DOI: 10.4274/tjcd.galenos.2025.2025-6-6
Turk J Colorectal Dis 2025;35(4):102-108

Enhancing Surgical Performance Through Automated Video Analysis Utilizing Computer Vision and Machine Learning

Alaa El-Hussuna[1], Muhammed Elhadi[1,2], Andreas Møgelmose[3], Hanan Aljuaid[4]

[1]OpenSourceResearch Collaboration, Aalborg, Denmark
[2]Houston Methodist Hospital, Clinic of Colon and Rectal Surgery, Houston, USA
[3]Aalborg University Faculty of Engineering, Department of Electronic Systems, Aalborg, Denmark
[4]Princess Nourah bint Abdulrahman University, College of Computer and Information Sciences, Department of Computer Sciences, Riyadh, Saudi Arabia

ABSTRACT

Aim: Recent advancements in computer vision have enabled the development of automated systems that can assess surgeons' expertise with high accuracy using automated performance metrics (APMs). This study aims to evaluate and enhance surgical performance through the use of APMs.

Method: This is a prospective, quality-control, multicenter international cohort study. The primary outcome is the improvement of APMs extracted from two-dimensional laparoscopic or robotic colorectal procedure video films after feedback to the surgeons. The secondary outcome is the development of new metrics to measure the model's performance beyond simple accuracy. The collaborators will send 2-3 real-world video films of colorectal procedures they have performed. They will then receive feedback on their films, including an APM data analytics report. After the feedback, the collaborators will send 2-3 videos of the same colorectal procedures. Data analysis of APMs comparing pre- and post-feedback operations will follow.

Conclusion: The study will enable efficient training programs within constrained working hours and address heightened ethical considerations regarding patient safety. Moreover, the training of surgeons in low- and middle-income countries will benefit from the results of this study, as they can improve their skills without the need to spend months to years training in developed countries.

Keywords: Automated performance metrics, training, surgery, performance, computer vision

Introduction

The accurate evaluation of surgical trainees' performance is essential for surgical training (i.e., acquiring surgical skills) and serves as a key component of proficiency-based training (i.e., mastering surgical skills).[1] To develop their skills, surgeons must regularly perform procedures under supervision. However, the growing complexity of modern healthcare, restrictions on working hours, and ethical concerns related to patient safety necessitate the development of efficient training programs that protect patients. Such programs should facilitate automated, objective, and data-driven assessments of surgical skills while offering meaningful feedback.[2] Recent shifts in surgical training, including self-directed learning and reflective practice, highlight the benefits of repetitive and independent practice, which have been enabled by objective evaluation tools.[3]

The potential for bias in surgical skill assessment has been widely debated in various studies.[4-6] A data-driven approach can provide an objective evaluation method, minimizing bias in assessing surgical proficiency (see the appendix for further in-depth discussion). Current methods for evaluating technical skills include task-specific checklists, global rating scales,

Address for Correspondence: Alaa El-Hussuna
OpenSourceResearch Collaboration, Aalborg, Denmark
E-mail: contact@osrc.network **ORCID ID:** orcid.org/0000-0002-0070-8362

Received: 25.06.2025 **Accepted:** 03.08.2025 **Publication Date:** 22.09.2025

Cite this article as: El-Hussuna A, Elhadi M, Møgelmose A, Aljuaid H. Enhancing surgical performance through automated video analysis utilizing computer vision and machine learning. Turk J Colorectal Dis. 2025;35(4):102-108

and technology-based performance measures. Although observer-based scoring metrics are cost-effective and easily accessible, they are prone to bias and can be time-consuming to implement. In contrast, technology-based performance measures offer a unique opportunity for detailed, automated, and objective assessments,[7] which can be integrated into the digital platforms connected to laparoscopic and robotic workstations.

A systematic review examining the objective assessment of robotic surgical techniques across different specialties[8] has revealed that manual and automated tools, such as the Objective Structured Assessment of Technical Skills and the Global Evaluative Assessment of Robotic Skills, still carry potential subjective bias. However, automated assessment tools, which utilize data from robotic workstations, provide more objective and comprehensive evaluations. The review highlights that a key issue is the lack of a universally accepted standard for assessment, resulting in variability in the focus, application, and effectiveness of existing tools.

Recent advancements in computer vision have enabled the development of automated systems capable of assessing surgeons' expertise with high accuracy using automated performance metrics (APMs). Studies have demonstrated that experts considerably outperform novices in areas such as instrument length, bimanual dexterity, instrument idle time, camera path length, and camera movements. Similar distinctions have also been observed between super experts and experts.[8] APMs may offer a more comprehensive and objective evaluation of a surgeon's skills than expert evaluators. However, most studies on APMs (Appendix 1 and Appendix 2) are based on small sample sizes, lack diversity in training datasets, and have no or limited validation datasets. There is a need to investigate the benefits of existing APMs using large, diverse, real-world video datasets.

This study aims to enhance the evaluation and improvement of surgical performance in colorectal procedures by using APMs extracted from laparoscopic and robotic surgical video analysis.

Materials and Methods

Designing the Study

The European Society of Coloproctology (ESCP) has successfully conducted many international prospective audits.[9,10] This study was presented during the annual conference of the ESCP in Thessaloniki at the cohort studies session on Wednesday, September 25, 2024. The study design, including the type of index procedure, the time interval between the index and the next procedure, how many procedures are expected between them, how data can be transferred, and other design-related questions, was then discussed with the audience. The audience then voted on these issues using the ESCP mobile phone application. The design of this study is based on these discussions and the subsequent voting.

To see this session and the voting, use this link:

https://vimeo.com/escp/review/1033584541/e8a4b81d1d

This is a prospective, randomized, multicenter international cohort study. The participants (surgeons) will be randomly assigned to one of two groups to ensure comparability and minimize selection bias. Group 1 will receive feedback based on video analyses of their performance, and Group 2 will serve as the control group and will not receive feedback. Randomization will be conducted using a computer-generated sequence, with allocation concealed until assignment.

The study will compare the same types of cases performed by the same surgeon over time to monitor whether feedback improves the surgeon's performance.

Primary Outcome

The primary outcome is the improvement of surgical performance, measured by improvement in APMs. APMs will be extracted from two-dimensional laparoscopic or robotic colorectal procedure video films after feedback to the surgeons.

Secondary Outcome

The secondary outcome is the measurement of the model's performance beyond simple accuracy, including the assessment of APMs using large, diverse, real-world video datasets.

Inclusion and Exclusion Criteria

The inclusion criteria are two-dimensional, real-world surgical video films recorded during elective laparoscopic or robotic colorectal procedures. Both procedures used for training and those not intended for training will be included.

Selection of the Colorectal Procedures

The authors' choice of colorectal procedures is a pragmatic one aimed at obtaining a homogeneous group of surgical procedures, enabling knowledge transfer from common to more complex procedures and promoting data efficiency. By selecting different colorectal procedures, the algorithm's applicability in medical practice and the scalability of the networks will be greatly improved.

Only elective curative procedures will be included, as the emergency setting may be affected by multiple factors that could introduce noise into the interpretation. Procedures in which conversion from the original plan (laparoscopic or robotic) to an open procedure occurs will also be included to train the algorithm to recognize non-progression in the surgical procedure.

The following index colorectal procedures will be included:

- Ileo-caecal and ileocolic resections
- Right hemicolectomy and extended right hemicolectomy, as defined in the ESCP 2015 audit[9]

- Left hemicolectomy and sigmoid colon resections, as defined in the ESCP 2017 audit[10,11]

These colorectal procedures are usually performed by supervised trainees and consultant surgeons. The procedures will be included regardless of indication (benign or malignant), provided that they are intended to be curative. There is no need for special adjustment for case mix or surgical complexity, as the surgeon will choose 2-3 video films of a procedure performed by them, followed by 2-3 video films of the same procedure after receiving APM-based feedback.

Appendix 3 shows the clinical report form (CRF) that will be attached to each film. This CRF has been kept simple to ensure basic information is provided for each procedure.

Quality Control of the Video Films

The video data will be recorded in high definition with a resolution of 1920 × 1080 pixels. However, 1280 × 720 pixels will be accepted in centers that cannot provide higher-resolution films.

Unedited video films will undergo quality control checks. Two authors will review the footage for overall quality, including blurriness, lack of focus, loss of fine details, stability, color accuracy, exposure, and clarity.

Phase Definition

The phase definitions for each laparoscopic or robotic colorectal procedure will follow the recommendations of leading international surgical societies, if available. For procedures with no well-defined phases, at least three expert surgeons will be consulted to define the procedure phases.

Preprocessing and De-identification of Surgical Video Data

Data and video files will be uploaded to and stored on a secure server provided by Aalborg University. To comply with data privacy regulations, the data will first be de-identified, with the removal of all metadata. Metadata includes all patient data, the date, time, and location of the operation, as well as information about the operating staff. No off-the-shelf solutions exist for such a setting, so tools tailored for this study will be developed.

Annotation of Surgical Video Data

The annotation process will encompass two key tasks: identification of surgical phases and tool positions. These annotations are critical for subsequent model training and analysis of APMs, and as such, rigorous procedures will be followed to ensure consistency and reliability.

Annotator Roles and Tools

- Surgical phase annotation will be performed by surgical residents in their 4th year or higher of clinical training, who possess adequate familiarity with the procedural workflow and phase definitions.
- Tool position annotation, which is comparatively more mechanical and less reliant on clinical judgment, will be conducted by trained undergraduate or graduate student assistants.

Annotations Will Be Conducted Using Established Tools Such as

- V7 (https://www.v7labs.com),
- Computer Vision Annotation Tool (https://www.cvat.ai/)
- Labelbox (https://labelbox.com)

All annotations will be exported and stored in multiple object tracking format,[12] a standard format suitable for downstream analysis and model training.

Annotator Training

All annotators will undergo a structured training program, including:

- A 2-hour initial training session introducing the annotation platform, guidelines, and tasks
- Annotated examples and a reference manual defining surgical phases and annotation criteria
- A pilot annotation set of 3-5 videos to be annotated during training, followed by a feedback session with an expert reviewer (a board-certified surgeon or senior research fellow)
- Certification, requiring annotators to achieve ≥85% agreement with expert labels on a pilot set before contributing to the main dataset

Inter-rater Reliability and Adjudication

To ensure consistency:

- A subset of 20% of the videos will be annotated independently by at least two annotators.
- Inter-rater reliability will be calculated using Cohen's kappa for phase annotations and Intersection over Union (IoU) for tool position bounding boxes.
- A minimum acceptable kappa score of 0.75 and IoU ≥0.5 will be enforced, and discrepancies will trigger review.

Although manual annotation inherently carries a degree of subjectivity -particularly in complex tasks such as surgical phase recognition- our protocol is specifically designed to minimize variability. Structured training, expert-reviewed feedback, a certification threshold, and ongoing inter-rater reliability checks help ensure consistency and mitigate annotator bias.

Quality Control and Adjudication

- A surgical expert will review a random sample of 10% of the annotations to validate correctness and completeness.
- If systematic errors or deviations are found, affected annotations will be re-reviewed or corrected.

- A weekly consensus meeting involving annotators and supervising experts will be held to discuss edge cases and update annotation guidelines as needed.
- A detailed annotation logbook will be maintained for each video, documenting the annotator ID, timestamp, tools used, and any issues or anomalies.

These measures will ensure that the dataset used for downstream machine learning (ML) model training is robust, consistent, and clinically reliable.

Study Stages

Stage One

A prospective audit will be conducted, in which laparoscopic and robotic real-world video films of colorectal procedures are collected from international collaborating centers. Any consultant or trainee can participate in the study. The collaborators will send 2-3 real-world video films of colorectal procedures that they have performed. The collaborators may choose how many procedures to send for analysis, provided that they submit at least two films of the same procedure.

Stage Two

The collaborators will then receive feedback on their films, including an APMs data analytics report. Based on the APMs report, the participating surgeon will receive an objective, data-driven technical assessment of performance adjusted for case difficulty. The confidential, password-protected report will highlight areas for performance improvement. This report will include the types of assessed APMs, their interpretation, and suggestions for improving performance. The report will be generated in a standardized format by the study team after a short pilot assessment.

Stage Three

The collaborators will send 2-3 videos of the same colorectal procedures that they performed in stage one after receiving the data analytics. These follow-up procedures must be performed within 6 months of the first (index) procedure. At least 10 procedures must be performed after the index procedure.

Stage Four

Data analysis of APMs will compare pre- and post-feedback operations. The collaborators will receive a detailed feedback report upon request. The confidential, password-protected report will include APMs from the two sets of films collected in stages one and three.

Data Collection

With a vast network of surgeons in the OpenSourceResearch Collaboration, ESCP, and American Society of Colon and Rectal Surgeons, it is expected that a large, generalizable, and diverse dataset will be obtained, which can be used to train the model.

Statistical Analysis

The data analysis consists of two parts:

- Extraction of surgical phases and tool tracks
- Computation of APMs from extracted data

Extraction of Surgical Phases and Tool Tracks

No off-the-shelf solutions for computing APMs exist, so a custom algorithm will be developed. Inspiration can be found in earlier approaches.[13,14] Even so, improved results should be obtainable using more modern transformer-based methods, including action recognition methods,[15] such as ASFormer surgical phase detection; object detection methods,[16] such as DEtection TRansformer[17] for tool detection; and CoTracker[18] for surgical tool tracking.

Fine-tuning of pre-trained algorithms will be leveraged to the fullest possible extent, but substantial amounts of training data must be manually annotated, as outlined in the previous section.

Accuracy will be reported using standard metrics: Mean over Frames[15] for surgical phase detection and Higher Order Tracking Accuracy[19] for tool tracking.

Computations of Automated Performance Metrics From Extracted Data

Computation of APMs will be performed using custom methods for each metric. All APMs are well-defined by mathematical formulas (Appendix 2), so no ML is required for this stage. It may be interesting to test an ML-based surgeon rater using the raw extracted data and compare its performance with the predefined APMs. However, that exercise is left for future work.

Evaluation

Apart from using standard evaluation metrics as mentioned above, an important aspect of modeling is out-of-sample validation, which involves partitioning the data into training, validation, and test sets-usually in a 70%:10%:20% split or similar. This project will follow this standard procedure for the computer vision field. If the amount of annotated data is insufficient to allow for such a split, N-fold cross-validation will be performed.

Statistical Analysis

Sample Size Calculation

Assuming a 20% change in APMs, the sample size was calculated as follows:

- Z is the Z-score corresponding to the desired confidence level (for a 95% confidence level, Z≈1.96).
- p is the estimated proportion at baseline (or for the control group).
- E is the margin of error expressed as a proportion (20%=0.2).
- p’ is the desired percentage change expressed as a decimal.

The sample size used the given values:

- Population size (N)=1,000
- Confidence level=95% ($Z \approx 1.96$)
- Margin of error=20% (0.2)
- Desired percentage change=20% (0.2)

The following formula was used: $n=(1.96^2 \times p \times [1 - p]) / (0.2^2 \times [p \times (1 + 0.2)])$

We can assume a conservative estimate of 0.5 for p:

$n=(1.96^2 \times 0.5 \times [1 - 0.5]) / (0.2^2 \times [0.5 \times (1 + 0.2)])$

$n=(3.8416 \times 0.25) / (0.04 \times 0.6)$

$n \approx 10.1056 / 0.024$

$n \approx 421.0667$

Based on this calculation, a sample size of approximately 422 individuals is needed to detect a 20% change in APMs.

Given the lack of prior evidence on which APMs are most responsive to feedback, this study adopts an exploratory approach, assessing multiple performance domains without designating a single primary outcome. A 20% relative improvement in any APM will be considered meaningful in this context.

Adjusting for Case Complexity

To address potential confounders related to case complexity, key patient-level variables that may influence surgical performance will be collected for each patient undergoing surgery. During the analysis phase, statistical methods such as multivariable regression or propensity score matching will be used to adjust for differences in case complexity between groups. This approach will help isolate the effect of feedback on the surgeon's performance, minimizing the risk of confounding due to patient-related factors. It allows for accounting for case complexity while maintaining the integrity of randomization and minimizing bias.

Analysis Plan

This study includes within-subject pre- and post-feedback comparisons for surgical performance, alongside classifier evaluation tasks for tool detection and phase recognition. Analyses will be conducted using SPSS and/or Python statistical libraries (e.g., SciPy, scikit-learn).

1. Descriptive Statistics

- Summary statistics (mean, median, standard deviation, range) will be provided for continuous variables, and frequencies and proportions will be provided for categorical variables.

2. Analysis of APMs

- For pre- vs. post-feedback comparisons within surgeons, paired t-tests (for normally distributed APMs) and Wilcoxon signed-rank tests (for non-parametric data) will be used.
- For comparing multiple time points or groups, repeated measures analysis of variance or linear mixed-effects models will be applied, allowing for both fixed effects (e.g., feedback, session) and random effects (e.g., surgeon ID).

3. Evaluation of Classifier Performance (Tool Detection and Phase Recognition)

- Receiver operating characteristic curves and the area under the curve (AUC) will be computed for binary classification tasks, including tool presence detection (whether a specific tool is in use at a given time) and phase classification performance (correct classification of surgical phase per video frame).
- For multi-class phase classification, macro- and micro-averaged AUCs will be reported.
- Precision, recall, F1-score, and confusion matrices will also be presented to provide a comprehensive evaluation of classification performance.

4. Handling of Missing or Ambiguous Data

- Incomplete annotations or ambiguous cases will be flagged and excluded from the primary analysis, but they may be included in sensitivity analyses.
- Multiple imputation will be considered if missing data exceeds 5% in any analytic subset.

5. Statistical Significance

- A two-sided p-value of <0.05 will be considered statistically significant. Where applicable, 95% confidence intervals will be reported alongside effect sizes.

Ethical Considerations

All data collected will reflect current practice, with no changes made to planned treatment pathways. As such, this study should be registered as an audit of current practice at each participating center. The local team at each site is responsible for ensuring that local audit approval (or equivalent) is obtained. Participating centers will be asked to confirm that they have received formal approval at their sites. Patients' consent to use the videos for research purposes will be obtained, including consent for the de-identified videos to be used in future studies without additional consent.

Discussion

This protocol presents a novel approach to surgical education and performance assessment, utilizing advanced computer vision and ML technologies. By focusing on APMs derived from laparoscopic and robotic surgical videos, the study aims to improve surgical training for trainees and enhance performance for specialists.

This approach is particularly relevant in the context of modern healthcare's evolving complexities, including the need for efficient training programs within constrained working hours and heightened ethical considerations around patient safety. The training of surgeons in low- and middle-income countries (LMICs) will benefit from the results of this study. If improvements in APMs lead to improved performance, surgeons from LMICs can enhance their skills without the need to spend months or years training in developed countries.

Real-world data on surgeons' performance can personalize training in precise and productive ways. Guided by surgical educators, ML models can identify performance qualities not necessarily evident to experienced trainers, potentially leading to more rapid skill acquisition. Automated surgical phase recognition is a foundational step for other applications that can create informative and focused educational material for students and residents.

Challenges such as non-static cameras resulting in abrupt viewpoint changes, inconsistent organs and instruments, variations in illumination, unfocused frames, and the presence of blood and smoke in the surgical field can be addressed through iterative refinement of the models to improve image analysis.

Perspectives

In the future, APMs might be correlated with different post-operative outcomes (functional, oncological, patient-reported outcomes, etc.), opening a new era in surgical research as objective measures are integrated with clinical assessments and patient-reported outcome measures. An artificial intelligence system capable of recognizing surgical phases may be used for numerous tasks, including quality measurement, adverse event recording and analysis, education, statistics, and surgical performance evaluation.[20]

High-volume simulator training based on real procedures will be possible, as anonymized procedures can be transformed into surgical simulators for training and experimentation with innovative modifications to traditional techniques. The efficiency of producing surgical reports is an additional benefit.

For hospital administration, operating room scheduling is challenging because pre-operative estimates of procedure duration are often inaccurate. This inaccuracy stems from considerable variability in how procedures unfold. Real-time information about the progress of surgeries is crucial for effectively adjusting the daily operating room schedule. Ideally, this information should be objective, automatically accessible, and available in real time to predict the remaining duration of surgeries. Such data would enable optimal planning and utilization of operating theatre resources, ensuring they are used to their fullest capacity.[20]

Ethics

Ethics Committee Approval: Not applicable.

Informed Consent: Not applicable.

Acknowledgement

Princess Nourah bint Abdulrahman University Researchers Supporting Project number PNURSP2025R54, Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

The OSRC is an international research organization with a particular focus on implementing information technologies and AI in medical research. The role of the OSRC in this project was to recruit researchers and coordinate research work. More information about this organization is available on its website: https://www.osrc.network.

Footnotes

Authorship Contributions

Concept: A.E.H., A.M., Design: A.E.H., A.M., H.A., Data Collection or Processing: A.E.H., A.M., Analysis or Interpretation: A.E.H., A.M., Literature Search: A.E.H., M.E., A.M., Writing: A.E.H., M.E., A.M., H.A.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: Princess Nourah bint Abdulrahman University Researchers Supporting Project Number (PNURSP2025R54).

REFERENCES

1. Buckley CE, Kavanagh DO, Nugent E, Ryan D, Traynor OJ, Neary PC. Zone calculation as a tool for assessing performance outcome in laparoscopic suturing. Surg Endosc. 2015;29:1553-1559.
2. Ebina K, Abe T, Hotta K, Higuchi M, Furumido J, Iwahara N, Kon M, Miyaji K, Shibuya S, Lingbo Y, Komizunai S, Kurashima Y, Kikuchi H, Matsumoto R, Osawa T, Murai S, Tsujita T, Sase K, Chen X, Konno A, Shinohara N. Automatic assessment of laparoscopic surgical skill competence based on motion metrics. PLoS One. 2022;17:e0277105.
3. Ganni S, Botden SMBI, Chmarra M, Li M, Goossens RHM, Jakimowicz JJ. Validation of motion tracking software for evaluation of surgical performance in laparoscopic cholecystectomy. J Med Syst. 2020;44:56.
4. Helliwell LA, Hyland CJ, Gonte MR, Malapati SH, Bain PA, Ranganathan K, Pusic AL. Bias in surgical residency evaluations: a scoping review. J Surg Educ. 2023;80:922-647.
5. Alimi Y, Bevilacqua LA, Snyder RA, Walsh D, Jackson PG, DeMaria EJ, Tuttle JE, Altieri MS. Microaggressions and implicit bias in surgical training: an undocumented but pervasive phenomenon. Ann Surg. 2023;277:e192-e196.
6. Fitzgerald C, Hurst S. Implicit bias in healthcare professionals: a systematic review. BMC Med Ethics. 2017;18:19.
7. D'Angelo AL, Rutherford DN, Ray RD, Laufer S, Kwan C, Cohen ER, Mason A, Pugh CM. Idle time: an underdeveloped performance metric for assessing surgical skill. Am J Surg. 2015;209:645-651.
8. Chen J, Cheng N, Cacciamani G, Oh P, Lin-Brande M, Remulla D, Gill IS, Hung AJ. Objective assessment of robotic surgical technical skill: a systematic review. J Urol. 2019;201:461-469.

9. 2015 European Society of Coloproctology collaborating group. The relationship between method of anastomosis and anastomotic failure after right hemicolectomy and ileo-caecal resection: an international snapshot audit. Colorectal Dis. 2017.

10. ESCP Cohort Studies and Audits Committee. The 2017 European Society of Coloproctology (ESCP) international snapshot audit of left colon, sigmoid and rectal resections - study protocol. Colorectal Dis. 2018;20(Suppl 6):5-12.

11. Dendorfer P, Rezatofighi H, Milan A, Shi J, Cremers D, Reid I, Roth S, Schindler K, Leal-Taixe L. CVPR19 tracking and detection challenge: How crowded can it get? 2019; Available from: http://arxiv.org/abs/1906.04567

12. Kitaguchi D, Takeshita N, Matsuzaki H, Takano H, Owada Y, Enomoto T, Oda T, Miura H, Yamanashi T, Watanabe M, Sato D, Sugomori Y, Hara S, Ito M. Real-time automatic surgical phase recognition in laparoscopic sigmoidectomy using the convolutional neural network-based deep learning approach. Surg Endosc. 2020;34:4924-4931.

13. Eckhoff JA, Ban Y, Rosman G, Müller DT, Hashimoto DA, Witkowski E, Babic B, Rus D, Bruns C, Fuchs HF, Meireles O. TEsoNet: knowledge transfer in surgical phase recognition from laparoscopic sleeve gastrectomy to the laparoscopic part of Ivor-Lewis esophagectomy. Surg Endosc. 2023;37:4040-4053.

14. Ding G, Sener F, Yao A. Temporal action segmentation: an analysis of modern techniques. IEEE Trans Pattern Anal Mach Intell. 2024;46:1011-1030.

15. Zou Z, Chen K, Shi Z, Guo Y, Ye J. Object detection in 20 years: a survey. Proceedings of the IEEE. 2023;111:257-276.

16. Dai Z, Cai B, Lin Y, Chen J. UP-DETR: Unsupervised pre-training for object detection with transformers. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2021.

17. Karaev N, Rocco I, Graham B, Neverova N, Vedaldi A, Rupprecht C. CoTracker: It is better to track together. 2023. Available from: http://arxiv.org/abs/2307.07635

18. Luiten J, Os Ep AA, Dendorfer P, Torr P, Geiger A, Leal-Taixé L, Leibe B. HOTA: a higher order metric for evaluating multi-object tracking. Int J Comput Vis. 2021;129:548-78.

19. Golany T, Aides A, Freedman D, Rabani N, Liu Y, Rivlin E, Corrado GS, Matias Y, Khoury W, Kashtan H, Reissman P. Artificial intelligence for phase recognition in complex laparoscopic cholecystectomy. Surg Endosc. 2022;36:9215-9223.

20. Guédon ACP, Meij SEP, Osman KNMMH, Kloosterman HA, van Stralen KJ, Grimbergen MCM, Eijsbouts QAJ, van den Dobbelsteen JJ, Twinanda AP. Deep learning for surgical phase recognition using endoscopic videos. Surg Endosc. 2021;35:6150-6157.

Appendix 1-3 Link.
https://d2v96fxpocvxx.cloudfront.net/8a9ff4da-541a-42fa-9980-1a9a3ab6d6c5/content-images/54657bfa-ee96-494b-bf09-301400fde7e3.pdf