Box-Trainer Assessment System with Real-Time Multi-Class Detection and Tracking of Laparoscopic Instruments, using CNN

Fatemeh Rashidi Fathabadi, Janos L. Grantner, and Ikhlas Abdel-Qader

Department of Electrical and Computer Engineering, Western Michigan University, 1903 West Michigan Ave, Kalamazoo, MI 49008, USA
Fatemeh.rashidifathabadi@wmich.edu; janos.grantner@wmich.edu; ikhlas.abdelqader@wmich.edu

Saad A. Shebrain M.D.

Department of Surgery, Homer Stryker M.D. School of Medicine, Western Michigan University, 300 Portage Street, Kalamazoo, MI 49007, USA
saad.shebrain@med.wmich.edu

Abstract: In Minimally Invasive Surgery (MIS), surgeons need to acquire a specific set of skills, before carrying out a "real" operation. Training with the Laparoscopic Surgical Box-Trainer device helps in acquiring the needed skills for surgery residents which are traditionally not taught to them. Video recording of residents' performance and computer-assisted surgical trainers for MIS provide valuable information for resident's assessment. In this paper, we propose real-time detection and tracking of a multi-class of laparoscopic instruments for an intelligent box-trainer performance assessment system using SSD-ResNet50 V1 FPN architecture in TensorFlow backend. The dataset has been extracted from various laparoscopic box training videos. Using distance measurements and evaluation criteria constraints, we present an evaluation of the surgeon's performance. Based on the experimental result, the trained model could identify each instrument at the score of 90% fidelity, in each location, within a region of interest. This research is a result of a partnership between the Department of Electrical and Computer Engineering and the Department of Surgery, of the Homer Stryker M.D. School of Medicine, at Western Michigan University.

Keywords: Intelligent Laparoscopic Surgical Box-Trainer; Laparoscopic Surgical Tool Tip Tracking; Fuzzy Logic-Based Performance Assessment System

1 Introduction

Minimally invasive surgery (MIS) reduces complications and health risks with respect to traditional (open) surgery and decreases hospital stay. However, surgeons must acquire many skills before carrying out a real operation, for example, development of excellent eye-hand coordination while operating using visual information from two-dimensional monitor images and having confident control of the graspers and other laparoscopic surgery instruments, just to mention a few [1] [2]. To help in acquiring these skills various Virtual Reality (VR) trainers have been developed which can assist surgery interns to improve their skills [3] [4]. The assessment of surgical skills requires a considerable amount of time and effort. In recent decades, various training methods have been introduced to provide valuable feedback, expedite the development of surgery skills and assess the trainees' performance [5] [6] [7]. By monitoring a recorded video of a surgeon's performance or observing it in real-time in an Operating Room (OR) during laparoscopic procedures, the assessment procedure can be implemented. Furthermore, object detection and distance estimation concerning the laparoscopic surgical instruments and the test platforms are two fundamental factors for creating an intelligent performance assessment system in MIS [8]. In this paper, we propose a multi-class, real-time detection and tracking system for laparoscopic instruments using SSD-ResNet50 V1 FPN. It will enhance the capabilities of our intelligent box-trainer system [9] [10]. The paper is organized as follows: Section 2 reviews related work in this research area. Section 3 presents tools and utilities which were applied in this approach. Section 4 provides a detailed explanation of the methodology employed in this research. Sections 5 and 6 describe the tracking point location procedure and the model training and evaluation processes, respectively. In Section 7, the real-time tracking and assessment procedures are outlined. Finally, in Section 8, conclusions and plans for further research are given.

2 Related Works

In this section, we focus on the most recent work of researchers regarding surgeons' performance accuracy enhancement during MIS. As mentioned previously, laparoscopic instrument detection and tool-tip tracking contribute to the surgeon's performance assessment. In what follows, we review modern methods that have been proposed in the areas of object detection algorithms, tooltip tracking, and performance assessment [6-35]. Although in the field of MIS, researchers have proposed to apply texture features, color detection, Haar wavelets, and gradient-based features for both processing medical images and hardware-based simulators, there is an emerging trend in recent decades to utilize deep learning approaches [11]. For laparoscopic box-trainer systems, the approach to detect the surgical tools and the movements of tooltips in 3D space by using deep learning along with real-time performance assessment is relatively new.

In more recent studies scholars have been predominantly working on Deep Learning algorithms and Computer-Aided Diagnosis (CAD) system [12]. Yamazaki et al. [6] applied the open-source neural network platform YOLOv3 to detect the movements of surgical instruments in video recordings of laparoscopic gastrectomy procedures.

Namazi et al. [8] proposed a method to assess the surgeon's performance using a Deep Learning System (SPD-DLS) to identify the surgical phases from recorded videos of a laparoscopic procedure. They used a deep Convolutional Neural Network (CNN) followed by a Long Short-Term Memory (LSTM) model to consider both spatial and temporal information to identify the surgical phases in the video. Grantner et al. in [9] proposed an Intelligent Box-Trainer System (IBTS) to implement tooltip tracking tasks in 3D space, measurement of the forces applied by the grasper's jaws, and task execution time, and an assessment system for the laparoscopic surgeon's performance using fuzzy logic. They worked with a color-filtering algorithm for tool-tip tracking. Allen et al. [13] estimated the tooltip position by detecting the laparoscopic instrument's shaft in each image. They employed color space analysis to extract the instrument contours and then utilized line fitting to estimate the direction of movement for each laparoscopic instrument. In the end, to identify the position of the tool-tip of each instrument they employed a linear search. Perez-Escaminosa et al. [14] detected and tracked movements of laparoscopic instruments in a three-dimensional workspace using a sensor-free system based on green and blue color markers which were placed on the tip of the instruments [13] [15].

In [16], researchers developed a tracking algorithm using a sequence of image contrast enhancement, Sobel Filtering, and color-based segmentation. This algorithm extracts information obtained from the laparoscopic instrument's shaft edge to extract the motion fields of laparoscopic instruments via video tracking. Sun et al. [17] utilized an adaptive fusion kinematics method in an autonomous surgical instrument detection and tracking algorithm. They developed a fuzzy logic system to adjust the kinematics weights and laparoscopic information. Huang et. al. [18] proposed a method to estimate the position, velocity, and direction of laparoscopic instruments which were used in a tracking module. They utilized an Inertial Measurement Unit (IMU) providing direct motion information for the laparoscopic instrument tracking module. Moreover, an Extended Kalman Filter was employed to integrate the information from the different sources to compensate for the biases of the IMU in a unified framework.

Zahiri et al. [19] implemented an Image-Based Tool Tracking system using two-color markers placed on two graspers. Gautier et al. in [20] proposed a surgeon's performance assessment system using colored tapes attached to the end of the laparoscopic instrument. By tracking the colored tapes, frequency analysis and linear discriminant analysis of the 3D reconstructed trajectories of the instruments were extracted to assess the surgeon's skills. Partrige et al. [10] utilized a color-thresholding motion-tracking program to track the movement of colored laparoscopic instrument tips, providing objective performance feedback to a

portable laparoscopic box simulator. Dockter et al. [21] developed a 3D tracking algorithm at a high rate of computational speed to validate its performance in a da Vinci surgical endoscope. Islam et al. [22] presented a tool-tip tracking algorithm based upon a fuzzy logic assessment system by utilizing a web-based video telemonitoring system to monitor and track movements of surgeons' hands and the surgical tooltips. To assess the surgeon's skills, they used colored tool-markers to extract velocity, acceleration, jerkiness, and snaps of the tools' movements during laparoscopic procedures. In recent studies of laparoscopic surgical operations, researchers have been interested in working with Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) [23]. Using Region-Based Convolutional Neural Networks and a new dataset m2cai16-tool location, Jin et al. [24] used learning of instrument regions in cholecystectomy to detect and localize the region of interest of surgical tools in laparoscopic surgical videos.

In [25], Kletz et al. worked with a Deep Learning instance segmentation approach in recorded videos using a region-based Fully Convolutional Network. They managed to identify instruments as multi-class instance segmentation and determined each instrument classification. Zhang et al. [26] proposed a Modulated Anchoring Network for the detection of laparoscopic surgery tools based on the Faster R-CNN which was made up of a new anchoring scheme referred to as modulated anchoring and a relation module on an existing dataset (m2cai16-tool-locations) as well as a new private one (AJU-Set). Choi et al. [27] proposed real-time models for the detection of surgical instruments during laparoscopic surgery using a dataset that included information on the seven surgical tools for learning the CNN model. To track surgical instruments in real-time, the unified architecture of YOLO was applied to the models.

Wang et al. [28] proposed a multi-label classification deep learning method that combined two deep neural networks, VGGNet and GoogLeNet, to detect the surgical tools in laparoscopic videos. Colleoni et al. [29] proposed a Fully Convolutional Neural Networks (FCNNs) encoder-decoder architecture for surgical instrument joint detection and localization using three-dimensional convolutional layers to exploit spatio-temporal features from laparoscopic videos. The researchers used the EndoVis and UCL dVRK datasets for training testing procedures. Hasan et al. [30] presented a U-NetPlus model for the surgical tool segmentation which is the modification of the U-Net architecture by introducing both VGG-11 and VGG-16 as an encoder and redesigned the decoder part by replacing the transposed convolution operation with an up-sampling operation based on the nearest-neighbor (NN) interpolation followed by two convolution layers.

In the paper by Kurian et al. [31], researchers used the CNN architecture ResNet50 to recognize four surgical phases:

1) Preparation
2) Trocar placement
3) Clipping and cutting
4) Gallbladder retraction

They combined ResNet and temporal features in the form of I3D and LSTM. Kanakatte et al. [32] presented a deep architecture, in which a pixel-wise instance segmentation algorithm segmented and localized the surgical tool in cholecystectomy surgery videos. Jo et al. [33] presented a detection and classification surgical instruments algorithm in laparoscopic images which can work under real-time conditions, too. This algorithm is based on the object detection system YOLO9000. In the paper by Jonmohamadi et al. [34], the researcher used trained fully convolutional neural networks with U-net and U-net++ architectures to segment four key structures of the knee, such as Femur, ACL, Tibia, and Meniscus, in an automated fashion. Zhang et al. [35] developed a marker-free surgical instrument tracking framework based on object extraction using the LinkNet-18 network architecture which belongs to U-Net. In this work, the researchers used a masking method to segment each part of a laparoscopic instrument such as the end-effector, the shaft, and also the background. For real-time tracking, a target trajectory has been defined for the laparoscope-holder robot to be tracked. Using Euclidean Distance Transformation, the binary image was transformed to a distance.

Zijian et al [36] proposed an algorithm that tracks two parts of the surgical instrument: the end-effector and the shaft. In this approach, the shaft detection has been done by edge-points and line features and the trained CNN has been utilized to track and detect the end-effector. Zhu et al [37] proposed an end-to-end learning-based approach to predict distances for given objects in the RGB images. Their method includes three components: a feature extractor, a distance regressor and a multiclass classifier. In this method, a base model extracts features from images, then uses ROI pooling to generate a fixed-size feature vector for each object, and finally feeds the ROI features into a distance regressor to predict the distance for each object.

3   Tools and Utilities

In our study, we propose an intelligent box-trainer performance assessment system based upon real-time detection and tracking of multi-class of laparoscopic instruments. To detect and track the laparoscopic instruments, we used a deep learning approach. Our network is based on an open-source Tensorflow Object Detection Application Programming Interface (API[1]), and we used SSD-ResNet-50 [38] model Feature Pyramid Network (FPN) Architecture as a backbone of our network. TensorFlow is a frequently used software for Machine Learning (ML) applications that provides an interface to common ML algorithms and executable code for various models [39]. In this work, Tensorflow is the backend for object detection and image processing algorithm.

---

[1]   Application Programming Interface (API)

3.1   Experimental Setup

Our Intelligent Box-trainer System (IBTS) is depicted in Figure 1. It is our development platform to create hardware and software architectures and algorithms aiming for the development of an objective laparoscopic-surgery skills assessment system [9]. The main components of the IBTS are as follows: an FLS box-trainer device, two 5-megapixel USB 2.0 cameras with variable-focus lens, a 32' HD computer monitor to visualize the underlying test procedure carried out on a particular FLS platform, extra LED strips for better lighting conditions, a tablet which is used by the supervising medical personnel, a PC workstation to record the test videos and run the tracking and assessment programs and a router to implement wireless communications between the tablet and the PC workstation. One of the standard FLS pattern cutting tests was used in this study. In this test, the platform is an artificial tissue with a bold circle line on it. During the test, the surgeon holds the tissue in place by a grasper while using a pair of scissors to cut the tissue around the circle such that the cutting distance from the circle line should not exceed 5 mm, and cutting into the line is also considered as a failure of the test.

(a)   (b)

Figure 1

a) The IBTS System

b) Tracking the laparoscopic instruments and generating real-time performance assessment using the
IBTS System in the Homer Stryker M.D. School of Medicine, of WMU

3.2   Dataset

In this study, we used our custom dataset (IFCL.LBT100) that has been created for laparoscopic box trainer's performance assessment research. For this project, we have created a relatively large dataset using various laparoscopic training videos. Our custom dataset is composed of extracted frames from these videos. The frames have been manually annotated using the Image Annotation Tool LabelImg[2] which

---

2   http://tzutalin.github.io/labelImg/

is a free and open-source software. Each labeled image has its individual .xml which can be converted to .csv files and .tf.record files which are used during training processes. Having recorded and processed more test videos we plan to post our dataset online for other researchers in this field.

3.2.1   Partitioning of the Dataset

Once the annotation of our dataset was finished, we classified all the images and .xml, .csv, and .tf record files for the training, testing, and evaluation tasks. Typically, the ratio of this arrangement is 6:2:2, i.e., 60% of the images are used for training, 20% for testing, and the remaining 20% is used for evaluation purposes.

3.2.2   Creating the Label Map

To satisfy the training algorithm, we prepared a label map that maps each of the classes to an integer value. This label map is used both by the training and detection processes. A simple example of the label map for our dataset contains three labels: a scissor, a grasper, and a circle pattern on an artificial tissue which are considered as the laparoscopic instruments in a laparoscopic box-trainer. This label map file (with the extension of .pbtxt) is illustrated in Figure 2.

```
item {
   id: 1
   name: 'Scissor'
}

item {
   id: 2
   name: 'Grasper'
}
item {
   id: 3
   name: 'Circle'
}
```

Figure 2

Label map file example

3.2.3   Data Augmentation

The TensorFlow Object Detection API Image Preprocessor tool provides multiple data augmentation steps with variation and modification from the original data. Applying these augmentation steps to the dataset the neural networks can use more training data to achieve better performance. In our approach, to train the model based on SSD-ResNet50 V1 FPN, we adequately augmented the dataset using TensorFlow API data augmentation variables.

4   Methodology

We selected a collection of detection models and pre-trained them on the COCO
2017 dataset such as the EfficientDet D1 640x640, SSD MobileNet V1 FPN
640x640, and SSD-ResNet50 V1 FPN from TensorFlow 2 Detection Model Zoo
and Detecto Module in Pytorch [40]. These models are useful for initialization when
training on our new datasets. By comparing the performance of these models, we
have concluded that SSD-ResNet50 delivers better performance with respect to
real-time detection. We trained our model based upon the SSD-ResNet50 V1 FPN
Architecture. The entire workflow of the SSD-ResNet50 V1 FPN Architecture is
illustrated in Figure 3. SSD with the ResNet50 V1 FPN feature extractor in its
architecture is an object detection model that has been trained on the COCO 2017
dataset. A Momentum optimizer with a learning rate of 0.04 was used for the region
proposal and classification network, and the learning rate was reduced on the
plateau. As shown in Figure 3, the Feature Pyramid Network (FPN) generates the
multi-level features as inputs to the SSD-ResNet50 Architecture. The FPN is an
extractor and provides the extracted feature maps layers to an object detector. When
the model localizes any small object, it draws an object boundary box around it at
each location. After training the model, the testing procedure was carried out by
providing the surgical videos as input to the trained model. Afterward, we used
Tensorboard which is a suitable feature of the TensorFlow Object Detection API.
It allowed us to continuously monitor and visualize several different
training/evaluation metrics when our model was being trained. As the final step, we
obtained the output video containing the labeled surgical instruments and the
assessment results along with the log file. The generated log file records the surgical
assessment, the bounding box for each laparoscopic instrument, and the center point
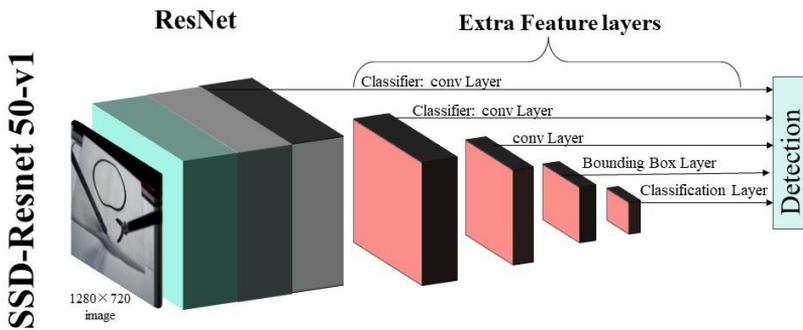of each laparoscopic instrument.

Figure 3
SSD-ResNet50 V1 FPN Architecture

5   Tracking Point Location Procedure

By providing the surgical videos as the input to the trained model, we detected and localized each laparoscopic instrument with high accuracy. After object extraction, we used Euclidean Distance Transformation[3] [35] to measure the distance between the center of the circle pattern and the center of the scissors' bounding box where the tissue is cut. To assess the surgeon's performance, we measure the distance between the spot where the tissue is cut and the closest area of the circle. This distance should not exceed 5 mm for passing the assessment test. The measurement procedures and formulations are defined by Eqs. (1) thru (8). Dis stands for the distance of two points in each frame and Pix stands for the pixel point set which contained all the pixel points of an extracted object of each frame.

$$D[A][B] = \min\{Dis[(A_x . A_y) . (B_x . B_y)] . (A. B) \in Pix\}$$

$$Dis[(A_x . A_y) . (B_x . B_y)] = \sqrt{(B_x - A_x)^2 + (B_y - A_y)^2}$$

(1)

Given a line $y = mx+b$, the slope m delineates the ratio between the change in x, defined by dx, and the change in y, defined by dy. Hence, the slope creates a relationship between a change in the y-values with respect to a change in the x-values which is a derivative of y to x:

$$dy = mdx$$

(2)

$$dy^2 = (B_y - A_y)^2$$

(3)

$$dx^2 = (B_x - A_x)^2$$

(4)

$$dx^2 + dy^2 = (B_x - A_x)^2 + (B_y - A_y)^2$$

$$dx^2 + dy^2 = Dis^2$$

(5)

By substituting Eq. (2) in Eq. (5) we obtain Eq. (6)

$$dx^2 + (mdx)^2 = Dis^2$$

(6)

Using Eq. (2) and Eq. (6), we can calculate the changes in the x-values by Eq. (7) and the y-values by Eq. (8) to obtain the distance between each of two points in our approach.

$$dx = \sqrt{\frac{Dis^2}{1 + m^2}}$$

(7)

---

[3]   https://github.com/alejandrods/Social-Distance-Using-TensorFlow-API-Object

$$dy = m\sqrt{\frac{Dis^2}{1+m^2}} \tag{8}$$

Using these equations, we measured the bounding boxes of each laparoscopic instrument in each frame. Based upon real measurements and the number of pixels in each frame, we calibrated the position of each instrument to a real value. The assessment measurement algorithm is illustrated in detail in Figure 4. There, A marks the center of the circle pattern, and B marks the center of the scissors box. Line d, which connects points A and B is defined as the distance of the scissors from the center of the circle. Using Eq. (1) to Eq. (8), we calculated the radius of the circle in each frame.

By subtracting the center of the scissors bounding box from the center of the circle pattern, the distance between the scissors and the center of the circle is calculated for the assessment procedure. These calculations may lead to inaccurate measurements under some circumstances, e.g., when the grasper wrinkles the artificial tissue, or when the trained model cannot recognize the instrument. In our research, there were some short periods, typically lasting for a few seconds, when the model couldn't find the circle which, lead to inaccurate measurement. In other cases, the model could detect the circle by keeping the reference center of the circle in its place by localizing the bounding box of the circle in each frame.
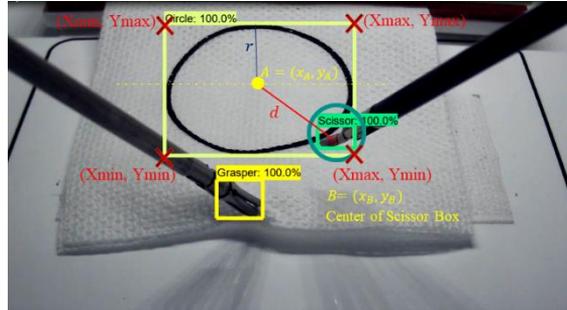

Figure 4

Illustration of the assessment measurement calculations

6   Model Training and the Evaluation Process

The classification loss, which is used to measure the model's confidence by classifying the pixel's region confined by the bounding box, is illustrated Figure 5. The localization loss that measures the geometric distance between the predicted bounding box and the ground truth annotation (validation bounding boxes) is depicted in Figure 6. The overall loss function or total loss is a weighted

combination of the classification loss and the localization loss [41]. It is depicted in Figure 7, which illustrates the performance of the model during training, i.e., what the network predicts for the image versus the allocated label of the image at the end of each epoch during the training process. The train-validation total loss, as it is shown Figure 7, is sometimes higher than the training loss but it decreases over time and, hence, it exhibits a satisfactory result.
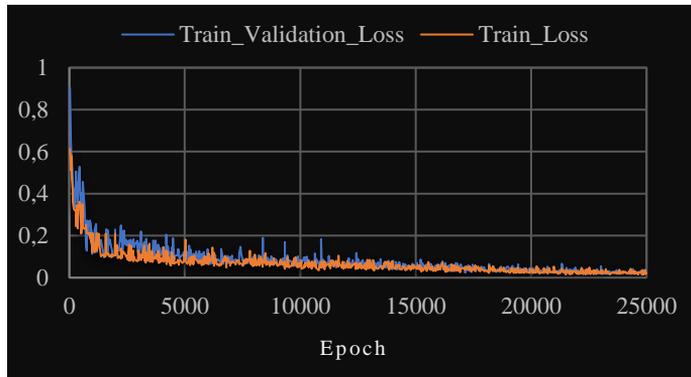

Figure 5

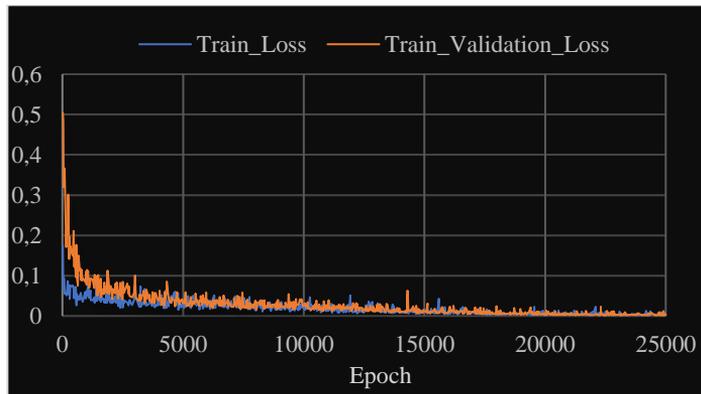The comparison of overall train-classification loss and train-validation classification loss

Figure 6

The comparison of overall train- localization loss and train-validation localization loss
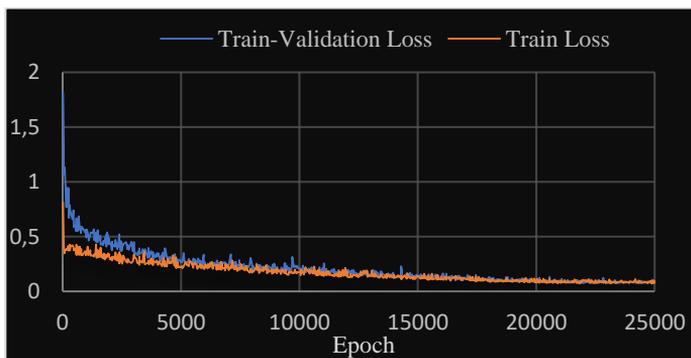
Figure 7
The comparison of overall train-total loss and train-validation total loss

7   Real-time Tracking and the Assessment System

To implement the tracking task, the tracking point has to be located frame-by-frame in the laparoscopic test videos. In our implementation, we analyzed tracking and generated the surgeons' performance assessment along with it.
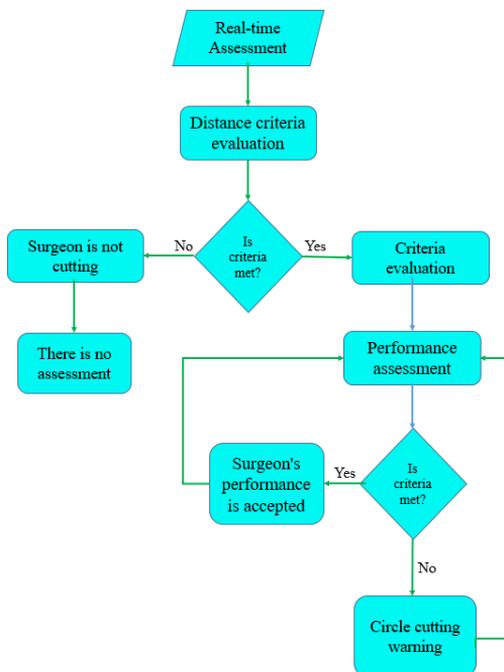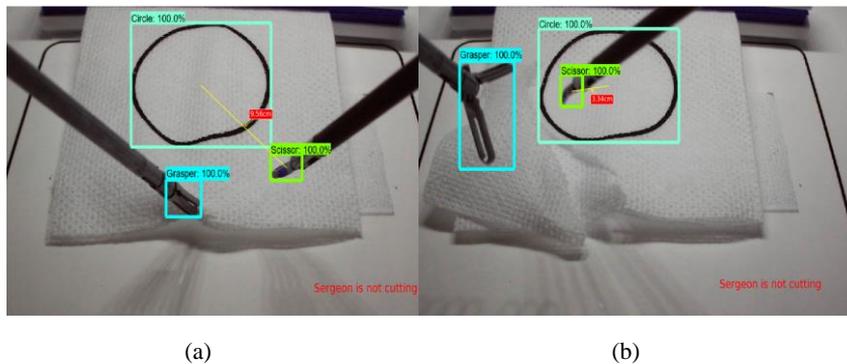

Figure 8
The real-time tracking and assessment system flowchart

In the tracking procedure, we expect that the network detects and localizes each laparoscopic instrument in real-time. However, the surgeon's performance will be assessed only during the actual cutting of the tissue. To clarify the reasons for this, we have simplified the assessment procedure to just three parts. In each part, many components are taken into account for the surgeon's performance assessment. Only the following processes are considered: (1) the surgeon is not cutting, (2) criteria evaluation and performance assessment, and (3) circle cutting warning if the established criteria are not met. The tracking and assessment system flowchart is depicted in Figure 8. In what follows, we investigate each of these processes.

7.1   Surgeon is not Cutting and the Procedure does not Commence

In this situation, either the surgeon has not started the cutting or the scissors are in the air, i.e., they are way above the artificial tissue. By defining different constraints for the distance of the center of the circle to the tips of the scissors, the network can recognize when the surgeon is not cutting. In Figure 9 (a to h), different conditions are illustrated when no cutting takes place. When the surgeon is about to start the cutting (a), there is no need for performance assessment. When the surgeon is in the middle of the cutting process but he temporarily stops doing it and releases the tissue to take a different approach for continuing the task, there is no cutting, either. The network recognizes it when a surgeon is not cutting (b, d, e, h). In two illustrated scenarios the surgeon intends to continue with the cutting but just a small section of the scissors is in the frame. In these cases, the network correctly recognizes the situation and decides "no cutting" is taking place (f, g).
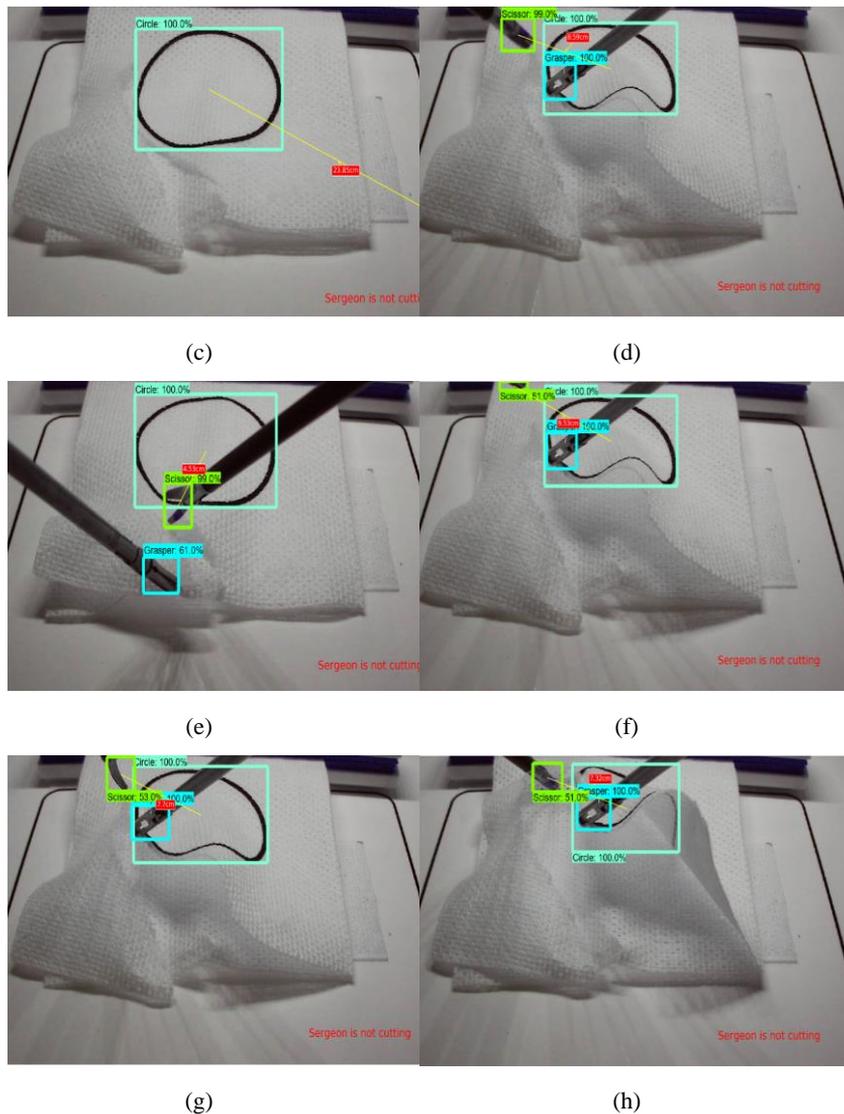

(a)   (b)

(c)　　　　　　　　　　　　　　　(d)

(e)　　　　　　　　　　　　　　　(f)

(g)　　　　　　　　　　　　　　　(h)

Figure 9
Surgeon is not cutting leading to no assessment

7.2　Performance Assessment is Active

The surgeon's performance will be assessed based upon an error distance, d_e, a distance between the spot where the scissors made the cut and the section of the circle which is at the closest point to the scissors' tips, as shown in Figure 10. The surgeon's performance is not acceptable unless the error distance $d_e < 5$ mm, at

all times. If this condition is not met, the system renders an assessment but it doesn't accept the surgeon's performance. In this case, there are three scenarios to consider.

1) In Figure 10 (a), the surgeon is cutting, however, he is only getting to a position in which he can actually begin with the task. The tips of the scissors are too far from the circle and the network should only deliver assessment during the pattern cutting task. Clearly, in this situation, we do not expect the network to deliver any valid assessment.

2) In Figure 10 (b), the tips of the scissors are close enough to the circle, so after this moment, the surgeon's performance will be monitored and assessed. Based upon Figs. 10 (c, d, e), the surgeon's performance is deemed good because the measured distance between the tips of the scissors and the circle line is less than 5 mm. In Figure 10 (f), the surgeon restarted cutting after he had changed the direction he wanted to move the scissors. As expected, in the case of each start, the distance of the scissors' tips from the circle line is typically larger than that when the surgeon is cutting continuously.

3) In Figs. 10 (g, h), the most challenging scenarios are illustrated: the tissue is wrinkled by the grasper but the reference center of the circle is still in its place, i.e., it is visible to the camera.

Having enough images in a dataset has a great impact on training a model. In our study, because the number of images was not as large as it should be, the model could not recognize the circle in few instances. In particular, when the tissue is wrinkled by the grasper such that the circle line disappears from the sight of the camera. In a situation like that not only the complete circle cannot be recognized by the model, but even an expert cannot recognize it as a circle. Therefore, in such moments, we have a pass-fail assessment. To address this problem, we have to record more videos including many frames of such cases intentionally containing this scenario. Under normal conditions, this situation rarely happens. The more we can train a model to understand this scenario, the better the prediction analysis will be. In addition, installing a third camera into the system, which is positioned directly above the platform, will help in resolving this problem, as we continue our research.
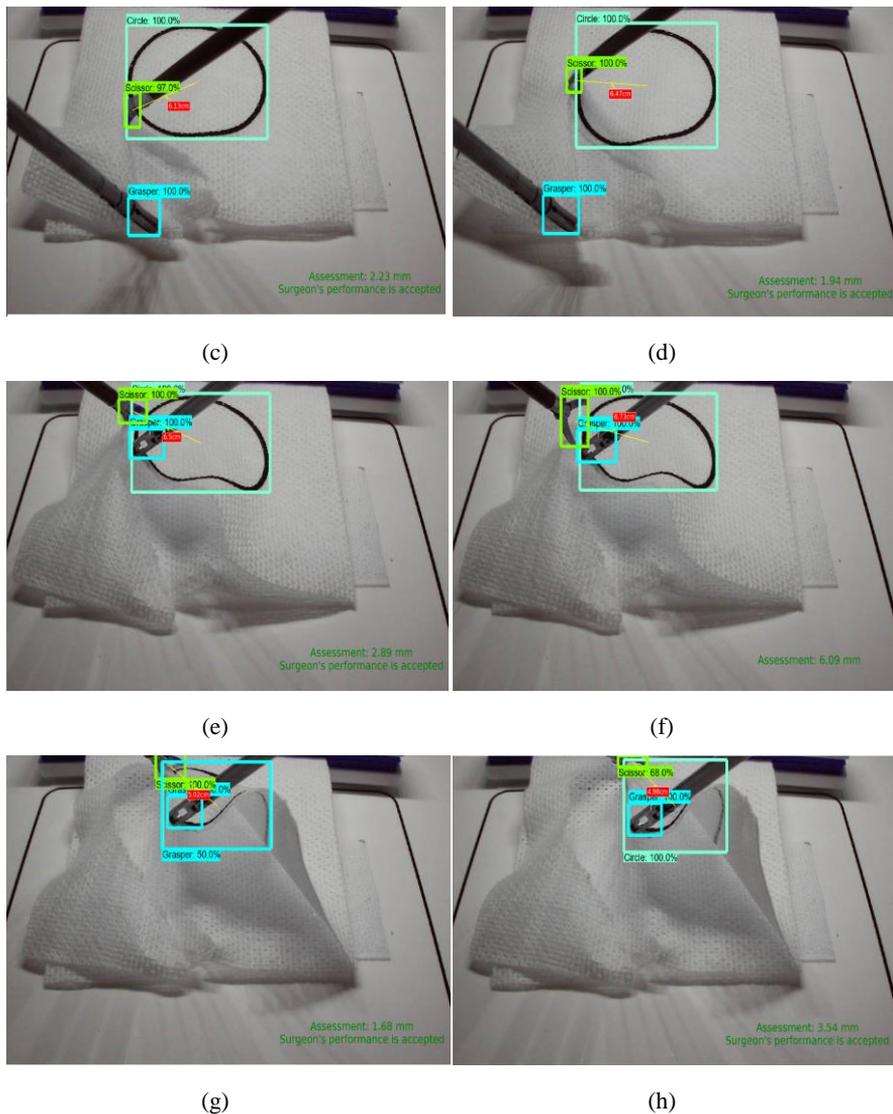

(a)   (b)

(c)   (d)

(e)   (f)

(g)   (h)
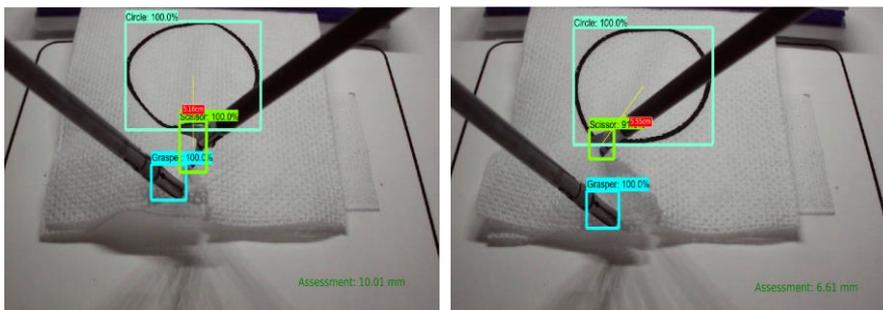
Figure 10

Surgeon's performance is assessed leading to performance acceptance or rejection

7.3 Circle Cutting Warning and Criteria is not Met

There is an additional requirement for passing the pattern cutting test. It is mandatory that the surgeon should not cut through the circle line. To help in meeting this constraint, we defined a rule in which if the tips of the scissors are too close to the circle line (defined as less than 0.5 mm) it will alert the surgeon about this

situation. This scenario is illustrated in Figure 11, when the surgeon was cutting too close to the circle line.
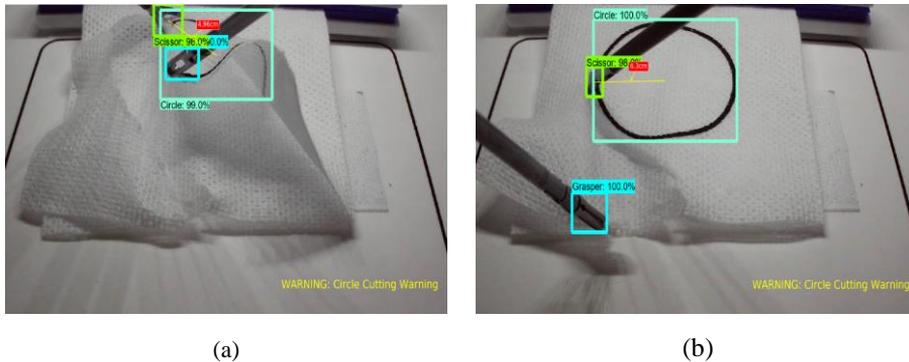

(a)   (b)

Figure 11

Surgeon is getting too close to the circle, leading to circle cutting

Conclusions and Future Research

In this paper, we proposed employing real-time detection and tracking of a multi-class of laparoscopic instruments for an intelligent box-trainer performance assessment system. We generated the dataset using extracted frames of various training videos using a laparoscopic box-trainer. Moreover, we added a distance measurement algorithm to the object detection algorithm in the TensorFlow backend using ResNet-50- architecture. The algorithm continuously measures changes in the distance of the center of the circle from the center of the scissors' tips and also the changes of the distance, where the tissue is been cut, from the circle line. Using distance measurements and evaluation criteria constraints, we assessed whether the surgeon's performance was accepted or not. Based on the experimental result, the trained model could identify each instrument at the score of 90% fidelity, in each location within a region of interest, and determine their relative distance with 65% reliability, under real-time conditions. There were few instances when the detection failed to lead to pass-failed assessment, in particular, when the tissue was wrinkled by the grasper. The error rate in carrying out these tasks was less than 20%. We assume that the performance measures of the system can be improved by adding an additional, top camera to the system and measure the distance from different perspectives. In future research, we plan to develop an automated performance assessment system, by tracking the laparoscopic instruments, under real-time conditions, measuring the test execution times and fusing the measured data with expert surgeon opinion, in the framework of a fuzzy logic-based intelligent decision support system.

Acknowledgement

(University of Tehran) for sharing his experience in this type of work, *hosseinrahmatpour@ut.ac.ir.*

References

[1]   A. Chellali *et al.*, "Achieving interface and environment fidelity in the Virtual Basic Laparoscopic Surgical Trainer," *Int. J. Hum. Comput. Stud.*, Vol. 96, pp. 22-37, 2016

[2]   D. Oh *et al.*, "Surgical techniques for totally laparoscopic caudate lobectomy," *J. Laparoendosc. Adv. Surg. Tech.*, Vol. 26, No. 9, pp. 689-692, 2016

[3]   B. S. Peters, P. R. Armijo, C. Krause, S. A. Choudhury, and D. Oleynikov, "Review of emerging surgical robotic technology," *Surg. Endosc.*, Vol. 32, No. 4, pp. 1636-1655, 2018

[4]   H. Jiang *et al.*, "Enhancing a laparoscopy training system with augmented reality visualization," in *2019 Spring Simulation Conference (SpringSim)*, 2019, pp. 1-12

[5]   B. W. King, L. A. Reisner, A. K. Pandya, A. M. Composto, R. D. Ellis, and M. D. Klein, "Towards an autonomous robot for camera control during laparoscopic surgery," *J. Laparoendosc. Adv. Surg. Tech.*, Vol. 23, No. 12, pp. 1027-1030, 2013

[6]   Y. Yamazaki *et al.*, "Automated Surgical Instrument Detection from Laparoscopic Gastrectomy Video Images Using an Open Source Convolutional Neural Network Platform," *J. Am. Coll. Surg.*, Vol. 230.5, pp. 725-732, 2020

[7]   S. Ahmadi and S. J. Azhari, "A LP, Very High-CMRR, Wide-Bandwidth FDCCII-Based CMIA Adapted to Both Current and Voltage Inputs," *Arab. J. Sci. Eng.*, Vol. 44, No. 8, pp. 6727-6740, 2019

[8]   B. Namazi, G. Sankaranarayanan, and V. Devarajan, "Automatic detection of surgical phases in laparoscopic videos," in *Proceedings on the International Conference on Artificial Intelligence (ICAI)*, 2018, pp. 124-130

[9]   J. L. Grantner, A. H. Kurdi, M. Al-Gailani, I. Abdel-Qader, R. G. Sawyer, and S. Shebrain, "Multi-Thread Implementation of Tool Tip Tracking for Laparoscopic Surgical Box-Trainer Intelligent Performance Assessment System," *Acta Polytech. Hungarica*, Vol. 16, No. 9, 2019

[10]   R. W. Partridge, M. A. Hughes, P. M. Brennan, and I. A. M. Hennessey, "Accessible laparoscopic instrument tracking ('InsTrac'): construct validity in a take-home box simulator," *J. Laparoendosc. Adv. Surg. Tech.*, Vol. 24, No. 8, pp. 578-583, 2014

[11]   X.-Y. Zhou, Y. Guo, M. Shen, and G.-Z. Yang, "Artificial Intelligence in Surgery," *arXiv Prepr. arXiv2001.00627*, 2019

[12]   P. Eskandari and S. B. Shokouhi, "DT-CWT: A New Feature for Tumor Classification in Breast DCE-MRI," *Mapta J. Electr. Comput. Eng.*, Vol. 3, No. 1, pp. 35-39, 2021

[13]   B. F. Allen, F. Kasper, G. Nataneli, E. P. Dutson, and P. Faloutsos, "Visual tracking of laparoscopic instruments in standard training environments.," in *MMVR*, 2011, pp. 11-17

[14]   F. Pérez-Escamirosa, I. Oropesa, P. Sánchez-González, J. Tapia-Jurado, J. Ruiz-Lizarraga, and A. Minor-Martínez, "Orthogonal cameras system for tracking of laparoscopic instruments in training environments," *Cir. Cir.*, Vol. 86, No. 6, pp. 548-555, 2019

[15]   C. B. Duane, "Close-range camera calibration," *Photogramm. Eng*, Vol. 37, No. 8, pp. 855-866, 1971

[16]   I. Oropesa *et al.*, "EVA: laparoscopic instrument tracking based on endoscopic video analysis for psychomotor skills assessment," *Surg. Endosc.*, Vol. 27, No. 3, pp. 1029-1039, 2013

[17]   Y. Sun, B. Pan, S. Zou, and Y. Fu, "Adaptive Fusion-Based Autonomous Laparoscope Control for Semi-Autonomous Surgery," *J. Med. Syst.*, Vol. 44, No. 1, p. 4, 2020

[18]   C.-C. Huang, N. M. Hung, and A. Kumar, "Hybrid method for 3D instrument reconstruction and tracking in laparoscopy surgery," in *2013 International Conference on Control, Automation and Information Sciences (ICCAIS)*, 2013, pp. 36-41

[19]   M. Zahiri, R. Booton, K.-C. Siu, and C. A. Nelson, "Design and evaluation of a portable laparoscopic training system using virtual reality," *J. Med. Device.*, Vol. 11, No. 1, 2017

[20]   B. Gautier, H. Tugal, B. Tang, G. Nabi, and M. S. Erden, "Laparoscopy instrument tracking for single view camera and skill assessment," in *2019 International Conference on Robotics and Automation (ICRA)*, 2019, pp. 5039-5045

[21]   R. Dockter, R. Sweet, and T. Kowalewski, "A fast, low-cost, computer vision approach for tracking surgical tools," in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2014, pp. 1984-1989

[22]   G. Islam, K. Kahol, B. Li, M. Smith, and V. L. Patel, "Affordable, web-based surgical skill training and evaluation tool," *J. Biomed. Inform.*, Vol. 59, pp. 102-114, 2016

[23]   A. P. Twinanda, S. Shehata, D. Mutter, J. Marescaux, M. De Mathelin, and N. Padoy, "Endonet: a deep architecture for recognition tasks on laparoscopic videos," *IEEE Trans. Med. Imaging*, Vol. 36, No. 1, pp. 86-97, 2016

[24]   A. Jin *et al.*, "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)*, 2018, pp. 691-699

[25]   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)*, 2019, pp. 1-6

[26]   B. Zhang, S. Wang, L. Dong, and P. Chen, "Surgical Tools Detection Based on Modulated Anchoring Network in Laparoscopic Videos," *IEEE Access*, Vol. 8, pp. 23748-23758, 2020

[27]   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)*, 2017, pp. 1756-1759

[28]   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 (ISBI 2017)*, 2017, pp. 620-623

[29]   E. Colleoni, S. Moccia, X. Du, E. De Momi, and D. Stoyanov, "Deep learning based robotic tool detection and articulation estimation with spatio-temporal layers," *IEEE Robot. Autom. Lett.*, Vol. 4, No. 3, pp. 2714-2721, 2019

[30]   S. M. K. Hasan and C. A. Linte, "U-NetPlus: A modified encoder-decoder U-Net architecture for semantic and instance segmentation of surgical instruments from laparoscopic images," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2019, pp. 7205-7211

[31]   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)*, 2020, pp. 928-931

[32]   A. Kanakatte, A. Ramaswamy, J. Gubbi, A. Ghose, and B. Purushothaman, "Surgical tool segmentation and localization using spatio-temporal deep network," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2020, pp. 1658-1661

[33]   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," *Appl. Sci.*, Vol. 9, No. 14, p. 2865, 2019

[34]   Y. Jonmohamadi *et al.*, "Automatic segmentation of multiple structures in knee arthroscopy using deep learning," *IEEE Access*, Vol. 8, pp. 51853-51861, 2020

[35]   J. Zhang and X. Gao, "Object extraction via deep learning-based marker-free tracking framework of surgical instruments for laparoscope-holder robots," *Int. J. Comput. Assist. Radiol. Surg.*, Vol. 15, No. 8, pp. 1335-1345, 2020

[36]   Z. Zhao, S. Voros, Y. Weng, F. Chang, and R. Li, "Tracking-by-detection of surgical instruments in minimally invasive surgery via the convolutional neural network deep learning-based method," *Comput. Assist. Surg.*, Vol. 22, No. sup1, pp. 26-35, 2017

[37]   J. Zhu and Y. Fang, "Learning object-specific distance from a monocular image," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 3839-3848

[38]   K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778

[39]   O. Alsing, "Mobile object detection using tensorflow lite and transfer learning" 2018

[40]   F. R. Fathabadi, J. L. Grantner, S. A. Shebrain, and I. Abdel-Qader, "Multi-Class Detection of Laparoscopic Instruments for the Intelligent Box-Trainer System Using Faster R-CNN Architecture," in *2021 IEEE 19th World Symposium on Applied Machine Intelligence and Informatics (SAMI 2021)*, pp. 149-154

[41]   Y. Wu *et al.*, "Rethinking Classification and Localization for Object Detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 10186-10195