psychomotor skill training in surgery. The American College of Surgeons (ACS) has created didactic instructions and manual skills to improve the basic laparoscopic surgery skills of surgical residents and practicing surgeons using the FLS box trainer [3].

The FLS box trainer, along with didactic instructions and manual skills, can help surgical residents and practicing surgeons improve their basic surgical skills. Current assessment methods focus on detecting surgical tools and analyzing motion, but it's also crucial to track surgical instruments during operations or training to analyze operations and assess training.

In a previous study [5], computer vision algorithms were used to assess performance during surgical tool detection, categorization, and tracking in real-time FLS surgical videos. An artificial neural network learned from expert and non-expert behaviors and a web-based tool was created for uploading MIS training videos securely and receiving evaluation scores with trainee performance analysis over time. The assessment used a multi-dimensional vector consisting of smoothness of motion, proficiency of surgical gestures, and number of errors.

Another study [6] presented a trainer for assessing laparoscopic surgical skills using computer vision, augmented reality, and AI algorithms on a Raspberry Pi programmed in Python. The assessment method employs an artificial neural network based on a predetermined threshold for the peg transfer task. A simulation of pattern-cutting was used to track laparoscopic instruments, while computer vision libraries counted the number of transferred points during the transfer task.

Recent advancements in deep learning, specifically CNN networks, have shown remarkable progress in computer vision tasks [11]. Various studies have implemented deep learning architectures to detect surgical tool presence in laparoscopic videos [12], [13], [14], and phase recognition [15].

Some projects have created systems that can detect laparoscopic instruments in real-time during robotic surgery, using the real-time detection algorithm of the CNN network. These systems are based on the object detection systems YOLO [16] and YOLO9000 [17], with a mean average precision of 84.7 for all tools. They also have a speed of 38 frames per second (FPS).

The available skill assessment frameworks have some limitations when it comes to evaluating fundamental laparoscopic skills based on globally accepted standards and criteria. These frameworks only focus on tool motion and do not consider surgical objects and their manipulation during training. For instance, the Objective Structured Assessment of Technical Skills (OSATS) [16] and the Global Operational Assessment of Laparoscopic Skills (GOALS) [17] are examples of such frameworks. To address this limitation, a new system was developed using a deep learning algorithm called YOLACT [18]. This system tracks surgical tool motion and detects surgical objects, including their deformability, shapes, and geometries in the surgical field of view. The system was tested on a modified FLS peg transfer exercise and provided a more comprehensive evaluation of laparoscopic skills beyond tool motion alone.

To evaluate the abilities of those performing intricate intracorporeal suturing, automated systems have been suggested. One such system, developed by the authors in [19], uses the latest versions of One-Stage-Object-Detectors like YOLOv4, Scaled-YOLOv4, YOLOR, and YOLOX. A dataset of suturing tasks was used to train this system, which strikes a balance between cutting-edge architectures. In [20], the authors proposed a skill evaluation system that employs Scaled-YOLOv4 and a centroid tracking algorithm.

The authors presented a fuzzy logic supervisor system for assessing surgical skills in [21]. This system used multi-class detection and tracking of laparoscopic instruments during standard FLS pattern-cutting tests. However, the system had limitations when the instruments or objects were not within view of the camera. To address this issue, a new autonomous evaluation system was proposed by authors in [22], which utilized two cameras and multi-thread video processing to detect laparoscopic instruments.

In addition, two fuzzy logic systems were implemented in parallel to evaluate left and right-hand movement. The authors in [23] have improved the YOLOV7x algorithm significantly for detecting surgical instruments. These enhancements effectively address concerns about dense arrangements, mutual occlusion, difficulty in distinguishing similar instruments, and varying lighting conditions.

To provide a more comprehensive assessment of surgical quality, our approach includes examining circle shape deformability, as well as laparoscopic tool tracking and detection in box trainer pattern cut test recorded videos.

3. Methodology

The FLS box trainer boasts multiple exercises, but we've focused on perfecting a pattern-cutting exercise. Our system centers around two circles printed on artificial tissue, with a radius of 2.5 and 3.0 centimeters for the inner and outer circles respectively Figure 1. Keeping the scissors within these circles is essential for this exercise, as crossing either circumference will result in an incorrect cut.

- In this paper, we propose a system that works seamlessly with our intelligent FLS box trainer. This system utilizes two cameras placed inside the box to record videos, which are then used to train a deep-learning object detector and tracker. We have employed powerful deep learning algorithms YOLOv7, which have set a new standard in real-time object detection. Our model has been trained, validated, and tested using a custom data set to track three objects - the scissors, the clipper, and the circle - within the box trainer's surgical view. The circle is the object that needs to be cut out, and this is achieved using laparoscopic scissors and clippers.
- Figure 2 illustrates the complete workflow of our method. First, we divide the recorded videos into frames and preprocess them before training the YOLOv7 network. After successfully training the model, it can accurately detect and track intended objects in both tested videos and real-time videos. Lastly, we calculate the performance evaluation parameters and generate an output video that displays the labeled surgical objects.

3.1. Network Architecture

The YOLO family of models has a long-standing association with the Darknet framework [24], tracing back to its inception in 2015 [25].
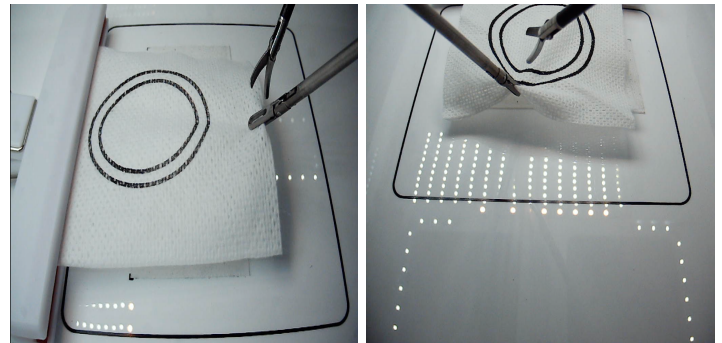

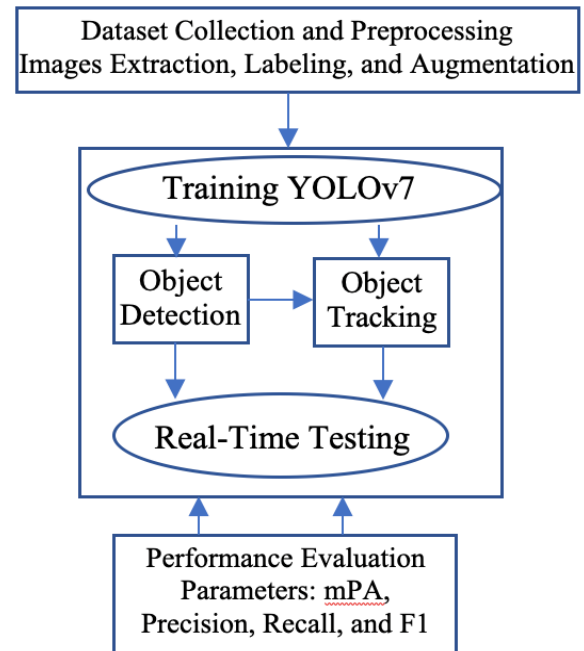Figure 1: FLS Pattern cutting test setup views as captured by the two cameras.

Figure 2: The proposed workflow chart.

The YOLO (You Only Look Once) detection layers rely on regression and classification optimizers to determine the necessary number of anchors. The image is divided into cells using a 19x19 grid, where each cell can predict up to five bounding boxes. However, some of these cells and boxes may not contain an object, so a probability of object presence (PC) is utilized to remove low-probability bounding boxes. Non-max suppression is subsequently used to select the bounding boxes with the highest shared area. YOLO has many versions and variants that enhance performance and efficiency.

The most recent official version, YOLOv7, was created by the original authors of the architecture. It is a single-stage real-time object detector and, according to the YOLOv7 paper [2], it is currently the fastest and most accurate real-time object detector available.

- E-ELAN (Extended Efficient Layer Aggregation Network): It is the computational block in the YOLOv7 backbone, in which the network learns faster by expanding, shuffling, and merging cards to continuously improve its ability to learn without destroying its gradient path.
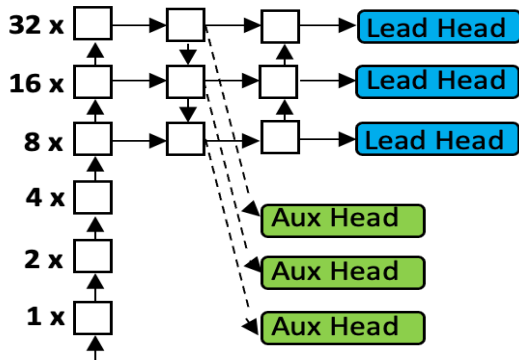
- Concatenation-based model scaling allows the model to maintain the properties that it had at the initial design and thus maintain the optimal structure.
- To replace the convolutional layer or residual with re-parameterized convolution, planned re-parameterized convolution without an identity connection uses RepConv.
- Coarse for auxiliary and Fine for lead loss: As YOLOV7 includes multiple heads, the Lead Head is responsible for the final output, and the Auxiliary Head is used to train in the middle layers as illustrated in Figure 3 (a). Loss assists with updating the weights of these heads, allowing for Deep Supervision and better model learning. A Label Assigner mechanism was introduced to enhance deep network training, which considers the network prediction results and ground truth before assigning soft labels as shown in Figure 3 (b). Unlike traditional label assignment methods that generate hard labels based on given rules by directly referring to the ground truth, reliable soft labels use calculation and optimization methods that also consider the quality and distribution of prediction output together with the ground truth.

YOLOv7 introduces important reforms that significantly improve real-time object detection accuracy while keeping inference costs low. Compared to state-of-the-art real-time object detections, YOLOv7 reduces parameter and computation costs by about 40% and 50%, respectively, resulting in faster inference speeds and higher detection accuracy [2]. YOLOv7 has a fast and robust network architecture that integrates features, provides better object detection performance, and employs an efficient model training process with a robust loss function and label assignment. Overall, YOLOv7 represents the best option for optimizing real-time object detection.

3.2. Dataset

Our laparoscopic detection and tracking project requires a dataset annotated with spatial bounds for objects and tools. To achieve this, we extracted 1572 labeled images from 13 videos ofthe FLS pattern-cutting test, recorded by an expert surgeon and residents from the School of Medicine at Western Michigan University [26]. These videos have a resolution of 640x480 pixels and a frame rate of 30 frames per second and were carefully selected to accurately depict various instrument scenarios, lighting conditions, and angles. We resized the images to 416x416 pixels with auto orientation as a preprocessing step.

Figure 4 displays a ground-truth example utilizing a free preprocessing tool from Roboflow [27]. This tool manually labels the three intended objects- circle, clipper, and scissors in each frame. To augment our dataset's sample size, we employed various techniques, such as affine transformations, rotations, cropping, shearing, hue saturation, and blurs, as depicted in Figure 5. To ensure accuracy, new pixels were filled with the average RGB value of the corresponding image, which has proven to be highly reliable [28]. Our augmentation step yielded 3,458 images, with 87% designated for training, 8% for validation, and 5% for testing. The dataset's distribution of instruments and objects is outlined in Table 1.

3.3. Evaluation Criteria

To evaluate our work, we used box loss, objectness loss, classification loss, precision, recall, and mean Average Precision(mAP) as performance metrics. Figure 6 depicts these metrics. By measuring box loss, we were able to determine the algorithm's ability to accurately locate the center of an object and ensure that its bounding box adequately covers it. Objectness measures the probability that an object exists within a proposed region, serving as a confidence metric. High objectness indicates a greater likelihood that an object will be visible in the image window. Classification loss evaluates the algorithm's ability to predict the correct object class. Ground-truth intersection over union (GIoU) refers to the overlap between the ground-truth region and the detection result region. In the context of GIoU judgment, precision is the ratio of true positives to total detections. Meanwhile, recall is the ratio of successful detections to the total number of classes. The mAP, on the other hand, displays our bounding box predictions based on various GIoU thresholds set at mAP@0.5:0.95 and mAP@0.5 on average. To obtain the final estimate, the AP value is computed for each class across all GIoU thresholds, and the mAP is averaged for all classes.

Table 1: Details of the dataset: The number of instances for each instrument type is shown, distributed over 1572 images, and the resulting augmented sample size is included.

Object type	Instant samples	Augmented sample size
Scissors	1253	3759
Clipper	1289	3867
Circle	1432	4296
Total instances	3974	11922
Images	1572	3458

(a) The addition of an auxiliary head is included in the model.
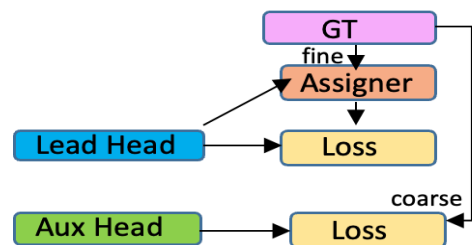
(b) Coarse-to-fine lead guide assigner.

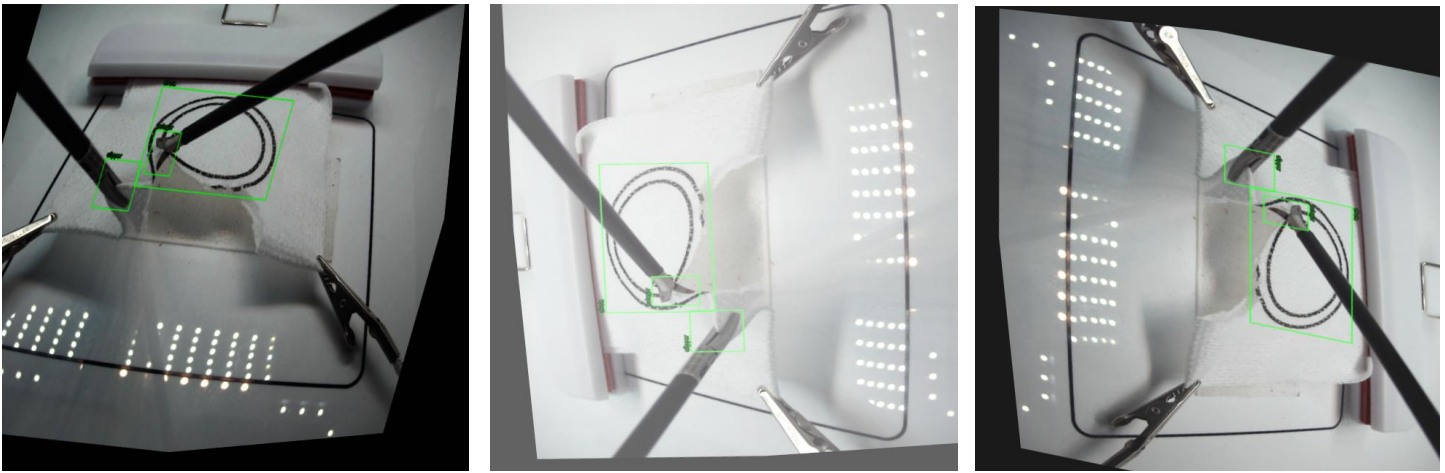Figure 3: YOLOV7 multiple heads and Label Assigner [2].

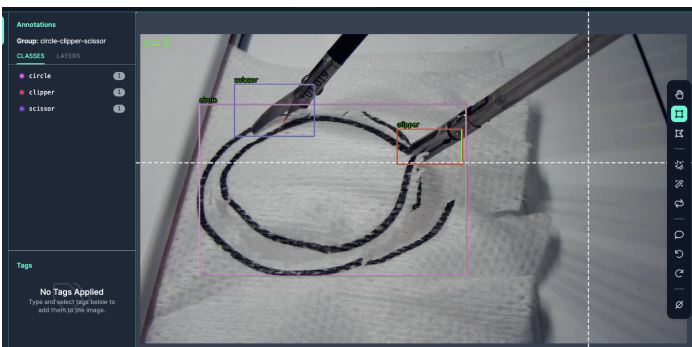Figure 5: Some sample images for preprocessing using augmentation techniques.

Figure 4: Image labeling process ground-truth example.

4. Experimental Results

This section outlines the setup used to train and evaluate CNN-based models for FLS trainer pattern-cutting test instrument recognition and tracking. We utilized a notebook developed by Roboflow.ai [29] to train YOLOv7 with our custom dataset and yolov7_training.pt pre-trained weights.

4.1 Training

To train YOLOv7, the proposed system used the Tesla P100-PCIE-16 GB GPUs with 56 processors and 16,280 MB of memory obtained from Google Colab [30]. To implement the process and validate the performance of the model scripts, several Python libraries were used (e.g., in [31] and [32]). The YOLOv7 model was trained with 416 x 416-pixel images, 16 batches, and 200 epochs. The training process took 3.357 hours with 10.6G of GPU memory.

4.1 Quantitative and Qualitative Results

In this subsection, we quantitatively and qualitatively evaluate the performance of our approach to detecting and tracking the FLS laparoscopic instruments and objects in the pattern-cutting test using YOLOv7.

The performance details during the training and validation phases are presented in Figure 6 and Table 2. Along with the losses, precision, recall, and mean average precision were calculated using GIoU thresholds of 50% and 50%:95% for up to 200 iterations. The model's precision, recall, and mean average precision exhibited rapid improvement, stabilizing after roughly 100 epochs with minor fluctuations at the start. Moreover, the validation data's classification loss decreased significantly until approximately epoch 50. To select weights, early stopping was employed.

Further, Figure 7 presents a precision-recall curve that provides a granular performance indicator for each class. The PR curve analysis indicated that the circle provided the highest performance (98.7%), followed by the clipper (95.9%), and the scissors (91.0%). It was expected since the scissors were not always clearly visible. They may be obscured by the gauze while cutting and exhibit different orientations as they move. Table 3 shows the precision, recall, and mean average precision values for each class.

Moreover, the F1 score measures the model accuracy by calculating the harmonic mean of precision and recall for the minority positive class. The harmonic mean emphasizes similar precision and recall values; the more precision and recall scores differ, the worse the harmonic mean. This score provides both recall and precision, which means that it captures both positive and negative cases. For all classes, the F1 score for the proposed model is 0.95 at a confidence level of 0.78 as shown in Figure 8.

Figure 6 showcases qualitative results for the suggested module. The top row displays the actual boxes, and the second row reveals the detection and classification outcomes obtained from YOLOv7. Despite the laparoscopic instruments and circles having deformities, varied orientations, locations, and some covering, the detection accuracy is quite high.

Moreover, this system has demonstrated an exceptional ability to detect and track targeted objects with a high degree of accuracy in pattern-cutting test videos, taking only 12.3 milliseconds per frame for processing. An example of a pattern-cutting test with active tracking and detection can be seen in Figure 9, which shows some selected frames from two videos recorded by two different cameras. The trained model can detect and track the scissors, graspers, and circles despite their varying orientations, locations, and coverage.
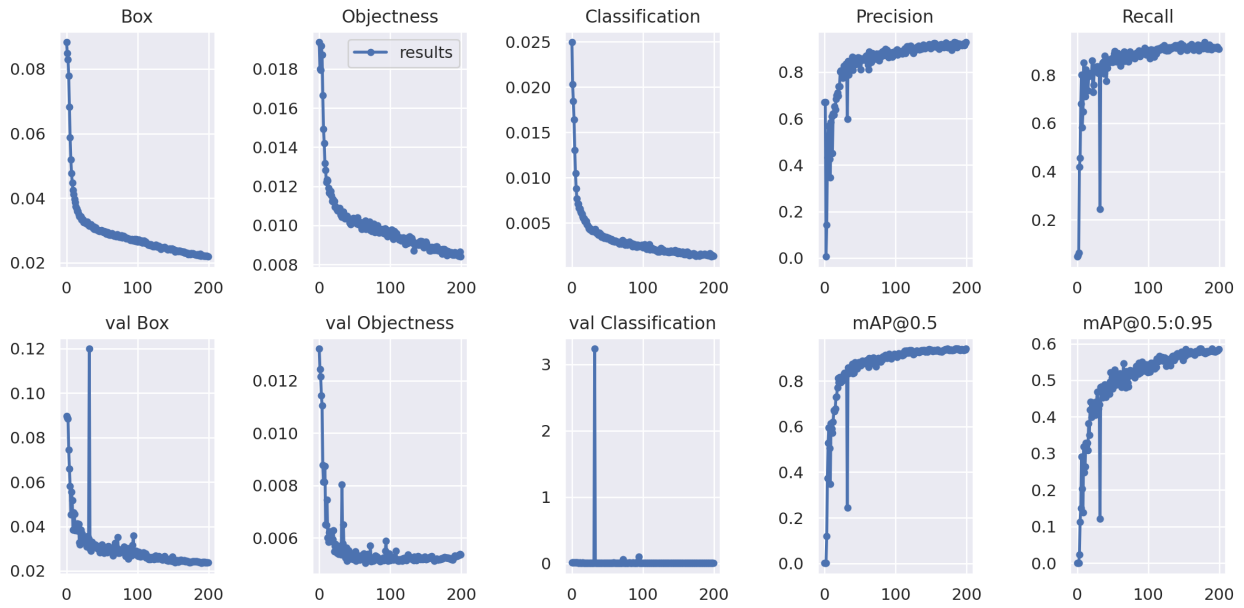
Figure 6: Plots of the box loss, objectness loss, classification loss, precision, recall, and mean average precision (mAP) for both the training and validation sets over the training epochs.

Table *2*: Losses, mean average precision, precision, and recall final value*s*.

Evaluation Criteria	Final Value
Box Loss	0.01391
Objectness	0.004788
Class. Loss	0.0003568
mAP@0.5	0.951
mAP@0.5:0.95	0.641
Precision	0.95
Recall	0.941

Table 3: Precision, recall, and mean average precision values for each class.

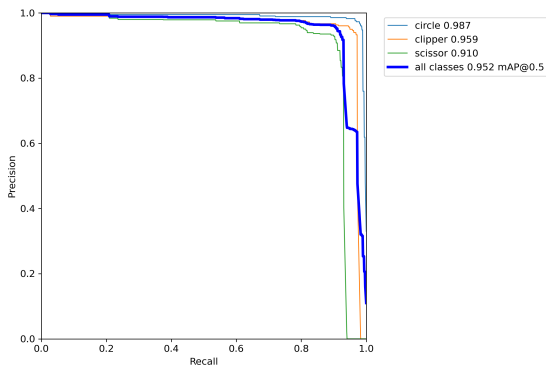| Evaluation Criteria	Circle	Clipper	Scissors
Precision	0.979	0.949	0.932
Recall	0.966	0.957	0.9
mAP@0.5	0.987	0.959	0.91
mAP@0.5:0.95	0.889	0.537	0.495

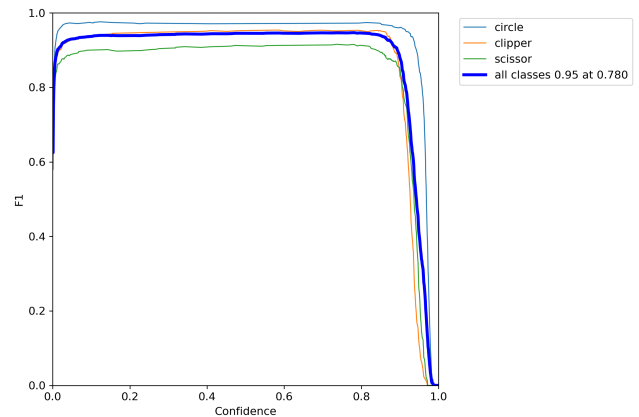Figure 7: Precision-recall (PR) curves for all classes.

Figure 8: F1 score curves for all classes.

Afterward, testing was conducted on the lab's main workstation. This workstation is equipped with a 2.5GHz Intel(R) Xeon(R) CPU E5-1650 v4 and 32.0GB of RAM. With a delay of 1.67 seconds, the model can detect and track the circle, the grasper, and the scissors. In this case, the delay was caused by the time it took to capture and process frames. To conduct real-time assessments, more powerful hardware is required.

Our study proposes a new approach for detecting and tracking laparoscopic instruments, using deep-learning neural networks.

We have compared our approach with other methods reported in the literature, summarizing the number of extracted images, labeling tools, model built, and results in Table 4. Our proposed approach outperforms previously reported models, achieving an F1 score of 0.95 at a confidence level of 0.78, and a mAP score of 95.2, 95.3 precision, 94.1 Recall, and real-time processing speed of 83.3 FPS, despite the limited number of videos. The models in previous studies may differ in construction, which could explain the differences in results. Overall, our study presents a reliable and efficient method for assessing the performance of trainers in laparoscopic instrument use.
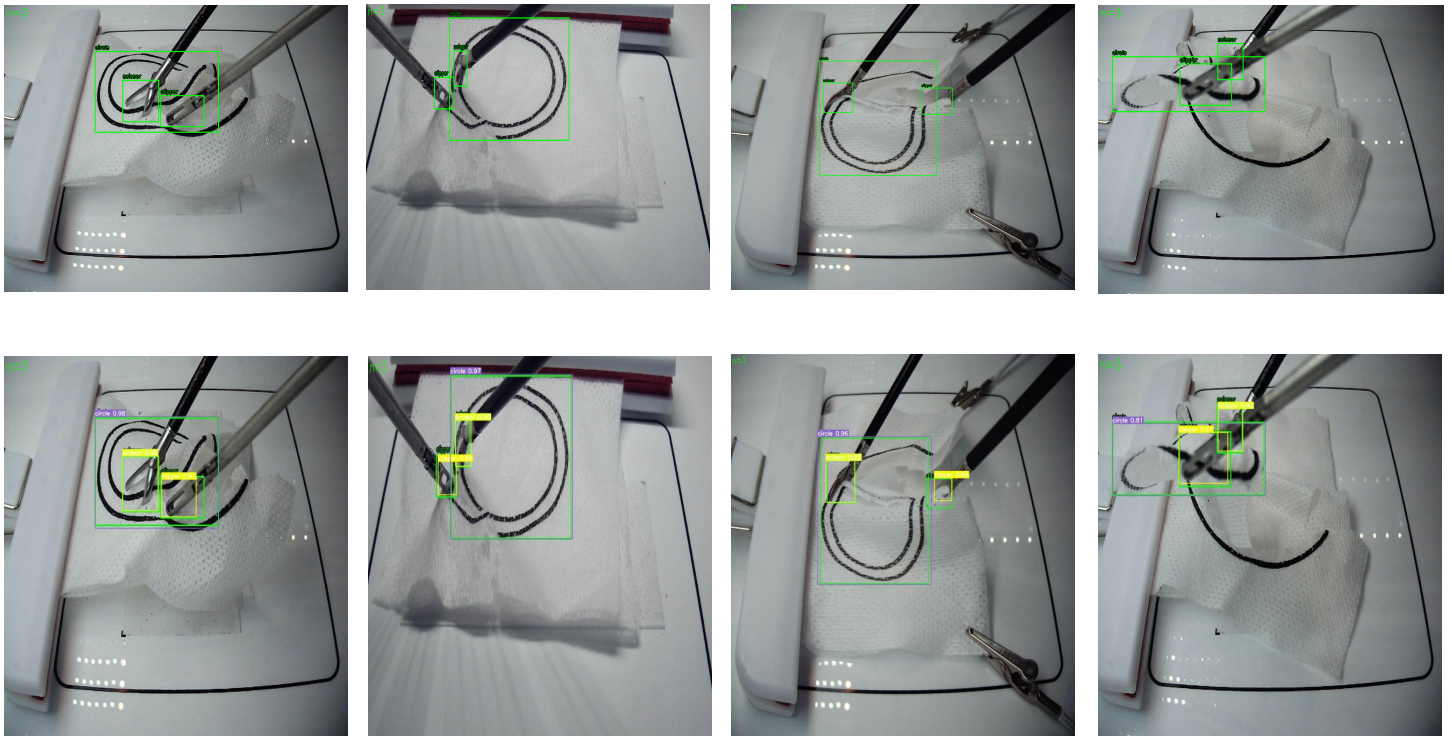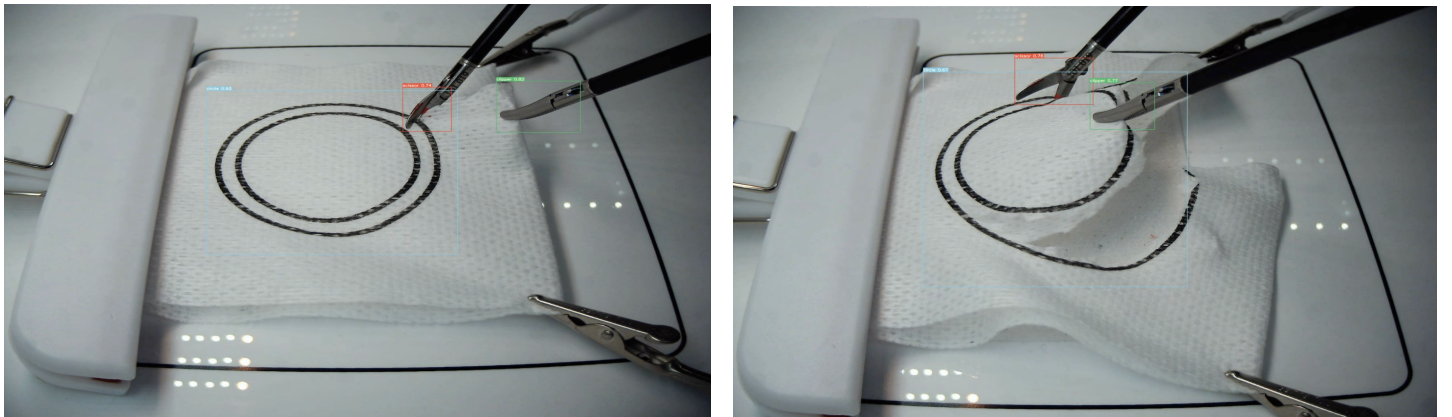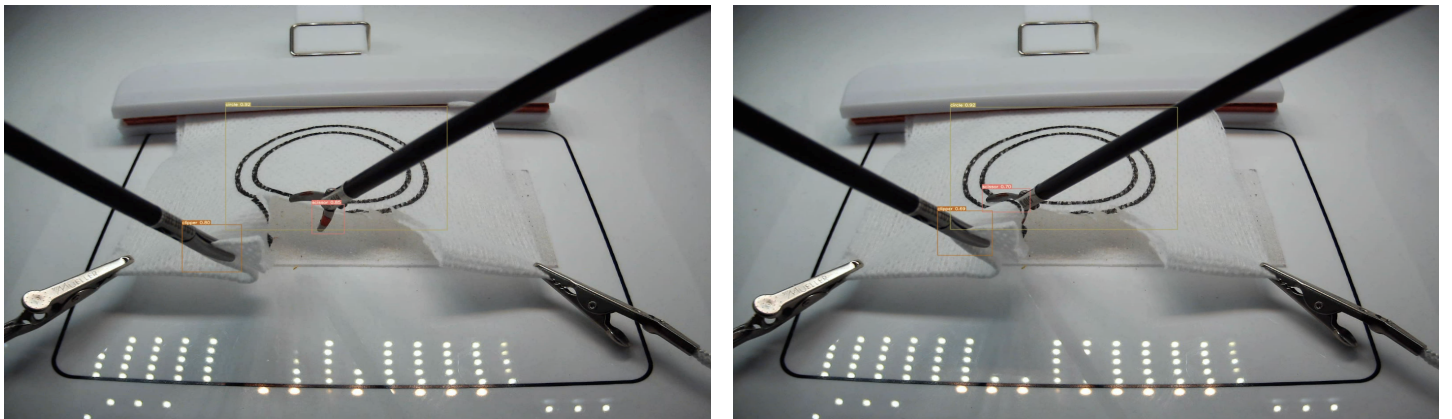
Figure 9: Qualitative results showing the detection of four test images.

(a) First camera tested video.

(b) Second camera tested video.

Figure 10: Qualitative results showing detection and tracking of laparoscopic instruments in two tested videos

One significant use of this model lies in the field of surgical education and performance evaluation. The output of the model, which includes images and videos, can serve as feedback for surgical performance or as a means of deliberate practice for cognitive behaviors during tests. Moreover, these results are accurate and can be utilized as input for further analysis.

5. Conclusions and Future Work

The aim of this study was to enhance laparoscopic surgical training and assessment by developing an extended dataset of instruments and objects for a box trainer pattern-cutting test and implementing a real-time object detection approach based on YOLOv7. Our findings demonstrate that our method effectively detects and tracks spatial tool and object movements and could be used to create a reliable real-time assessment system. Moving forward, we plan to integrate these results into a fuzzy logic decision support system to develop an automated GOALS assessment system.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

Table 4: A comparison summary of the proposed model with the related reported approaches in the literature.

Approach	Dataset	Labeling Tool	Model	Accuracy %	Precision %	Speed
The proposed model	1572 images extracted from our lab dataset [10]	Roboflow [27]	Yolov7	78	mAP 95.2	In real-time at 83.3 FPS
Surgical tools detection based on modulated anchoring network in laparoscopic videos [12]	5696 extracted from m2cai16-tool-locations and AJU-Set datasets	The data already labeled	Faster R-CNN	69.6% and 76.5% for each dataset	mAP 69.6% and 76.5% for each dataset	not reported
Real-Time Surgical Tool Detection in Minimally Invasive Surgery Based on Attention-Guided Convolutional Neural Network [13]	4011 extracted from EndoVis Challenge, ATLAS Dione, and Cholec80-locations datasets	not reported	ResNet50 with multi-scale pyramid pooling.	not reported	100, 94.05, and 91.65 for each dataset	In real-time at 55.5 FPS
Deep learning based multi-label classification for surgical tool presence detection in laparoscopic videos [14]	29478	Pixel Annotation Tool36	EndoNet	not reported	mAP 63.36	not reported
Identifying surgical instruments in laparoscopy using deep learning instance segmentation [15]	333	not reported	Mask R-CNN	not reported	AP 81	not reported
Surgical-tools detection based on convolutional neural network in laparoscopic robot-assisted surgery [17]	M2CAI 2016 Challenge videos	not reported	YOLO	not reported	mAP 72.26	48.9 FPS
Robust real-time detection of laparoscopic instruments in robot surgery using convolutional neural networks with motion vector prediction [18]	7492 extracted from m2cai16-tool-locations dataset	not reported	YOLO9000	not reported	mAP 84.7,	38 FPS
Instrument Detection for the Intracorporeal Suturing Task in the Laparoscopic Box Trainer Using Single-stage object detectors [19]	900 images extracted from our lab dataset [10]	Roboflow [27]	YOLOv4, Scaled-YOLOv4, YOLOR, and YOLOX	not reported	mAP50 0.708 0.969 0.976 0.922 for each model	not reported
Surgical Skill Assessment System Using Fuzzy Logic in a Multi-Class Detection of Laparoscopic Box-Trainer Instruments [21]	950 images extracted from our lab dataset [10]	not reported	SSD ResNet50 V1 FPN and SSD Mobilenet V2 FPN	not reported	65% and 80% reliability, 70 and 90 of fidelity for each architecture	not reported
Surgical Instrument Detection Algorithm Based on Improved YOLOv7x. [23]	452	LabelImg	Improved YOLOV7x algorithm based on RepLK Block and ODConv	94.7	not reported	not reported

References

[1] K. Alkhamaiseh, J. Grantner, S. Shebrain and I. Abdel–Oader, "Towards Automated Performance Assessment for Laparoscopic Box Trainer using Cross-Stage Partial Network," 2021 Digital Image Computing: Techniques and Applications (DICTA), 2021, 01-07, doi: 10.1109/DICTA52665.2021.9647393.

[2] Wang, Chien-Yao, A. Bochkovskiy, and H. Liao., " YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," (2022), arXiv. https://doi.org/10.48550/arXiv.2207.02696

[3] "Fls homepage." [Online]. Available: https://www.flsprogram.org. ˜ Last accessed 2 Dec 2022.

[4] R. Aggarwal, T. Grantcharov, K. Moorthy, J. Hance, and A. Darzi, "A competency-based virtual reality training curriculum for the acquisition of laparoscopic psychomotor skill," American journal of surgery, **191**(1), 128—133, January 2006. [Online]. Available: https://doi.org/10.1016/j.amjsurg.2005.10.014

[5] G. Islam, K. Kahol, B. Li, M. Smith, and V. L. Patel, "Affordable, web-based surgical skill training and evaluation tool," Journal of Biomedical Informatics, **59**, 102–114, 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1532046415002397

[6] G. A. Alonso-Silverio, F. P´erez-Escamirosa, R. Bruno-Sanchez, J. L. Ortiz-Simon, R. Mu˜noz-Guerrero, A. Minor-Martinez, and A. Alarc´on-Paredes, "Development of a laparoscopic box trainer based on open source hardware and artificial intelligence for objective assessment of surgical psychomotor skills," Surgical Innovation, **25**(4), 380–388, 2018, pMID: 29809097. [Online]. Available: https://doi.org/10.1177/1553350618777045

[7] N. J. Hogle, W. D. Widmann, A. O. Ude, M. A. Hardy, and D. L. Fowler, "Does training novices to criteria and does rapid acquisition of skills on laparoscopic simulators have predictive validity or are we just playing video games?" Journal of surgical education, **65**(6), 431–435, 2008.

[8] M. A. Zapf and M. B. Ujiki, "Surgical resident evaluations of portable laparoscopic box trainers incorporated into a simulation-based minimally invasive surgery curriculum," Surgical innovation, **22**(1), 83– 87, 2015.

[9] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning. nature 521 (7553), 436444," Google Scholar Google Scholar Cross Ref Cross Ref, 2015.

[10] https://drive.google.com/drive/folders/1F97CvN3GnLj-rqg1tk2rHu8x0J740DpC

[11] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein et al., "Imagenet large scale visual recognition challenge," International journal of computer vision, **115**(3), 211–252, 2015.

[12] B. Zhang, S. Wang, L. Dong, and P. Chen, "Surgical tools detection based on modulated anchoring network in laparoscopic videos," IEEE Access, 8, 23 748–23 758, 2020.

[13] Shi, P., Zhao, Z., Hu, S. and Chang, F., "Real-Time Surgical Tool Detection in Minimally Invasive Surgery Based on Attention-Guided Convolutional Neural Network," IEEE Access **8**, 228853-228862, 2020.

[14] S. Wang, A. Raju, and J. Huang, "Deep learning based multi-label classification for surgical tool presence detection in laparoscopic videos," in 2017 IEEE 14th International Symposium on Biomedical Imaging. IEEE, 620–623, 2017.

[15] S. Kletz, K. Schoeffmann, J. Benois-Pineau, and H. Husslein, "Identifying surgical instruments in laparoscopy using deep learning instance segmentation," in 2019 International Conference on Content-Based Multimedia Indexing (CBMI). IEEE, 1–6, 2019.

[16] E. Kurian, J. J. Kizhakethottam, and J. Mathew, "Deep learning based surgical workflow recognition from laparoscopic videos," in 2020 5th International Conference on Communication and Electronics Systems (ICCES). IEEE, 928–931, 2020.

[17] B. Choi, K. Jo, S. Choi, and J. Choi, "Surgical-tools detection based on convolutional neural network in laparoscopic robot-assisted surgery," in 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 1756–1759, 2017.

[18] K. Jo, Y. Choi, J. Choi, and J. W. Chung, "Robust real-time detection of laparoscopic instruments in robot surgery using convolutional neural networks with motion vector prediction," Applied Sciences, **9**(14), 2865, 2019.

[19] M. Mohaidat, J.L. Grantner, S.A. Shebrain, and I. Abdel-Qader, " Mohaidat M, Grantner JL, Shebrain SA, Abdel-Qader I. Instrument detection for the intracorporeal suturing task in the laparoscopic box trainer using single-stage object detectors," In2022 IEEE International Conference on Electro Information Technology (eIT) May 19, 455-460, 2022.

[20] M. Mohaidat, J.L. Grantner, S.A. Shebrain, and I. Abdel-Qader, "Multi-Class Detection and Tracking of Intracorporeal SuturingInstruments in an FLS Laparoscopic Box Trainer Using Scaled-YOLOv4," In Proceedings of the Advances in Visual Computing:17th International Symposium, ISVC 2022, San Diego, CA, USA, 3–5 October 2022

[21] F.R. Fathabadi, J.L. Grantner, S.A. Shebrain, and I. Abdel-Qader, "Surgical skill assessment system using fuzzy logic in a multi-class detection of laparoscopic box-trainer instruments," In 2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 1248-1253, 2021.

[22] F.R. Fathabadi, J.L. Grantner, S.A. Shebrain, and I. Abdel-Qader, "3D Autonomous Surgeon's Hand Movement Assessment Using a Cascaded Fuzzy Supervisor in Multi-Thread Video Processing," Sensors, 23, 2623, 2023. https://doi.org/10.3390/s23052623

[23] R. Boping, B. Huang, Sh. Liang, and Y. Hou, "Surgical Instrument Detection Algorithm Based on Improved YOLOv7x," Sensors (Basel, Switzerland) 23,11 5037. 24 May. 2023, doi:10.3390/s23115037

[24] https://pjreddie.com/darknet/

[25] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in Proceedings of the IEEE conference on computer vision and pattern recognition, 779–788, 2016.

[26] "School of medicine." [Online]. Available: https://med.wmich.edu/. ˜ Last accessed 2 Dec 2022.

[27] [Online]. Available: https://aroboflow.com/. Last accessed 2 Dec 2022.

[28] A. M. Obeso, J. Benois-Pineau, M. G. Vázquez, and A. R. Acosta, "Saliency-based selection of visual content for deep convolutional neural networks," Multimedia Tools and Applications, 78(8), 9553–9576, 2019.

[29] https://colab.research.google.com/drive/1X9A8odmK4k6l26NDviiT6dd6 TgR-piOa#scrollTo=GD9gUQpaBxNa

[30] https://colab.research.google.com/signup\

[31] Keras image preprocessing. Available at: march 2015; accessed 3 november 2021.

[32] Googleresearch, "tensorflow: large-scale machine learning on heterogneous systems," google res., 2015.