International Journal of Computer Assisted Radiology and Surgery (2021) 16:2045–2054
https://doi.org/10.1007/s11548-021-02434-w

ORIGINAL ARTICLE

Object and anatomical feature recognition in surgical video images based on a convolutional neural network

Yoshiko Bamba[1] · Shimpei Ogawa[1] · Michio Itabashi[1] · Hironari Shindo[2] · Shingo Kameoka[3] · Takahiro Okamoto[4] · Masakazu Yamamoto[1]

Received: 5 February 2021 / Accepted: 17 June 2021 / Published online: 24 June 2021

Abstract

Purpose Artificial intelligence-enabled techniques can process large amounts of surgical data and may be utilized for clinical decision support to recognize or forecast adverse events in an actual intraoperative scenario. To develop an image-guided navigation technology that will help in surgical education, we explored the performance of a convolutional neural network (CNN)-based computer vision system in detecting intraoperative objects.

Methods The surgical videos used for annotation were recorded during surgeries conducted in the Department of Surgery of Tokyo Women's Medical University from 2019 to 2020. Abdominal endoscopic images were cut out from manually captured surgical videos. An open-source programming framework for CNN was used to design a model that could recognize and segment objects in real time through IBM Visual Insights. The model was used to detect the GI tract, blood, vessels, uterus, forceps, ports, gauze and clips in the surgical images.

Results The accuracy, precision and recall of the model were 83%, 80% and 92%, respectively. The mean average precision (mAP), the calculated mean of the precision for each object, was 91%. Among surgical tools, the highest recall and precision of 96.3% and 97.9%, respectively, were achieved for forceps. Among the anatomical structures, the highest recall and precision of 92.9% and 91.3%, respectively, were achieved for the GI tract.

Conclusion The proposed model could detect objects in operative images with high accuracy, highlighting the possibility of using AI-based object recognition techniques for intraoperative navigation. Real-time object recognition will play a major role in navigation surgery and surgical education.

Keywords Image-guided navigation technology · Surgical education · Convolutional neural network · Computer vision · Object detection

Introduction

Optimal and safe surgical methods and effective surgical education for young surgeons are challenges in surgical practice. Surgical techniques in open surgery have been considered tacit knowledge and are not available on storage devices. Digitizing surgical techniques using the latest technology is expected to play a major role in surgical evaluation and education.

With rising costs and lack of resources, the medical community is facing a challenge in providing medical practitioners with high-quality training materials. Dealing with the inadequacy of the training process, young surgeons as well as experts are relying more on alternative preparatory resources, such as surgical videos, to develop and improve their skills [1, 2]. Though the utility of video recordings is proven, their manual annotation and analysis require considerable experience, take a relatively long time and are associated with a high cost [3, 4]. Moreover, the handcrafted method cannot work with high performance on raw, preprocessed samples. With the advent of artificial intelligence

✉ Yoshiko Bamba
bamba.yoshiko@twmu.ac.jp

[1] Department of Surgery, Institute of Gastroenterology, Tokyo Women's Medical University, 8-1, Kawadacho Shinjuku-ku, Tokyo 162-8666, Japan

[2] Otsuki Municipal Central Hospital, Yamanashi, Japan

[3] Ushiku Aiwa Hospital, Ibaraki, Japan

[4] Department of Breast Endocrinology Surgery, Tokyo Women's Medical University, Tokyo, Japan

(AI), a shift in workflow and productivity in the medical field has begun and surgical practices and education stand to gain from the current technological revolution [4, 5]. Several groups have demonstrated the feasibility of different AI-based automatization approaches for video and medical image analysis for varying purposes, such as recognition of operative steps, identifying and tracking surgical tools and diagnosis [6–8]. As AI-enabled approaches can process huge amounts of surgical data, they can be used to recognize or predict adverse events, enable "navigation in surgery" by addressing various anatomical orientation questions and important decision-making tools and contribute to training and education. [9].

Computer vision (CV) includes the study of machine-mediated understanding of images. It includes image acquisition and interpretation and has been explored in areas such as image-guided diagnosis and surgery or virtual colonoscopy [10]. However, the success of medical image analysis remains limited by large variations in occlusions, viewpoints and lighting conditions during surgical processes. In the field of CV, deep learning technology substantially improved the traditional machine learning process [11]. The convolutional neural network (CNN), a prominent representative network of models in the field of deep learning, is gaining importance in current medical image processing, recognition and classification [12–14]. A review of the literature indicates a requirement for further studies on machine learning applications for intraoperative image analysis.

To explore the possibility of AI-driven applications in surgical education, our group developed an image-guided surgical navigation technology. This study investigated whether CNN-based CV could be utilized for efficient detection of both specific anatomical features and surgical tools during surgery.

Methods

Institutional approval

All datasets were deidentified, and the study protocol was exempt from institutional review board review at Tokyo Women's Medical University.

Datasets

The surgical videos used for annotation were recorded during surgeries carried out in the Department of Surgery, Institute of Gastroenterology at Tokyo Women's Medical University from January 2019–August 2020. Abdominal endoscopic images were cut out from 9 manually captured surgical videos for the training model, and additional images were cut out from other videos for validation (Fig. 1). The images varied in nature, representing different surgeries (colorectomy, rectal surgery, hernia, sigmoid resection), and duplicate images were excluded from the assessment. Any frame from a video in the training set was excluded from the test set. The images were manually annotated one by one by marking each visible tool or anatomical feature. During the annotation process, polygons were drawn, delimiting each object or anatomical feature in every video image. During the training process, every polygon signified a foreground mask and the rest of the image represented the background. The annotations were validated by experts in the field.

A total of 1070 images were cut out for training in an object recognition model, using IBM Visual Insights (Power SystemAC922), which includes 400 images from 2 right colorectomies, 510 images from 4 rectal surgeries, 110 images from 2 hernia surgeries and 50 images from 1 sigmoid resection surgery. Eight objects were selected for this annotation. The objects and the numbers of each object annotated in the images were as follows: GI tract, 1781; port, 861; forceps, 1873; gauze, 1016; vessels, 352; blood, 208; clips, 760; and uterus, 63 (Table 1 and Fig. 2 a, b). Instead of using similar images, we used a wide variety of images from various situations when selecting items for both training and validation. The model was deployed, and the other 200 images were used as input in the deployed model to verify its diagnostic accuracy. A surgical video with a 40 s run time was extracted from the other videos and used to verify the model.

Deep neural network training for automated object identification.

Deep neural networks are the most effective available techniques for solving object detection and instance segmentation tasks. In this study, an open-source programming framework for CNN was used to design a model that could recognize and segment objects in real time. The model was trained by and implemented through IBM Visual Insights (Power SystemAC922).

Analysis

IBM Visual Insights includes the most popular open-source deep learning frameworks and tools. The model types included in the software are GoogLeNet, Faster R-CNN, tiny YOLO V2, YOLO V3, Detectron, Single Shot Detector (SSD) and structured segment network (SSN). The model is built for easy and rapid deployment. Moreover, it also provides complete end-to-end workflow support for CV deep learning models that includes complete lifecycle management from installation and configuration, data labeling and model training, to inference and moving models into production. The default value of the ratio is 80/20, resulting in 80%

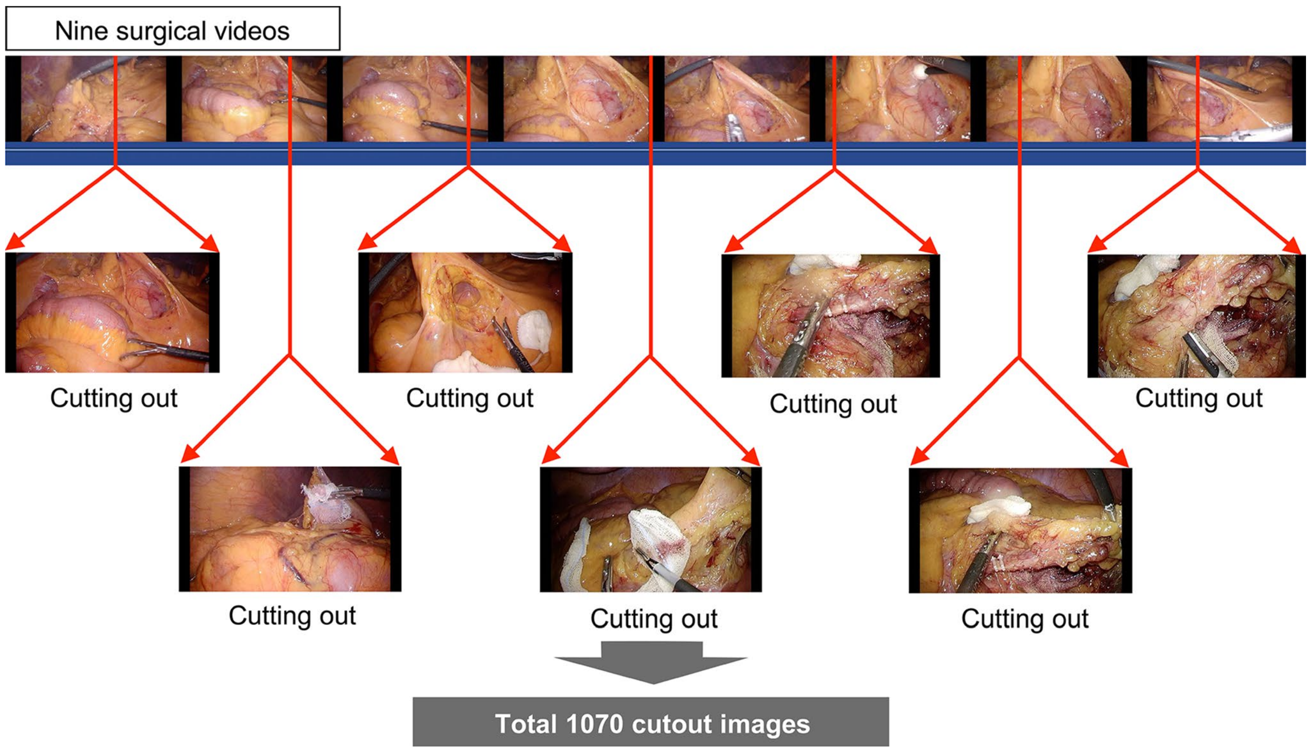


Fig. 1 Process of making still images for data labeling. A total of 1070 images were cut out from 9 surgical videos including 2 right colorectomies, 4 rectal surgeries, 2 hernia surgeries and 1 sigmoid resection surgery performed in the Department of Surgery at Tokyo Women's Medical University. Objects are labeled in these images

Table 1 Objects and numbers of individual objects annotated in the images

	Right colorectomy 1	Right colorectomy 2	Rectal surgery 1	Rectal surgery 2	Rectal surgery 3	Rectal surgery 4	Hernia 1	Hernia 2	Sigmoid resection	Total
GI tract	182	478	418	334	68	61	65	123	52	1781
Port	431	346	18	25	14	0	7	15	5	861
Forceps	426	376	353	366	93	64	61	88	46	1873
Gauze	107	95	204	413	4	46	6	94	47	1016
Vessel	49	125	34	95	0	6	0	0	43	352
Blood	25	15	8	0	7	89	12	0	52	208
Clip	122	457	126	17	0	4	0	0	34	760
Uterus	0	0	7	0	48	8	0	0	0	63
Total	1342	1892	1168	1250	234	278	151	320	279	

of the test data (at random) being used for training and 20% being used for measurement/validation. Figure 3 shows the flow of analysis using IBM Visual Insights.

Evaluation criteria

To quantitatively assess the performance of the designed network, accuracy, recall and precision were used as evaluation criteria in the image recognition field.

Accuracy is defined as the measurement of the percentage of correct image labels. It is calculated by (true positives + true negatives)/all cases. Recall is the percentage of images labeled as an object compared to all images that contain the object. It is calculated as true positives/(true positives + false negatives). Precision is the percentage of images that were correctly labeled as an object compared to all images labeled as that object. It is calculated by true positives/(true positives + false positives). Mean average precision (mAP) is the calculated mean of the precision for each

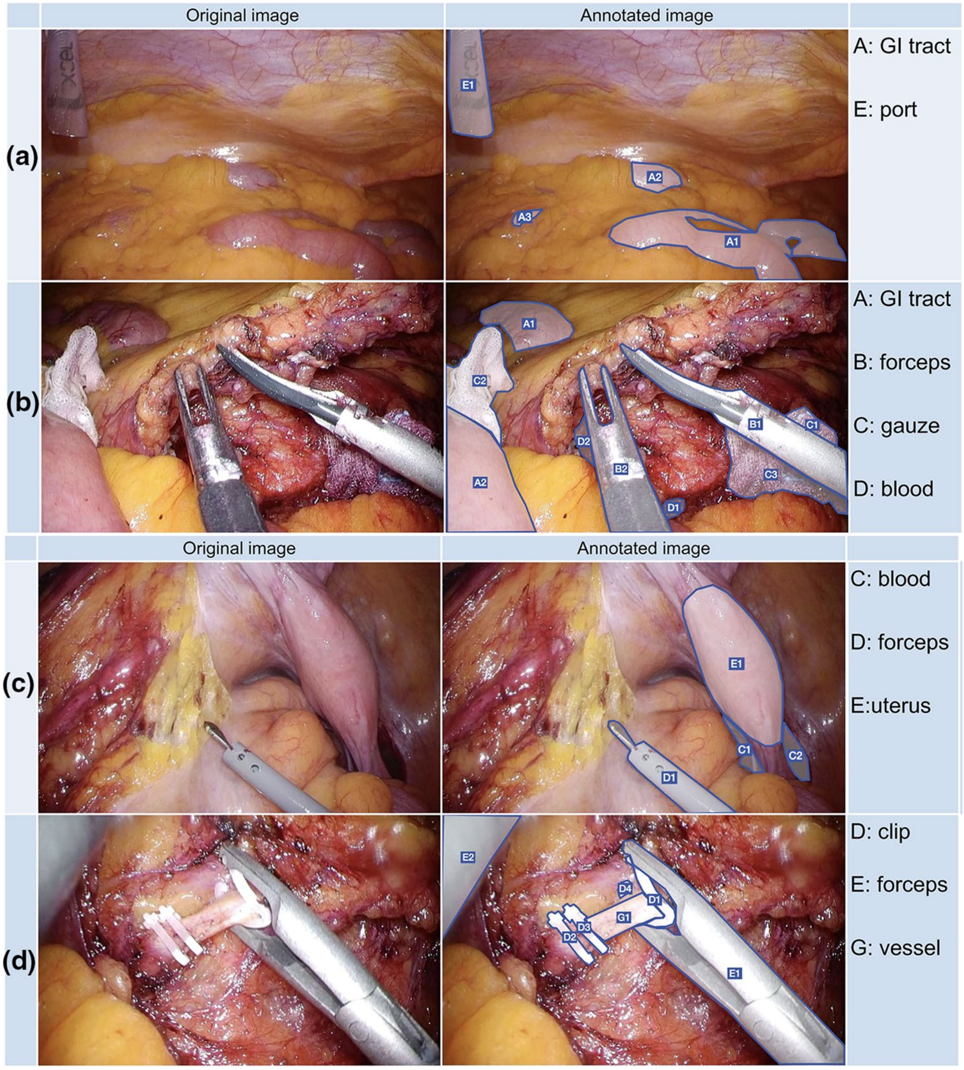


Fig. 2 Example images of labeling objects. A total of 8 objects, forceps, GI tract, port, gauze, clip, blood, vessel and uterus, were selected and labeled in the images to create an object recognition model. The left-side images are original, and the right-side images show labeled objects. Each object was surrounded carefully with a line for shape recognition. **a** GI tract and port are labeled. **b** GI tract, forceps, gauze and blood are labeled. **c** Blood, forceps and uterus are labeled. **d** Clip, forceps and vessel are labeled

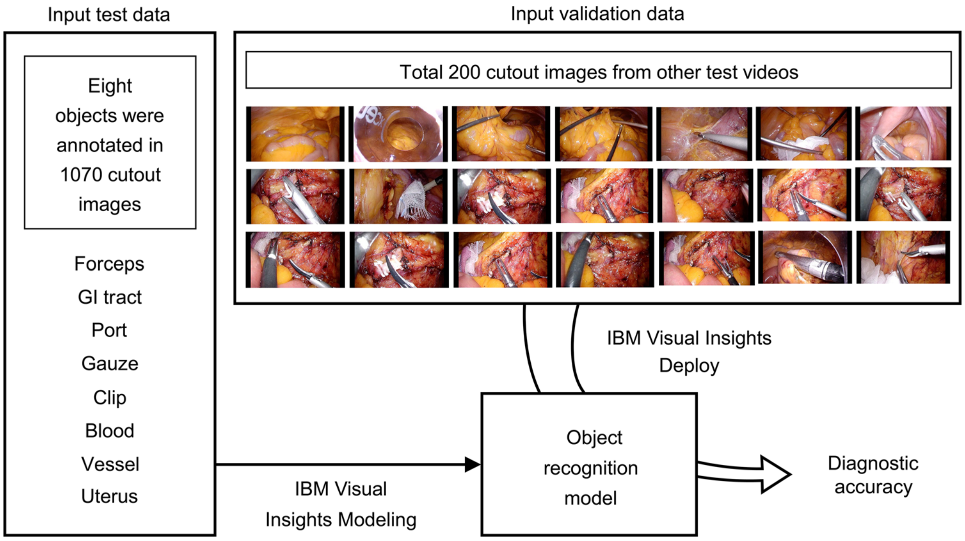
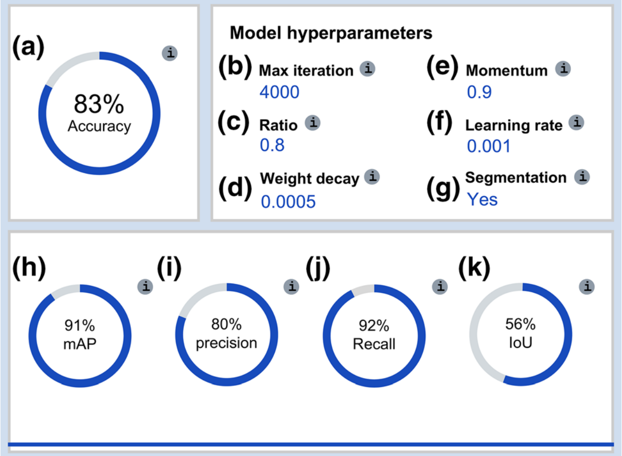

Input test data
Eight objects were annotated in 1070 cutout images
Forceps
GI tract
Port
Gauze
Clip
Blood
Vessel
Uterus
Input validation data
Total 200 cutout images from other test videos
IBM Visual Insights Deploy
Object recognition model
IBM Visual Insights Modeling
Diagnostic accuracy

Fig. 3 Flow of analysis using IBM Visual Insights. The 8 selected objects were labeled in a total of 1070 images that were cut out for creating an object recognition model. The other 200 images for validation were input into the model to verify whether each object was recognized accurately

object. Intersection over union (IoU), the location accuracy of the image label boxes, is calculated by the intersection (overlap) between a hand-drawn bounding box and a predicted bonding box divided by the union (combined area) of both bounding boxes.

Other hyperparameters in Fig. 4 set during the training process were Max iteration (the maximum number of times the data are passed through the training algorithm), weight decay (specifies regularization in the network, protects against over-fitting and is used to multiply the weights when training), momentum (increases the step size used when trying to find the minimum value of the error curve; a larger step size can keep the algorithm from stopping at a local minimum instead of finding the global minimum), learning rate (determines how much the weights in the network are adjusted with respect to the loss gradient; a correctly tuned value can result in a shorter training time) and segmentation (specifies whether segmentation was used to train the model).


(a) 83% Accuracy

Model hyperparameters

(b) Max iteration 4000

(c) Ratio 0.8

(d) Weight decay 0.0005

(e) Momentum 0.9

(f) Learning rate 0.001

(g) Segmentation Yes

(h) 91% mAP

(i) 80% precision

(j) 92% Recall

(k) 56% IoU

Fig. 4 Details of the training model. **a** Accuracy. **b** Max iteration. **c** Ratio. **d** Weight decay. **e** Momentum. **f** Learning rate. **g** Segmentation. **h** Mean average precision. **i** Precision. **j** Recall. **k** Intersection over union

Results

Figure 4 shows the details of the training model with 1070 images cut out. The accuracy of the model was 83%. Precision was 80%. Recall, the percentage of the images that were labeled as an object compared to all images that contain that object, was 92%. The mAP, the calculated mean of the precision for each object, was 91%. The IoU, the location accuracy of the image label boxes, was 56%.

The recall and precision for the detection of each object category in the model are shown in Table 2. 913 objects in eight categories were detected in 200 test images. Among the total number of detected objects, 834 objects were detected correctly. The number of objects not detected was 79. The number of false positives was 59. Figure 5 shows

Table 2 Recall and precision for each object

Object	Number of objects identified in images	Number of objects identified correctly	False negative	False positive	Recall % (95% CI)	Precision % (95% CI)
Forceps	347	334	13	7	96.3 (94.3–98.3)	97.9 (96.4–99.5)
GI tract	282	262	20	25	92.9 (89.9–95.9)	91.3 (88.0–94.6)
Port	31	27	4	4	87.1 (75.3–98.9)	87.1 (75.3–98.9)
Gauze	78	67	11	9	85.9 (78.2–93.6)	88.2 (80.9–95.4)
Clip	126	108	18	7	85.7 (79.6–91.8)	93.9 (89.5–98.3)
Blood	8	4	4	1	50.0 (15.4–84.6)	80.0 (44.9–115.1)
Vessel	29	23	6	5	79.3 (64.6–94.1)	82.1 (68.0–96.3)
Uterus	12	9	3	1	75.0 (50.5–99.5)	90.0 (71.3–108.6)
Total	913	834	79	59	91.3 (89.5–93.2)	93.4 (91.8–95.0)

False negative: Number of objects not identified

False positive: Number of objects identified despite absence of the objects

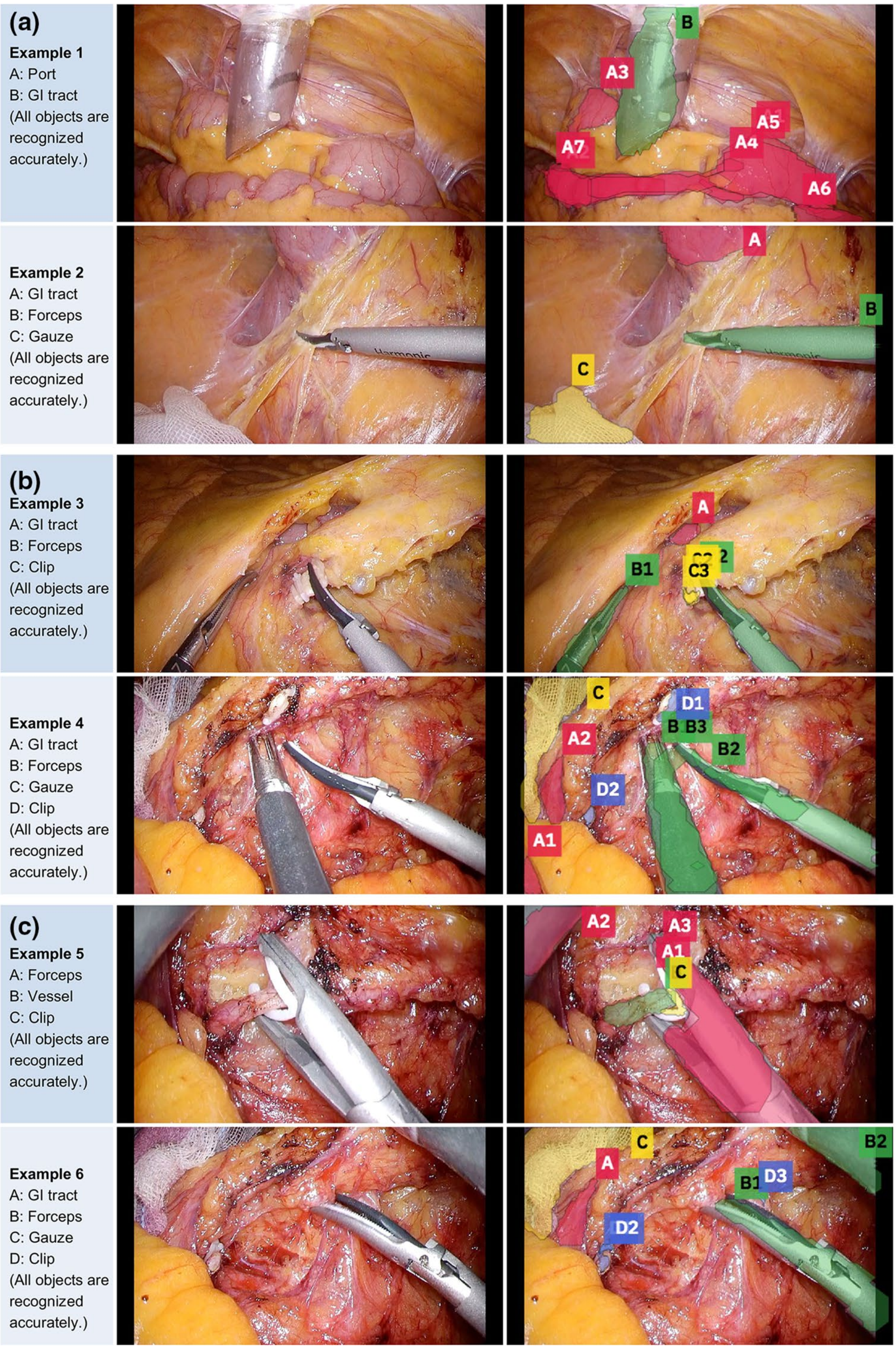


Fig. 5 Examples of object detection in surgical images. **(5a, Ex. 1)** The GI tract and port were recognized accurately. **(5a, Ex. 2)** The GI tract, forceps and gauze were recognized accurately. **(5b, Ex. 3)** The GI tract, forceps and clips were recognized accurately. **(5b, Ex. 4)** The GI tract, forceps, gauze and clips were recognized accurately. **(5c, Ex. 5)** The forceps, vessel and clips were recognized accurately. **(5c, Ex. 6)** The GI tract, forceps, gauze and clips were recognized accurately

examples of correct detection of objects in various categories. Figure 6 shows examples of false negative detection error, when the object was present in the image but was not detected. Figure 7 shows examples of detection error when one object was identified as another object.

A surgical video with a 40 s run time was used to test the model, with the results indicating that the object was detected accurately.

Discussion

This study demonstrated object recognition in surgical images using deep learning. In most cases, all objects were identified correctly (Fig. 5 a, b, c). The recall and precision of each object showed high accuracy (Table 2). The general framework for video image analysis involves structural-units segmentation, feature extraction for presenting

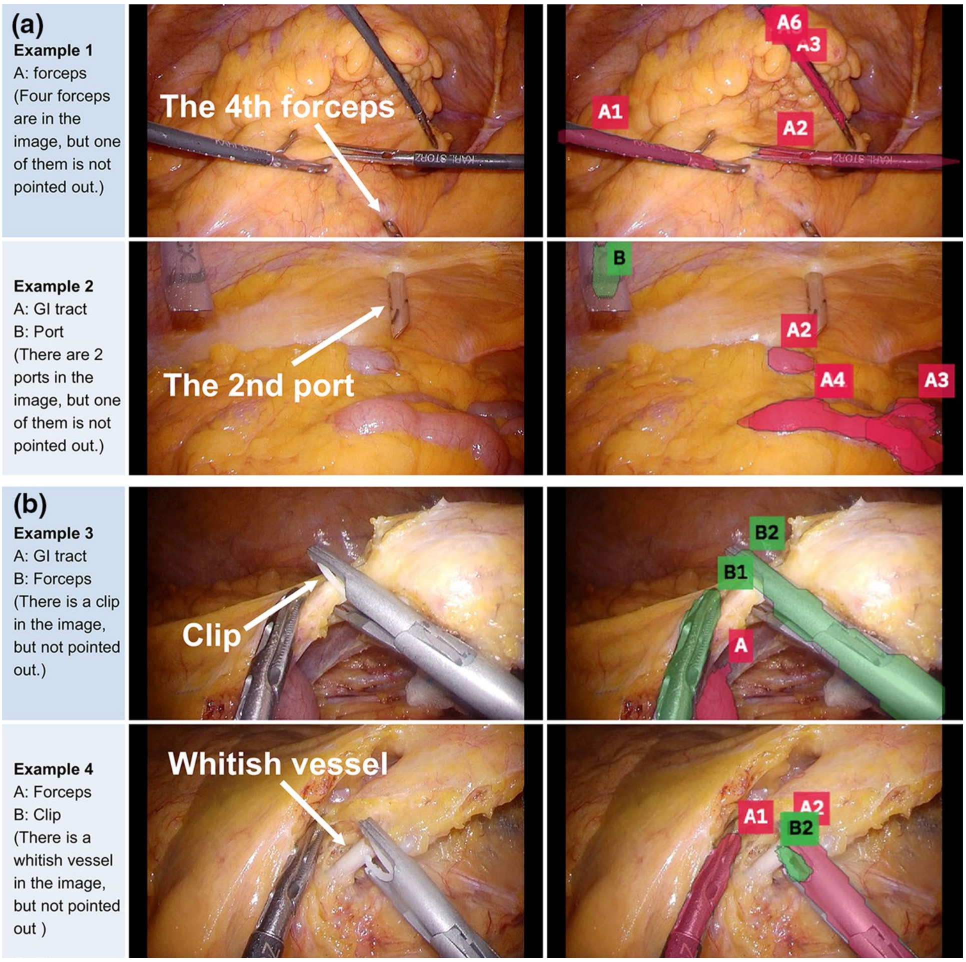

(a)
Example 1
A: forceps
(Four forceps are in the image, but one of them is not pointed out.)
The 4th forceps
A6
A3
A1
A2
Example 2
A: GI tract
B: Port
(There are 2 ports in the image, but one of them is not pointed out.)
The 2nd port
B
A2
A4
A3
(b)
Example 3
A: GI tract
B: Forceps
(There is a clip in the image, but not pointed out.)
Clip
B2
B1
A
Example 4
A: Forceps
B: Clip
(There is a whitish vessel in the image, but not pointed out)
Whitish vessel
A1
A2
B2

Fig. 6 Example of false negative detection error when the object present in the image was not detected. **(6a, Ex. 1)** An example with a mistake. There are 4 forceps in the image, but the 4th one was not identified. **(6a, Ex. 2)** There are 2 ports in the image, but one of them was not identified (**6b, Ex. 3**) An example with a mistake. There is a clip in the image, but it is recognized as a part of a forceps. (6**b, Ex. 4**) There is a whitish vessel in the image, but it was not identified

a specific object or activity using the extracted features for data mining, annotation/classification for developing a semantic video index and searching the video database with a distance similarity measure [4]. In our study, a CNN-based algorithm was designed and verified for its applicability in identification of both anatomical features and surgical tools. In comparison with other techniques, CNN can deal with a larger number of features during training. When compared to recent studies that also used CNN-based surgical image or action detection/classification, our method showed comparable or superior results. For example, in a study that analyzed laparoscopic intraoperative videos, the automatic surgical phase and action classification task showed overall accuracies of 81.0% and 83.2%, respectively, and the mean IoU for the automatic tool segmentation task for surgical tools was 51.2% [15]. A greater or similar value of these parameters was achieved. This is conceivable, because IBM Visual Insights is made from complex models to achieve better results. This is a preliminary report, and future research is needed. Our results support the view that surgeons can rely on AI-based analysis of population and patient-specific data to improve each phase of intervention and care and to provide a rapid analysis of large numbers of preoperative images and intraoperative scenes to improve the decision-making process dramatically [6].

Analysis of anatomic structures during surgery or diagnosis is relevant for documenting the details of a disease and its treatment and also for medical research and teaching purposes. Based on the videos of colorectal and hernia surgeries, the GI tract, blood, vessels and uterus were studied. For the GI tract, recall and precision were 92.9% and 91.3%, respectively. Twenty GI tracts were not identified due to unsharp images and somewhat darker colors than in other images. In most false positive cases of the GI tract, the peritonea are recognized as the GI tract when their color and gloss are similar (Fig. 7 a, Ex.1). In most of the recent research on the GI tract, the objective was limited to identification of a specific disease, such as early gastric cancer or the existence of polyps [16–20] instead of anatomical detection of the GI tract. However, the success of the CNN structure is highly correlated with the number of samples used for training [21]. For blood, recall and precision were 50% and 80%, respectively. Blood is not pointed out, perhaps, when there are fewer heliotropes than in the

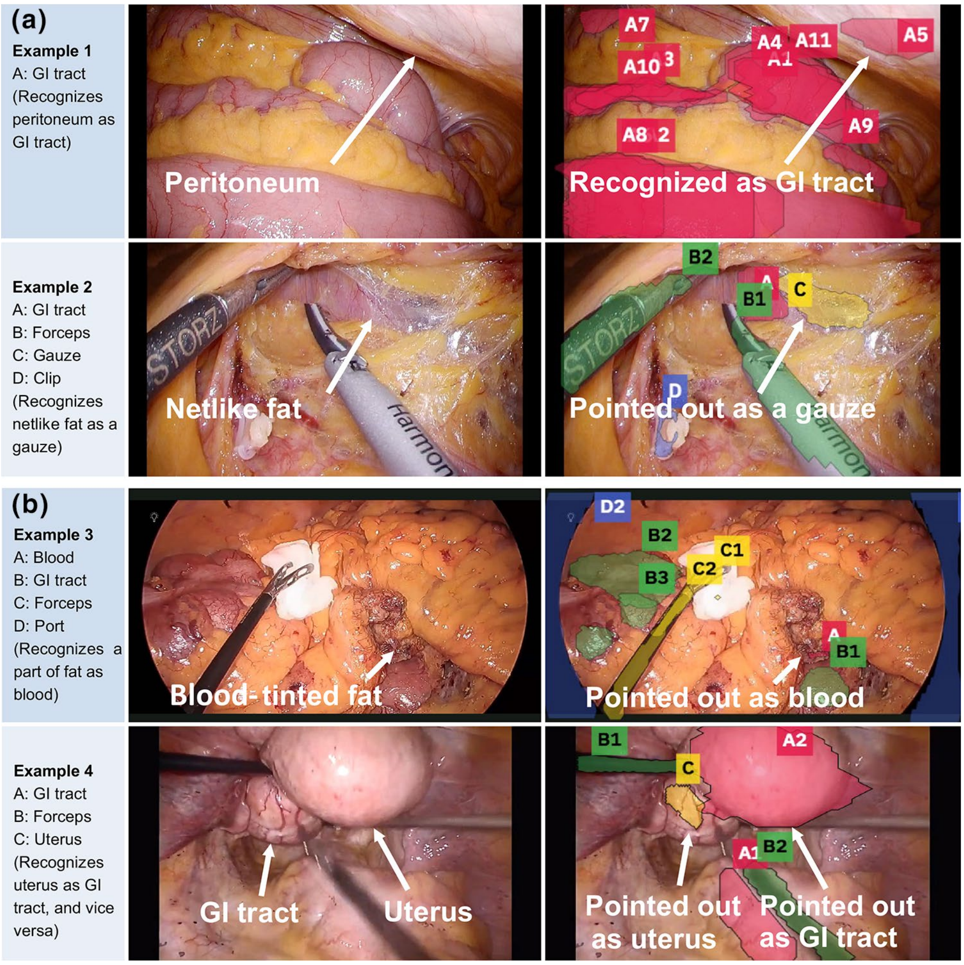


Fig. 7 Example of false positive detection error when one object was detected as another object. **(7a, Ex. 1)** All GI tracts were recognized accurately, but an intestinal wall was also identified as a GI tract. **(7a, Ex. 2)** There is no gauze in the image, but a part of the netlike fat is recognized as gauze. **(7b, Ex. 3)** Part of the fat is recognized as blood **(7b, Ex. 4)** The uterus is recognized as the GI tract and vice versa

model. In most false positive cases of blood, blood-tinted fat is recognized incorrectly as blood (Fig. 7 b, Ex. 3). In contrast, a few state-of-the-art deep learning-based systems have recently been reported to be capable of automatic detection of gastrointestinal bleeding with more than 98% recall by individual still image analysis [19]. For vessels, recall and precision were 79.3% and 82.1%, respectively. Vessels are not identified if their color is different from those in the training model (Fig. 6 b, Ex. 4). In false positive cases of vessels, reddish fat is also recognized incorrectly as a vessel. For the uterus, recall and precision were 75.0% and 90.0%, respectively. The uterus was not identified or mistakenly recognized if the GI tract or other organ had similar color and gloss (Fig. 7 b, Ex. 4). A previous study that explored the performance of two well-known CNN architectures, AlexNet and GoogLeNet, for detecting anatomical structures during gynecologic surgery including the uterus, the mean recall value was 78.2% and 61.5%, respectively and for the uterus, the recall value was 80.1% [22]. Our study showed a higher average recall value than that study. In a study of gynecological shot classifications using CNN-based architecture, the average precision and recall values were 42% and 43%, respectively. The accuracy achieved in that study was 48.67%, which was much lower than the accuracy of our method [23]. In a surgical action video scene, the interaction of various surgical instruments with tissues and organs represented the technicalities of the process and the analysis of these scenes is important for documentation and quality control, as well as training. Our study included gauze, clips, forceps and ports, some of the most frequently used tools, for detection. The common challenges encountered with image-based methods for the identification and tracking of surgical instruments are high deformation or artifacts, blurred surgical scenes due to camera movement and gas generated by the equipment and occlusion due to blood stains on the camera lens [24]. The initial methods depended on low-level handcrafted features, such as the amalgamation of features related to shape, color and texture [25]. Recent studies have focused on exploring the usage of CNNs in learning more discriminative visual features. When the performance on intraoperative tool detection in terms of the mAP in previous studies was compared with that in the current study, our model showed better performance. mAPs achieved in the earlier studies were 63.8% [26], 54.5%

[27], 52.5% [28]), 81% [29], 81.8% [30], 72.26% [31], 84.7% [32], which are lower than the mAP of 91% achieved in our study. In our study, for forceps, recall and precision were 96.3% and 98%, respectively. Thirteen forceps were not identified due to slightly blurred images or only a small part of the forceps being visible in the image (Fig. 6 a, Ex.1). Examples of forceps false positives were ports, long and narrow-shaped fat, or clips, which tended to be recognized as forceps. For port, both recall and precision were 87.1%. Four ports were not identified because they were transparent and were assimilated into other objects (Fig. 6 a, Ex. 2). In false positive cases of ports, reflection of light affected false recognition as other objects. For gauze, recall and precision were 85.9% and 88.2%, respectively. Eleven gauzes were not identified when the mesh of the gauze was not clear due to unsharp images. There were some false identifications of gauze when the shapes of other objects like fat were similar to gauze (whitish and netlike) (Fig. 7 a, Ex. 2). For clips, recall and precision were 85.7% and 93.9%, respectively. Eighteen clips were not identified; in most cases, they were recognized as a part of a forceps (Fig. 6 b, Ex. 3). In false positive cases of clips, other objects like fat and vessels were recognized incorrectly as clips because of their colors and shapes. We observed that the quality of the prediction varied based on the sharpness (clearness) of images, which considerably affected the outcome of validation. The more samples we enter into the model, the better the results that can be achieved.

This study aimed to build a navigation or object detection system during surgery. Given the promising results of our study, we believe that the model could ultimately be used to automatically evaluate surgical skills using CV analysis. The results of our study can contribute to the field of automatization of surgical assistance that can manage, deliver and retrieve surgical instruments for surgeons upon request [7]. Moreover, as the current coronavirus 2019 (COVID-19) crisis has accelerated and enhanced the requirement of e-learning solutions, our study contributes to the global effort of developing new training methods to optimize complex surgical education [33].

This study has several limitations. First, it was retrospective in nature. Second, it was performed with a limited number of surgical videos of colorectal and hernia surgeries. Despite these limitations, our results add substantial value to the field of intraoperative detection of anatomical features and surgical tools.

Conclusion

We propose a real-time detection model for identifying surgical instruments and anatomical features during various gastrointestinal surgeries with a CNN system. The proposed model could detect objects with high accuracy and performed comparably to other studies. Real-time object recognition will play a major role in surgical education and navigation surgery, and the technology has the potential to expand significantly by storing large amounts of data, although we encountered the problem of erroneous object detection due to the limited number of images used. Further studies are warranted to improve the data preprocessing and augment the tracking algorithm.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11548-021-02434-w.

Acknowledgements Editorial support, in the form of medical writing, assembling tables and creating high-resolution images based on authors' detailed directions, collating author comments, copyediting, fact checking and referencing, was provided by Editage, Cactus Communications. I would like to thank Ms. Junko Machida for study support.

Authors' contributions All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Yoshiko Bamba. The first draft of the manuscript was written by Yoshiko Bamba, and all authors commented on subsequent versions of the manuscript. All authors read and approved the final manuscript.

Funding This work was supported in part by a research grant from TWMU Career Development Center for Medical Professionals and a NAKAYAMA KOMEI Research Fellowship Grant.

Availability of data and material The data that support the findings of this study are available from the corresponding author upon reasonable request.

Declarations

Conflicts of interest None of the authors has any conflicts of interest.

Ethics approval The protocol for this study was reviewed and approved by the Tokyo Women's Medical University Review Board (Protocol No: 5380).

Consent to participate NA (Images from surgical videos were used, and there was no concern about identifiable information).

References

1. Grenda TR, Pradarelli JC, Dimick JB (2016) Using Surgical Video to Improve Technique and Skill. Ann Surg 264:32–33. https://doi.org/10.1097/SLA.0000000000001592
2. Karic B, Moino V, Nolin A, Andrews A, Brisson P (2020) Evaluation of surgical educational videos available for third year medical students. Med Educ Online 25:1714197. https://doi.org/10.1080/10872981.2020.1714197
3. Levin M, McKechnie T, Khalid S, Grantcharov TP, Goldenberg M (2019) Automated methods of technical skill assessment in surgery: a systematic review. J Surg Educ 76:1629–1639. https://doi.org/10.1016/j.jsurg.2019.06.011

4. Loukas C (2018) Video content analysis of surgical procedures. Surg Endosc 32:553–568. https://doi.org/10.1007/s00464-017-5878-1
5. Hashimoto DA, Rosman G, Rus D, Meireles OR (2018) Artificial intelligence in surgery: promises and perils. Ann Surg 268:70–76. https://doi.org/10.1097/SLA.0000000000002693
6. Hashimoto DA, Rosman G, Witkowski ER, Stafford C, Navarette-Welton AJ, Rattner DW, Lillemoe KD, Rus DL, Meireles OR (2019) Computer vision analysis of intraoperative video: automated recognition of operative steps in laparoscopic sleeve gastrectomy. Ann Surg 270:414–421. https://doi.org/10.1097/SLA.0000000000003460
7. Zhou T, Wachs JP (2017) Needle in a haystack: Interactive surgical instrument recognition through perception and manipulation. Rob Auton Syst 97:182–192. https://doi.org/10.1016/j.robot.2017.08.013
8. Komura D, Ishikawa S (2018) Machine learning methods for histopathological image analysis. Comput Struct Biotechnol J 9:34–42. https://doi.org/10.1016/j.csbj.2018.01.001
9. Bonrath EM, Gordon LE, Grantcharov TP (2015) Characterising "near miss" events in complex laparoscopic surgery through video analysis. BMJ Qual Saf 24:516–521. https://doi.org/10.1136/bmjqs-2014-003816
10. Volkov M, Hashimoto DA, Rosman G, Meireles OR, RusD (2017) Machine learning and coresets for automated real-time video segmentation of laparoscopic and robot-assisted surgery. In:2017 IEEE International Conference on Robotics and Automation. pag. 754–759. Available at https://doi.org/10.1109/ICRA.2017.7989093
11. Abdelhafiz D, Yang C, Ammar R, Nabavi S (2019) Deep convolutional neural networks for mammography: advances, challenges and applications. BMC Bioinform 20(Suppl 11):281
12. Li L, Chen Y, Shen Z, Zhang X, Sang J, Ding Y, Yang X, Li J, Chen M, Jin C, Chen C, Yu C (2020) Convolutional neural network for the diagnosis of early gastric cancer based on magnifying narrow band imaging. Gastric Cancer 23:126–132. https://doi.org/10.1007/s10120-019-00992-2
13. Wahab N, Khan A, Lee YS (2017) Two-phase deep convolutional neural network for reducing class skewness in histopathological images based breast cancer detection. Comput Biol Med. 85:86–97
14. Orturk S, Akdemir B (2019) A convolutional neural network model for semantic segmentation of mitotic events in microscopy images. Neural Comput Appl 31(8):3719–3728
15. Kitaguchi D, Takeshita N, Matsuzaki H, Oda T, Watanabe M, Mori K, Kobayashi E, Ito M (2020) Automated laparoscopic colorectal surgery workflow recognition using artificial intelligence: Experimental research. Int J Surg 79:88–94. https://doi.org/10.1016/j.ijsu.2020.05.015
16. Öztürk Ş, Özkaya U (2020) Gastrointestinal tract classification using improved LSTM based CNN. Multimed Tools Appl 79:28825–28840. https://doi.org/10.1007/s11042-020-09468-3
17. Horiuchi Y, Aoyama K, Tokai Y, Hirasawa T, Yoshimizu S, Ishiyama A, Yoshio T, Tsuchida T, Fujisaki J, Tada T (2020) Convolutional neural network for differentiating gastric cancer from gastritis using magnified endoscopy with narrow band imaging. Dig Dis Sci 65:1355–1363. https://doi.org/10.1007/s10620-019-05862-6
18. Hirasawa T, Aoyama K, Tanimoto T, Ishihara S, Shichijo S, Ozawa T, Ohnishi T, Fujishiro M, Matsuo K, Fujisaki J, Tada T (2018) Application of artificial intelligence using a convolutional neural network for detecting gastric cancer in endoscopic images. Gastric Cancer 21:653–660. https://doi.org/10.1007/s10120-018-0793-2
19. Karnes WE, Alkayali T, Mittal M,Patel A, Kim J, Chang JK, Ninh AQ, Urban G, Baldi P (2017) Su1642 Automated polyp detection using deep learning: leveling the field. GastrointestEndosc 85(5): AB376–AB377. DOI: https://doi.org/10.1016/j.gie.2017.03.871
20. M Kirkerød RJ Borgli V Thambawita S Hicks RieglerMA. Halvorsen P (2019) Unsupervised preprocessing to improve generalisation for medical image classification. In, 2019 13th International Symposium on Medical Information and Communication Technology (ISMICT) [Internet], pag 1–6 Available at: https://doi.org/10.1109/ismict.2019.8743979
21. Aoki T, Yamada A, Kato Y, Saito H, Tsuboi A, Nakada A, Niikura R, Fujishiro M, Oka S, Ishihara S, Matsuda T, Nakahori M, Tanaka S, Koike K, Tada T (2020) Automatic detection of blood content in capsule endoscopy images based on a deep convolutional neural network. J Gastroenterol Hepatol 35:1196–1200. https://doi.org/10.1111/jgh.14941
22. Petscharnig S, Schöffmann K (2018) Learning laparoscopic video shot classification for gynecological surgery. Multimed Tools Appl 77:8061–8079.https://doi.org/10.1007/s11042-017-4699-5
23. Petscharnig S, Schöffmann K (2017) Deep Learning for Shot Classification in Gynecologic Surgery Videos. In: Amsaleg L, Guðmundsson G, Gurrin C, Jónsson B, Satoh S. (eds) MultiMediaModeling. MMM 2017. Lecture Notes in Computer Science, vol 10132. Springer, Cham. https://doi.org/10.1007/978-3-319-51811-4_57
24. Jin Y, Li H, Dou Q, Chen H, Qin J, Fu C, Heng P (2020) Multi-task recurrent convolutional network with correlation loss for surgical video analysis. Med Image Anal 59:101572
25. Lalys F, Riffaud L, Bouget D, Jannin P (2012) A framework for the 980 recognition of high-level surgical tasks from video images for cataract surgeries. IEEE Trans Biomed Eng 59:966–976. https://doi.org/10.1109/TBME.2011.2181168
26. Raju A, Wang S, Huang J (2016) M2cai surgical tool detection challenge report. http://camma.u-strasbg.fr/m2cai2016/reports/Raju-Tool.pdf
27. Sahu M, Mukhopadhyay A, Szengel A, Zachow S (2016b) Tool and phase recognition using contextual CNN features. arXiv preprint
28. Twinanda AP, Mutter D, Marescaux J, de Mathelin M, Padoy N (2016) Single- and multi-task architectures for tool presence detection challenge at M2CAI 2016. Preprint at
29. Twinanda AP, Shehata S, Mutter D, Marescaux J, De Mathelin M, Padoy N (2017) Endonet: a deep architecture for recognition tasks on laparoscopic videos. IEEE Trans Med Imaging 36:86–97. https://doi.org/10.1109/tmi.2016.2593957
30. Jin A, Yeung S, Jopling J, Krause J, Azagury D, Milstein A, Fei-Fei L (2018) Tool detection and operative skill assessment in surgical videos using region-based convolutional neural networks. In: WACV, pp 691–699. Preprint at https://arxiv.org/abs/1802.08774
31. Choi B, Jo K, Choi S, Choi J (2017) 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) [Internet]. Seogwipo: IEEE; 2017 [citato 21 gennaio 2020]. pag. 1756–1759. Available at: https://ieeexplore.ieee.org/document/8037183/
32. Jo K, Choi Y, Choi J, Chung JW(2019) Robust real-time detection of laparoscopic instruments in robot surgery using convolutional neural networks with motion vector prediction. Appl Sci 9:2865.https://doi.org/10.3390/app9142865
33. García Vazquez A, Verde JM, Dal Mas F, Palermo M, Cobianchi L, Marescaux J, Gallix B, Dallemagne B, Perretta S, Gimenez ME (2020) Image-Guided Surgical e-Learning in the Post-COVID-19 Pandemic Era: What Is Next? J Laparoendosc Adv Surg Tech A 30:993–997. https://doi.org/10.1089/lap.2020.0535