An AI system for evaluating pass fail in fundamentals of laparoscopic surgery from live video in realtime with performative feedback

1st Yunzhe Xue
Department of Data Science
New Jersey Institute of Technology
Newark, NJ, USA
yx277@njit.edu

3th Rohit Muralidhar
General Surgery
Robert Wood Johnson Hospital
New Brunswick, NJ, USA
reachrohit92@gmail.com

5th Advaith Bongu
Transplant Surgery Division
Robert Wood Johnson Hospital
New Brunswick, NJ, USA
Advaith.Bongu@rwjbh.org

2nd Andrew Hu
General Surgery
Robert Wood Johnson Hospital
New Brunswick, NJ, USA
ah1425@rwjms.rutgers.edu

4rd Justin W. Ady
Vascular and Endovascular Surgery
Robert Wood Johnson Hospital
New Brunswick, NJ, USA
jwa60@rwjms.rutgers.edu

6th Usman Roshan
Department of Data Science
New Jersey Institute of Technology
Newark, NJ, USA
usman@njit.edu

***Abstract*—Medical students preparing to be a surgeon are required to demonstrate proficiency in laparoscopic surgery as part of their training. This is done via simulation on a Fundamentals of Laparoscopic Surgery (FLS) kit where the student has to use graspers to transfer six rings from one set of pegs to another and then again back to the original set of pegs without dropping the peg. A peg dropped outside the box is considered a fail and inside is the box is allowed as long as the transfer resumes with the correct grasper. We present an AI system that automatically determines if a student has passed or failed the FLS test. Our system uses an underlying YOLO v8 model to detect the FLS box, the left and right graspers, and the FLS pegs and rings. We then use logic on top of this to detect events such as pick peg from ring, transfer peg between grapsers, and place peg on ring. We are also able to detect if the grasper drops the peg inside or outside the box (the latter being an automatic fail) and if the dropped peg was picked with the correct or wrong grasper. Our system detects these events in realtime without looking into future frames of the video - this means it can give performative feedback to the student as they are performing the task. To evaluate our system we trained it on 6 videos of junior medical residents performing FLS containing several instances of dropping pegs plus 1 video of a fake FLS showing deliberate drops across the board - this is so that the model can learn drops. To evaluate our model we tested it on 14 videos on which our model correctly predicted pass fail on 11 - giving an accuracy of 78.6%. Compared to previous work our system requires only one camera, detects drops inside and outside the FLS box, produces a fully automated pass fail determination, and gives live performative feedback as the student is performing the task - thus informing them of their mistakes in realtime.**

***Index Terms*—YOLO, fundamentals of laparoscopic surgery, automatic evaluation, video AI**

I. Introduction

Medical students preparing for a career in surgery have to demonstrate proficiency in basic surgical skills. Evaluating these skills in the operating room require considerable resources and increases risk of surgical complications [1], [2]. Simulation has emerged as a cost effective and accurate alternative [3]–[6].

The Fundamentals of Laparoscopic Surgery (FLS) simulation kit is designed to train and evaluate students in laparoscopic surgery [7]. While there are many training videos and instructions available online there is no automatic evaluation to help students know if they are performing the tasks correctly. In particular, there is no system that gives automatic and accurate realtime performative feedback.

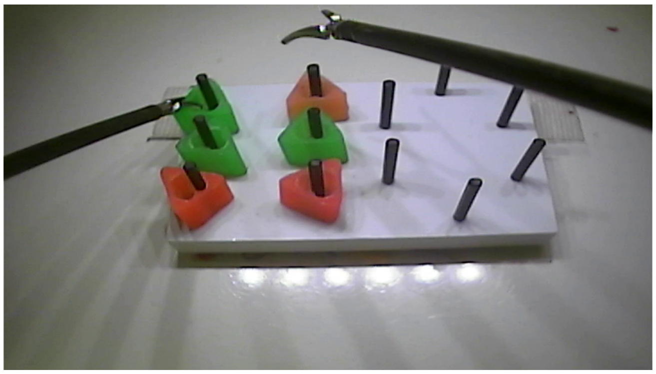
Previous work has seen the use of a YOLO bounding box model [8] to identify different parts of the FLS kit [9]. The authors determine the path and total time to pick a ring from a peg, transfer it to another grasper, and drop the rink to a peg. Their work is mostly analysis and correlates the time the model calculates to the level of the student. Another study uses three cameras and the ResNet50 bounding box model [10] to determine different components of the FLS kit [11]. The authors then use this information to determine the time for a peg transfer. In both previous studies their models do not detect cases when the peg is dropped outside the FLS box. They also do not detect cases where the peg is dropped inside the box and then resumed with the wrong grasper. Their focus is mainly on the transfer time.

Our work in comparison goes further - we aim to give a fully automated pass or fail determination with performative feedback in realtime. Below we describe our data followed by model and experimental results demonstrating the performance of our model.

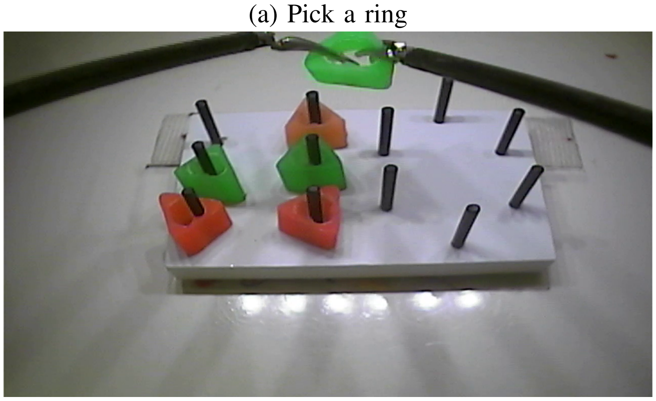
II. METHODS

A. Data

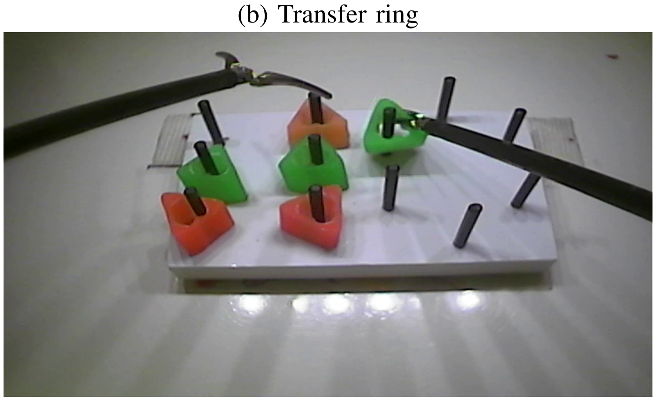
We collected videos of students performing FLS training at the Robert Wood Johnson Hospital in New Brunswick, New Jersey, USA. Below are screen snapshots of the FLS videos showing the left grasper picking a ring (Figure 1(a)), a ring transfer between left and right graspers in (b), and placing a ring on the peg in (c).



(a) Pick a ring



(b) Transfer ring



(c) Place ring on peg

Fig. 1. Examples of picking a ring, transferring a ring, and placing one on the peg

For each video we determined a pass fail ground truth using the following rules set by the medical team at Robert Wood Johnson Hospital.

B. Rules for determining pass or fail

1) Timing Starts when first object is touched and last object is dropped. Total time should not exceed 300s.
2) Count drops outside of the box, if > 0 automatic fail.
3) Count 6 successful consecutive Left to Right Completed tasks. A completed task is defined as grasp1-pick-transfer-grasp2-drop. If n=6, go to next step.
4) For each peg if a drop inside the box is demonstrated it can still count towards a completed task if picked up with grasper 1.
5) Count 6 successful consecutive Right to Left Completed tasks. A completed task is defined as grasp2-pick-transfer-grasp1-drop. If n=6, AND total time $<$ max then pass.
6) For each peg if a drop inside the box is demonstrated it can still count towards a completed task if picked up with grasper 2.

Based on the above rules we label each video with a ground truth pass or fail. In Table I we list the videos that we used to train our model and the test videos to evaluate our system. One of our videos called 'Simulated' contains only examples of drops outside the FLS box so that the model can learn to identify them.

TABLE I
MEDICAL RESIDENT LEVEL, GROUND TRUTH DETERMINATION, AND LENGTH OF EACH FLS TRAINING VIDEO

Resident level	**Ground truth**	**Video length (seconds)**
	Training videos	
1st year	Pass	202
1st year	Fail	97
1st year	Fail	193
2nd year	Fail	158
2nd year	Fail	239
2nd year	Fail	160
Simulated drops	Fail	73
	Test videos	
1st year	Pass	195
1st year	Fail	212
1st year	Fail	151
1st year	Fail	197
2nd year	Fail	120
2nd year	Fail	179
3rd year	Pass	104
3rd year	Pass	185
4th year	Fail	105
4th year	Pass	68
4th year	Fail	80
5th year	Pass	110
5th year	Pass	69
5th year	Pass	89

C. Models

Underlying our AI system is the YOLO model [8] shown in Figure 2. YOLO is a popular model for identifying objects in an image.

We randomly selected a single frame from each second of each video in our training dataset. We then trained the YOLO model with the selected images to identify the two graspers,

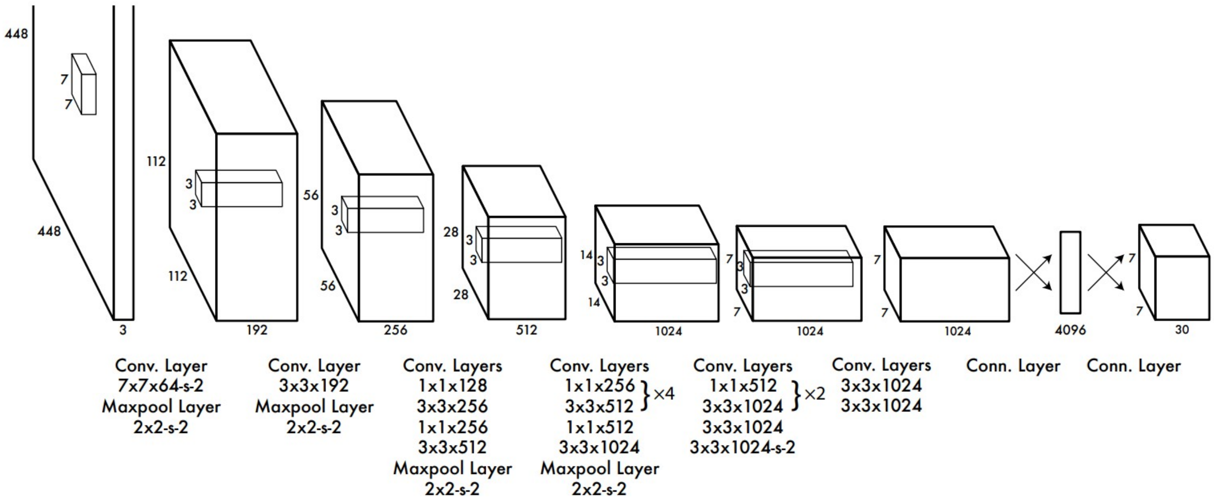


Fig. 2. YOLO model [8]

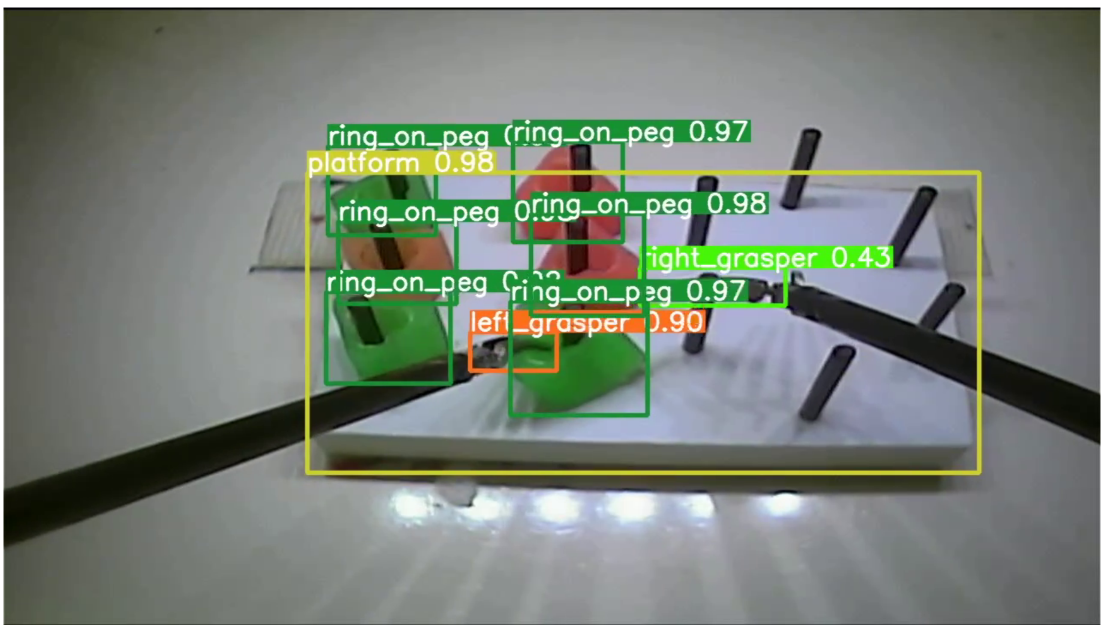


Fig. 3. Identification of FLS box, graspers, rings, and pegs as given by YOLO

the FLS box, all six rings, and all six pegs. See Figure 3 for an example of the YOLO model output.

D. AI system

Our overall approach is to first pass a video through our trained YOLO model. This helps the model understand what a grasper, peg, and ring is. We then apply logic on top of this to follow the rules given earlier (see above Subsection II-B) to determine if a resident passed or failed the FLS peg transfer task. See Figure 4 for an overview of our system.

III. Results

A. Overall metrics

In Table II below we show the ground truth and prediction of each test video as given by our AI system.

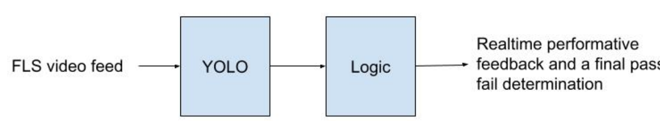


Fig. 4. Overview of our AI system

B. Realtime performative feedback

Our AI system not only gives a pass fail determination but also provides performative feedback and measures the time for each transfer. In Figure 5 we see the system's identification of various actions. The system identified when the grasper picks a ring, it identifies a transfer, and also recognizes a placement

TABLE II
MEDICAL RESIDENT LEVEL, GROUND TRUTH, AND PREDICTION AS GIVEN BY OUR AI SYSTEM

Resident level	**Ground truth**	**Prediction**
	Test videos	
1st year	Pass	Fail
1st year	Fail	Fail
1st year	Fail	Fail
1st year	Fail	Fail
2nd year	Fail	Fail
2nd year	Fail	Fail
3rd year	Pass	Pass
3rd year	Pass	Pass
4th year	Fail	Fail
4th year	Pass	Pass
4th year	Fail	Fail
5th year	Pass	Pass
5th year	Pass	Fail
5th year	Pass	Pass

of the ring.

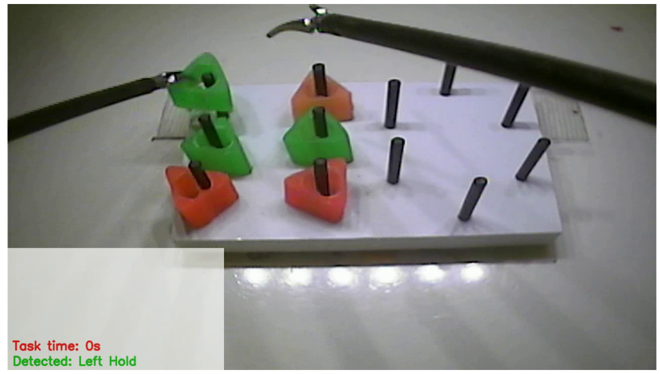
In Figure 6(a) and (b) we see the system outputs the total time for each transfer. In Figure model2(c) we see the model can identify when the ring falls outside the FLS box - this is considered an automatic fail.

IV. DISCUSSION

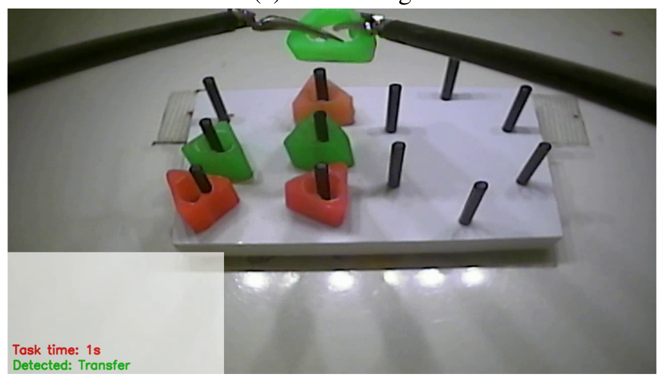
Our model correctly identifies pass fail in 11 of the 14 videos - thus giving an accuracy of 78.6%. In the three videos it gives a wrong determination due to errors made by the model in identifying the action the resident is performing. These mistakes are due to bounding box predictions made by the underlying YOLO model - we expect this to improve as we increase the size of the training set.

Note that our model is trained only on junior residents mostly with fail determination. This is so that the model has exposure examples where the resident drops the ring insides or outside the box. The fact that our model can successfully classify videos of senior residents is an indication of its generalization.

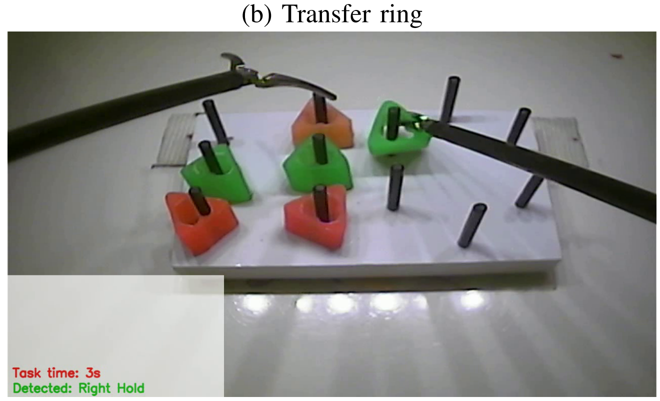
Our system does not identify actions when the resident is simultaneously performing a pick and drop. We will add logic to identify that in a future iteration. Overall our system is the first to give a fully automated pass or fail outcome of the FLS task with realtime performative feedback.



(a) Pick a ring



(b) Transfer ring



(c) Place ring on peg

Fig. 5. Examples of our system identfying the action being performed

REFERENCES

[1] M. Bridges and D. L. Diamond, "The financial impact of teaching surgical residents in the operating room," *The American Journal of Surgery*, vol. 177, no. 1, pp. 28–32, 1999.

[2] R. W. Allen, M. Pruitt, and K. M. Taaffe, "Effect of resident involvement on operative time and operating room staffing costs," *Journal of Surgical Education*, vol. 73, no. 6, pp. 979–985, 2016.

[3] V. N. Palter and T. P. Grantcharov, "Individualized deliberate practice on a virtual reality simulator improves technical performance of surgical novices in the operating room," *Journal of the American College of Surgeons*, vol. 213, no. 3, p. S126, 2011.

[4] S. R. Dawe, G. Pena, J. A. Windsor, J. Broeders, P. C. Cregan, P. J. Hewett, and G. J. Maddern, "Systematic review of skills transfer after surgical simulation-based training," *Journal of British Surgery*, vol. 101, no. 9, pp. 1063–1076, 2014.

[5] B. Zendejas, D. A. Cook, J. Bingener, M. Huebner, W. F. Dunn, M. G. Sarr, and D. R. Farley, "Simulation-based mastery learning improves patient outcomes in laparoscopic inguinal hernia repair: a randomized controlled trial," *Annals of surgery*, vol. 254, no. 3, pp. 502–511, 2011.

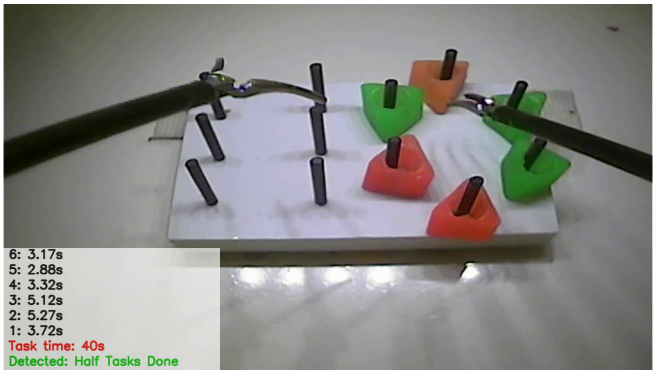
[6] T. Cox, N. Seymour, and D. Stefanidis, "Moving the needle: simulation's impact on patient outcomes," *Surgical Clinics*, vol. 95, no. 4, pp. 827–838, 2015.

[7] N. J. Soper and G. M. Fried, "The fundamentals of laparoscopic surgery: its time has come," *Bull Am Coll Surg*, vol. 93, no. 9, pp. 30–32, 2008.

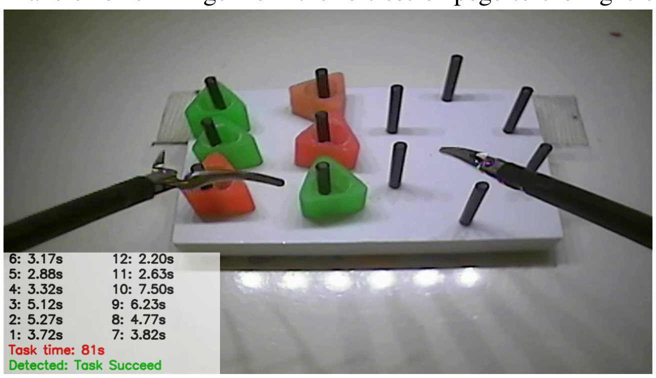
[8] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.

[9] A. Lazar, G. Sroka, and S. Laufer, "Automatic assessment of performance in the fls trainer using computer vision," *Surgical Endoscopy*, pp. 1–7, 2023.

[10] 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.



(a) Transfer of six rings from the left set of pegs to the right ones



(b) Transfer of six rings back to the left set of pegs from the right ones

(c) Drop outside the FLS box

Fig. 6. Examples of our system identifying the total time for transfers and identifying a drop outside the box.

[11] F. R. Fathabadi, J. L. Grantner, S. A. Shebrain, and I. Abdel-Qader, “Surgical skill training and evaluation for a peg transfer task of a three camera-based laparoscopic box-trainer system,” in *2021 International Conference on Computational Science and Computational Intelligence (CSCI)*. IEEE, 2021, pp. 1146–1151.