Surgical Endoscopy
https://doi.org/10.1007/s00464-025-12203-4

Artificial intelligence-assisted scar visualization under intraoperative bleeding using CycleGAN and uncertainty fusion in laparoscopic cholecystectomy

Tatsushi Tokuyasu[1] · Subal Ikeda[1] · Hiroki Orimoto[2] · Teijiro Hirashita[2] · Yuichi Endo[2] · Masafumi Inomata[2]

Received: 20 June 2025 / Accepted: 30 August 2025

Abstract

Background Accurate intraoperative identification of scar tissue is essential for preventing bile duct injury during laparoscopic cholecystectomy (LC), especially under visually impaired conditions caused by bleeding. This study aimed to develop an artificial intelligence (AI)-based framework to enhance scar region prediction in such challenging surgical environments.

Methods A hybrid approach was proposed, combining Cycle-Consistent Generative Adversarial Network-based image translation with uncertainty-aware fusion. Bleeding-contaminated laparoscopic images were translated into pseudo non-bleeding representations using unpaired domain adaptation. Segmentation results obtained from the original and translated images were then fused based on pixel-wise entropy to improve robustness.

Results The system was evaluated using 99 representative images from 20 surgical patients. Compared with conventional segmentation methods, the proposed framework significantly improved Dice coefficients across all three board-certified endoscopic surgeons who served as expert annotators, with all improvements demonstrating significance ($P < 0.001$). Subjective evaluations by the same surgeons confirmed high clinical utility, particularly in scar visibility and boundary delineation. The framework achieved near real-time inference speed (0.06 s per frame on an RTX A5000 GPU).

Conclusion This AI-assisted framework improved the accuracy and robustness of scar tissue detection during LC, even in bleeding-compromised fields. Its real-time capability and strong clinical validation indicate substantial potential for intraoperative application and enhancement of surgical safety.

Keywords Artificial intelligence · Intraoperative bleeding · Intraoperative image enhancement · Laparoscopic cholecystectomy · Scar tissue identification · Real-time systems

Laparoscopic cholecystectomy (LC) is the most commonly performed procedure for the treatment of acute cholecystitis (AC) and cholecystolithiasis [1]. However, in patients with severe inflammation, fibrosis, and scarring (particularly within Calot's triangle), the procedure becomes technically challenging and is associated with a significantly increased risk of bile duct injury (BDI) [2, 3]. In such cases, precise intraoperative identification of scarring tissue is essential to guide dissection planes and prevent complications, including conversion to open surgery or the need for bailout procedures [4].

Achieving the critical view of safety (CVS) is a widely accepted standard for safe LC and is essential for preventing BDI [5]. The CVS requires clear identification of the cystic duct and artery following complete dissection of Calot's triangle, as well as the separation of the gallbladder from the liver bed [6]. However, in patients with severe inflammation or fibrosis, anatomical landmarks are often obscured, making the establishment of CVS challenging. A recent study reported that up to 15% of laparoscopic cholecystectomies fail to achieve CVS due to dense adhesions or scarring, which significantly increases the risk of BDI and necessitates alternative strategies such as the "fundus-first" approach or subtotal cholecystectomy [4, 7]. In such settings, reliable identification of fibrotic and scarred regions

✉ Tatsushi Tokuyasu
tokuyasu@fit.ac.jp

[1] Department of Information Systems and Engineering, Faculty of Information Engineering, Fukuoka Institute of Technology, 1-30-1 Wajiro higashi, Higashi-ku, Fukuoka, Fukuoka 811-0295, Japan

[2] Department of Gastroenterological and Pediatric Surgery, Faculty of Medicine, Oita University, Oita, Japan

Published online: 15 September 2025

is crucial to guide safe dissection and prevent intraoperative complications.

Recent advances in artificial intelligence (AI) have enabled real-time intraoperative image analysis, including anatomical landmark recognition and surgical phase detection [8–11]. In LC, AI-based segmentation models have been applied to enhance the visualization of critical structures and guide safe dissection [12–14]. These approaches primarily focus on identifying and evaluating structures suitable for dissection. However, in patients with high-difficulty presentations, inflammation, scarring, and fibrosis resulting from chronic or past inflammatory episodes can obscure anatomical structures, thereby complicating surgical navigation. In such cases, information indicating "where to avoid dissecting" is often more directly relevant to surgical safety than information indicating "where to dissect." Building on this concept, our previous study by Orimoto et al. demonstrated the potential of a fibrosis prediction model that visualizes scarred regions using expert-annotated laparoscopic images [15]. Although promising, the model's performance was substantially reduced in bleeding-contaminated scenes—a common intraoperative challenge—where occlusion and color distortion hinder human and AI recognition. In clinical practice, intraoperative bleeding poses a serious risk by obscuring anatomical details at precisely the moment when accurate recognition is most critical [16]. The present study addresses this limitation by implementing an AI-assisted image enhancement strategy specifically tailored for challenging intraoperative environments. The AI targets were limited to scarred regions formed by chronic inflammation or previous inflammatory episodes. In this context, scar refers to tissue alterations caused by advanced inflammation that render dissection difficult [17], often suggesting challenges in achieving the CVS and in maintaining dissection along a safe anatomical plane [18]. Although fibrosis can be considered a precursor to scarring, scarred regions in this study are defined as those with distinct morphological features—such as whitish discoloration, surface irregularity, or tissue retraction—that are visually discernible intraoperatively and directly relevant to surgical decision-making. These intraoperative definitions are conceptually distinct from histopathological confirmation, which refers to microscopic structural changes observed in resected specimens.

Instead of attempting direct detection of scarred regions under compromised visibility, a generative approach was employed to improve the visual clarity of surgical scenes affected by bleeding [19]. Specifically, cycle-consistent generative adversarial network (CycleGAN)-based image translation was used to convert laparoscopic images impacted by intraoperative bleeding into pseudo non-bleeding representations [20, 21]. These transformed images were subsequently processed using a pretrained scar prediction model. To integrate predictions from both the original and pseudo non-bleeding images, an uncertainty-aware fusion strategy was introduced, in which outputs are adaptively weighted based on the confidence (entropy) of each prediction. This mechanism enhances the visibility and accuracy of scar detection, even under visually compromised conditions caused by bleeding [22, 23]. The final results are overlaid onto the original image, providing surgeons with real-time guidance that is both familiar and clinically actionable.

The system was evaluated using quantitative segmentation metrics (e.g., the Dice coefficient) and qualitative assessment by board-certified surgeons. The results demonstrate that this approach improves the interpretability of scarring regions under challenging surgical conditions and supports safe dissection in patients with high-risk LC. This technique may serve as a practical intraoperative support tool, particularly for general surgeons managing AC with poor visibility.

Materials and methods

A schematic workflow diagram (Fig. 1) is presented to illustrate the overall process, from algorithm development to final analysis. The numbered phases in the diagram outline the sequential flow of the study and are referenced where relevant in the following subsections.

Study design, ethics, and overall workflow

A retrospective analysis was conducted using a previously published dataset compiled by Orimoto et al. [15], which included LC procedures performed at Oita University Hospital. The original study aimed to train and validate a scar prediction model using expert-annotated surgical videos. The present study builds upon this foundation by implementing advanced image enhancement and fusion strategies to address performance limitations observed under bleeding conditions.

All data were fully anonymized prior to analysis. The study was approved by the Institutional Review Boards of Oita University Faculty of Medicine (IRB no. 2032) and Fukuoka Institute of Technology (IRB no. hm02-22). The study was conducted in accordance with the principles of the Declaration of Helsinki and its subsequent amendments.

Clinical dataset preparation

The training dataset consisted of patients with LC treated at Oita University Hospital, in whom the presence or absence of scarring could be corroborated by pathological data. These patients were organized sequentially by case ID (SC000–SC020). In this study, the term "scar region"

Phase	Research flow	Description
1. Preparation	Clinical video acquisition	Training dataset: Laparoscopic cholecystectomy (LC) videos with pathological specimens from Oita University Hospital Test dataset: LC videos performed at other institutions
	Annotation	Manual labeling of scar regions in endoscopic images
2. Model construction	Model training	Style transfer to non-bleeding surgical fields using CycleGAN Scar region prediction using semantic segmentation
3. AI processing	Scar region prediction	Prediction of scar regions from original endoscopic images
	CycleGAN + Scar region prediction	Style transfer of endoscopic images to non-bleeding appearance, followed by scar region prediction.
	Uncertainty-based image fusion	Integration of prediction results based on uncertainty estimation
4. Evaluation	Objective evaluation	Dice coefficient calculation comparing predictions with annotations from three experienced surgeons
	Subjective evaluation	Visual assessment by experienced surgeons comparing scar predictions on original and style-transferred images
	Statistic analysis	Statistical comparison of Dice coefficients between the proposed method and conventional approaches

Fig. 1 Overview of the workflow summarizing the process from algorithm development to final analysis, comprising four sequential phases: (1) dataset preparation, (2) AI model construction and training, (3) inference workflow, and (4) evaluation

refers to intraoperative areas of scarring as defined in the Introduction.

From these selected patients, expert-annotated surgical videos focusing on the region surrounding Calot's triangle were used to generate a training dataset for model development. A total of 2119 frames were extracted, with approximately 80–100 frames obtained from each patient. To reduce redundancy caused by the high similarity of consecutive frames in laparoscopic videos, frames were initially sampled at intervals of approximately 30 frames. Among these candidates, only frames with no or minimal motion blur were retained. For additional clinical evaluation, a separate set of 100 representative images was independently selected from 20 patients with LC (SC021–SC040) obtained from an external institution. Pathological confirmation was not available for these patients nor were detailed data regarding surgical settings, patient demographics, or imaging devices. Instead, patients were selected in whom scarred areas could be clearly identified by consensus between two board-certified surgeons (certified by the Japan Society for Endoscopic Surgery). These test data were specifically used for external evaluation, enabling assessment of the clinical utility of the proposed framework through both objective evaluation using the Dice coefficient and subjective evaluation based on video review.

As a statistical measure for both the training and test datasets, the scar area ratio was calculated for each patient. The corresponding distributions are presented as box plots in Figs. 2 (training dataset) and 3 (test dataset). In each box plot, the central line represents the median, whereas the green triangle denotes the mean. The whiskers extend to 1.5 times the interquartile range, and circles indicate outliers. The results revealed considerable variability: in the training dataset (21 patients), the mean scar area ratio was 10.3% with a standard deviation of 7.4% (range, 0–52%); in the test dataset (20 patients), the mean was 22.3% with a standard deviation of 10.9% (range, 3.5–56.5%). These findings indicate substantial variability in the presence and extent of scarring across patients and frames, suggesting that the model is unlikely to rely on a single pattern during training.

AI model development and training

The proposed workflow integrates three key AI components. First, a CycleGAN model translates bleeding-contaminated laparoscopic images into pseudo non-bleeding representations. Second, scar regions are predicted using the pretrained HyperSeg model [24]. Finally, predictions from both the original and translated images are merged through an uncertainty-aware fusion process, which employs entropy-based confidence weighting.

CycleGANs enable unpaired image-to-image translation, a feature particularly advantageous in surgical contexts where identical pre- and post-bleeding views are unavailable.

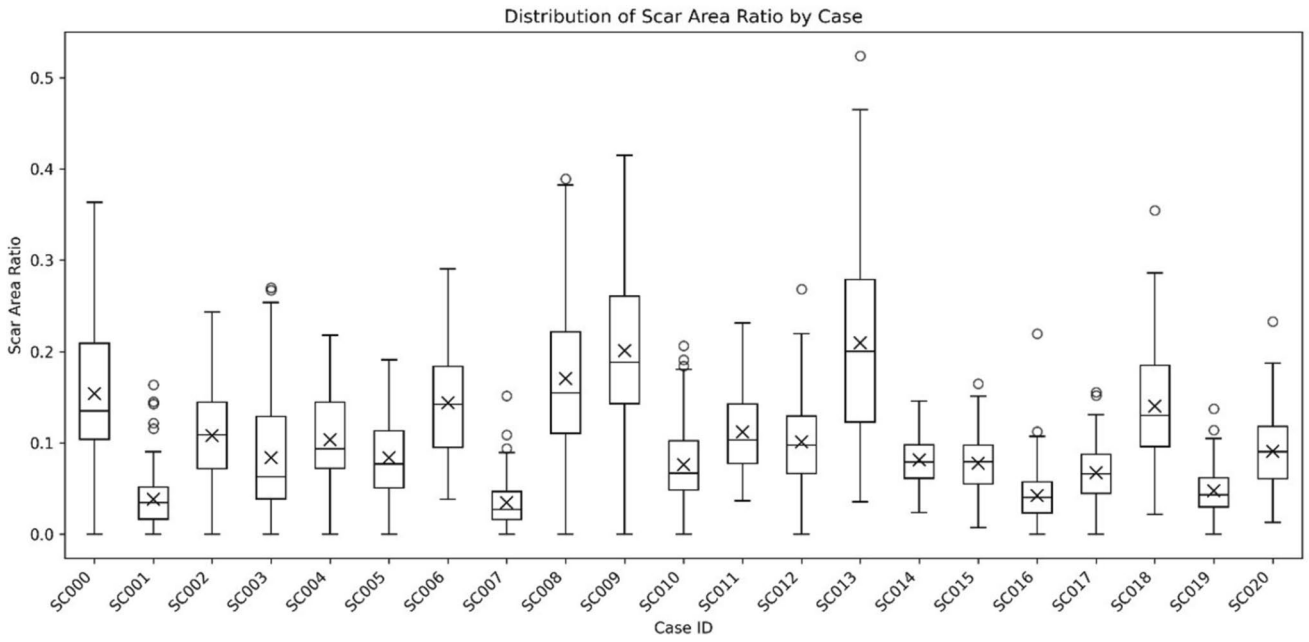

Fig. 2 Distribution of scar area ratio by case in the training dataset

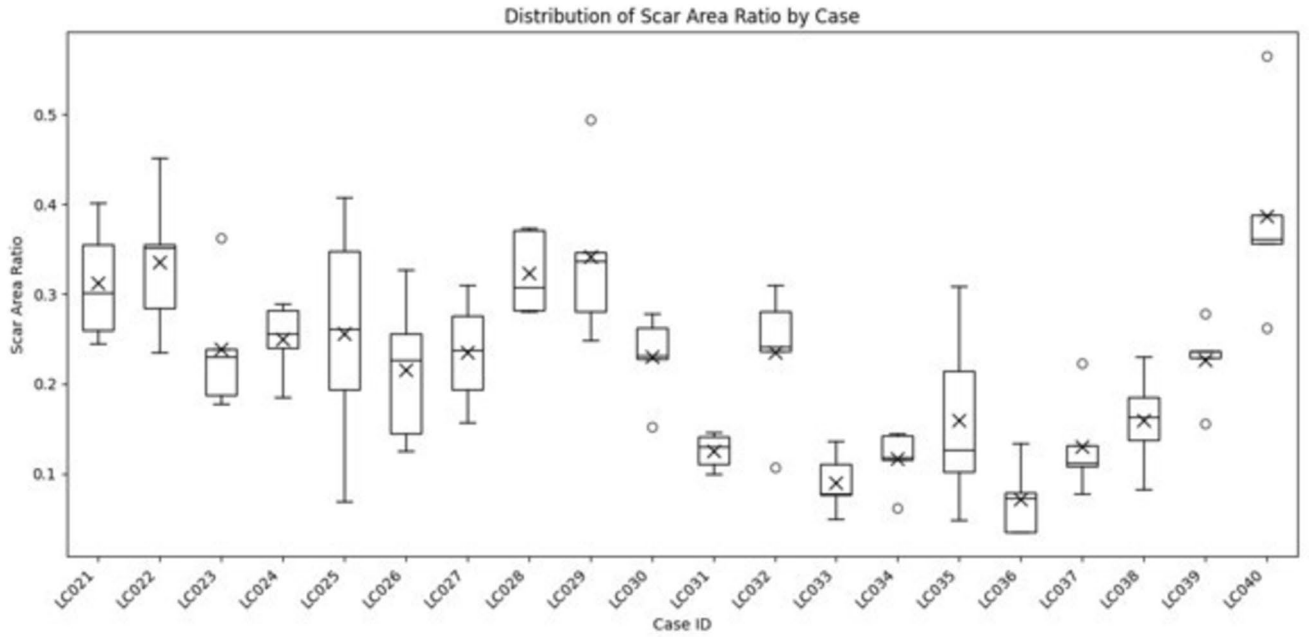

Fig. 3 Distribution of scar area ratio by case in the test dataset

The model was trained on 2000 laparoscopic images automatically selected from 21 patients with LC obtained from an external institution. These images were used exclusively for selecting bleeding and non-bleeding frames and were not included in the test dataset of the present study. Frame selection was performed using a Vision Transformer (ViT)-based binary classifier that discriminated between bleeding and non-bleeding frames [25]. Model training employed adversarial loss, cycle-consistency loss, and perceptual loss [26] to preserve fine anatomical structures.

Scar region prediction was performed using the HyperSeg model, a real-time patch-wise hypernetwork architecture that demonstrates robust generalization across domains. To integrate predictions from the original and translated images, an

uncertainty-aware fusion strategy was implemented. Entropy maps were computed for each prediction and subsequently used as confidence-based weights during pixel-wise fusion. Although Bayesian models were not directly employed, the uncertainty-aware fusion strategy was conceptually inspired by approaches such as Monte Carlo Dropout [27] and probabilistic segmentation methods, including the Probabilistic U-Net [28], both of which are designed to enhance robustness under conditions of uncertainty.

Inference workflow and output generation

During inference, both the original and CycleGAN-translated images were processed by the HyperSeg model to generate scar probability maps (SPMs), denoted as $p_A(x)$ and $p_B(x)$, respectively. The raw network outputs were normalized using the softmax function, converting logits into probabilities ranging from 0 to 1, with the sum across classes equal to 1. These probability maps were subsequently combined using an entropy-based, uncertainty-aware fusion strategy to generate a final prediction map, thereby enhancing scar visualization even in bleeding-compromised areas.

Let $p(x)$ denote the predicted probability map obtained after applying the HyperSeg model to an image. Pixel-wise uncertainties, $u_A(x)$ and $u_B(x)$, were defined as the entropies of the corresponding probability maps (Eq. 1). The associated confidence maps were calculated as described in Eq. 2. The fused probability map, $p_{fused}(x)$, was generated through confidence-weighted integration (Eq. 3). A final binary mask was produced by applying a confidence threshold q to the fused probability map (Eq. 4), with $q = 0.37$ determined empirically from preliminary validation. Higher values of q yielded more conservative predictions by excluding low-confidence areas, whereas lower q values increased detections at the risk of introducing false positives.

$$u(x) = -(p(x)\log p(x)) + (1 - p(x))\log(1 - p(x))) \tag{1}$$

$$w(x) = \frac{1}{u(x) + \varepsilon} \tag{2}$$

$$p_{fused}(x) = \frac{w_A(x) \cdot p_A(x) + w_B(x) \cdot p_B(x)}{w_A(x) + w_B(x)} \tag{3}$$

$$M(x) = \begin{cases} 1 & if\ p_{fused}(x) > \theta \\ 0 & otherwise \end{cases} \tag{4}$$

Figure 4 presents illustrative examples of the original image (B), the translated pseudo non-bleeding image (D), the corresponding probability maps (B_SPM and D_SPM), and the fused scar prediction results. The scar probability maps (SPMs) represent model-generated scores for scarring produced by HyperSeg rather than direct probabilities. Low-score regions (cool colors) surrounded by high-score regions (warm colors) may still be inferred as scarring following uncertainty-aware fusion. For clarity, the boundaries and overall visibility of the scarred regions inferred by the proposed method were enhanced to facilitate interpretation.

In this study, inference was performed on all incoming endoscopic video frames, including those affected by motion blur, defocus, smoke, or blood occlusion. Such degradations

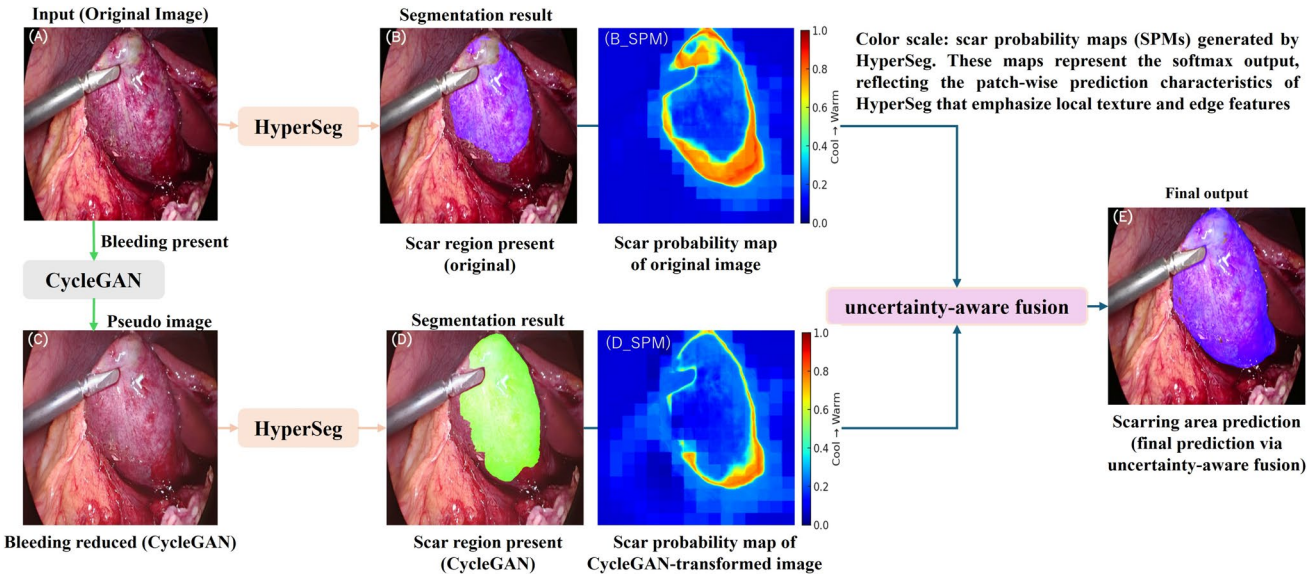

Fig. 4 Overview of the proposed AI-assisted workflow for scar region visualization under intraoperative bleeding conditions. AI artificial intelligence

were assumed to be temporary rather than persistent. Although the model continuously processed frames in real time, the instructional information generated was intended to be referenced primarily when the surgical field was stable. Accordingly, the short-term presence of degraded frames was considered to exert only a limited impact on the accuracy and reliability of the guidance provided.

Evaluation strategy

Objective evaluation was performed using the test dataset, which consisted of external-institution images independently annotated by three experienced surgeons. All annotators were board-certified endoscopic surgeons specializing in hepatobiliary-pancreatic surgery, with 18, 22, and 30 years of clinical practice, respectively. For ground-truth creation, the test dataset originally comprised 100 frames; however, one corrupted frame (along with its annotation) was excluded, leaving 99 frames for analysis. Scar regions were annotated independently by each surgeon to minimize potential bias.

Segmentation performance was assessed by comparing the predicted scar regions with these ground-truth annotations using the Dice coefficient, as defined in Eqs. (5)–(7). In this context, *TP* denotes true positives, *FP* denotes false positives, *TN* denotes true negatives, and FN denotes false negatives.

$$\text{Dice} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{5}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{6}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{7}$$

Comparisons were made between conventional predictions and the proposed fused predictions. Statistical significance was assessed using the Wilcoxon signed-rank test, with a P value of < 0.05 considered significant. To evaluate the influence of bleeding severity, frames were stratified into high-bleeding (probability > 0.5) and low-bleeding (probability ≤ 0.5) groups according to the ViT-based classifier. Separate Wilcoxon tests were conducted for each subgroup.

Subjective evaluation was performed by the same three surgeons who annotated the test dataset. Each surgeon was presented with paired video sequences containing original predictions (Images A and B) and fused predictions (Image E). A 5-point Likert scale was used to assess overall visibility, boundary delineation, and clinical usability.

Results

Visual example of improvement

Figure 5 illustrates a representative patient in whom intraoperative bleeding significantly reduced the visibility of scarred tissue within Calot's triangle. In this example, conventional AI-based segmentation (B) failed to detect a substantial portion of the scarred area due to visual obstruction.

By contrast, the CycleGAN-translated pseudo non-bleeding image (C) improved structural clarity, enabling the segmentation model to more accurately identify scarred features (D). The final uncertainty-aware fusion output (E) effectively combined predictions from both the original and

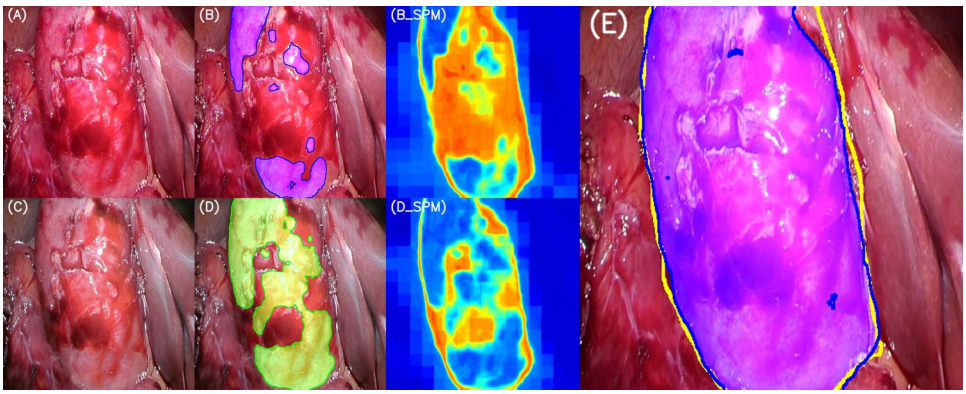

Fig. 5 Representative example demonstrating improvement in scarring detection using the proposed method. **A** Original laparoscopic image with bleeding contamination. **B** Prediction results from conventional segmentation AI applied to Image A, with corresponding scar probability map (B_SPM). **C** CycleGAN-translated pseudo non-bleeding image. **D** Prediction results from the AI segmentation applied to the translated image, with the corresponding scar probability map (D_SPM). **E** Final fused result, highlighting scarred regions in blue and overlaid with yellow contours from the ground-truth annotation. AI artificial intelligence, CycleGAN cycle-consistent generative adversarial network, SPM scar probability map

translated images, yielding results that closely approximated the expert-annotated ground truth. This case highlights the potential of the proposed system to support safer dissection decisions under bleeding conditions that are typically challenging for both surgeons and AI models. The color scale in the probability maps (B_SPM, D_SPM) represents model-assigned scores for scarring generated by HyperSeg and does not directly correspond to the final fused probabilities. Owing to the non-linear integration in the uncertainty-aware fusion process, regions in B_SPM and D_SPM with low scores (cool colors) that are adjacent to or surrounded by high-score regions (warm colors) may be elevated in value and consequently inferred as scarring in the fused result.

The proposed system achieved an average processing time of 0.06 s per frame on an NVIDIA RTX A5000 GPU. This measurement encompassed the entire workflow, including CycleGAN-based image translation (conversion of bleeding images to non-bleeding), segmentation of scarred regions, uncertainty-aware fusion of multiple inference results, and real-time overlay display onto the endoscopic video. These results demonstrate that the system operates at a speed suitable for real-time intraoperative use, without disrupting the surgical workflow.

Quantitative evaluation

To quantitatively assess the clinical utility of the proposed method, segmentation performance was compared before and after applying the uncertainty-aware fusion strategy across 99 evaluation frames independently annotated by three board-certified endoscopic surgeons who served as evaluators. The Dice coefficient was used as the primary metric, and statistical significance was determined using the Wilcoxon signed-rank test.

Evaluator 1: The mean Dice coefficient improved from 0.643 (95% confidence interval (CI) 0.603–0.683) to 0.792 (95% CI 0.754–0.830), with a median increase of 0.121 (95% CI 0.103–0.142), $P = 8.15 \times 10^{-13}$, Cohen's $d = 0.85$.
Evaluator 2: The mean Dice coefficient improved from 0.626 (95% CI 0.584–0.668) to 0.721 (95% CI 0.685–0.757), with a median increase of 0.066 (95% CI 0.032–0.104), $P = 1.13 \times 10^{-8}$, Cohen's $d = 0.62$.
Evaluator 3: The mean Dice coefficient improved from 0.704 (95% CI 0.663–0.745) to 0.874 (95% CI 0.852–0.896), with a median increase of 0.117 (95% CI 0.096–0.140), $P = 9.66 \times 10^{-16}$, Cohen's $d = 0.92$.

These results demonstrate that integrating CycleGAN-based image translation with uncertainty-aware fusion significantly improves the accuracy of scarred region detection in challenging intraoperative scenes. Figure 6 presents boxplots comparing the Dice scores before and after fusion across all evaluators, with Wilcoxon signed-rank tests confirming significant improvements ($P < 0.001$).

To further examine the effect of bleeding severity, the evaluation frames were stratified into high-bleeding (predicted probability > 0.5) and low-bleeding (≤ 0.5) subgroups using a ViT-based image classifier. Within both subgroups, the proposed method consistently outperformed conventional segmentation, as summarized in Table 1. Improvements

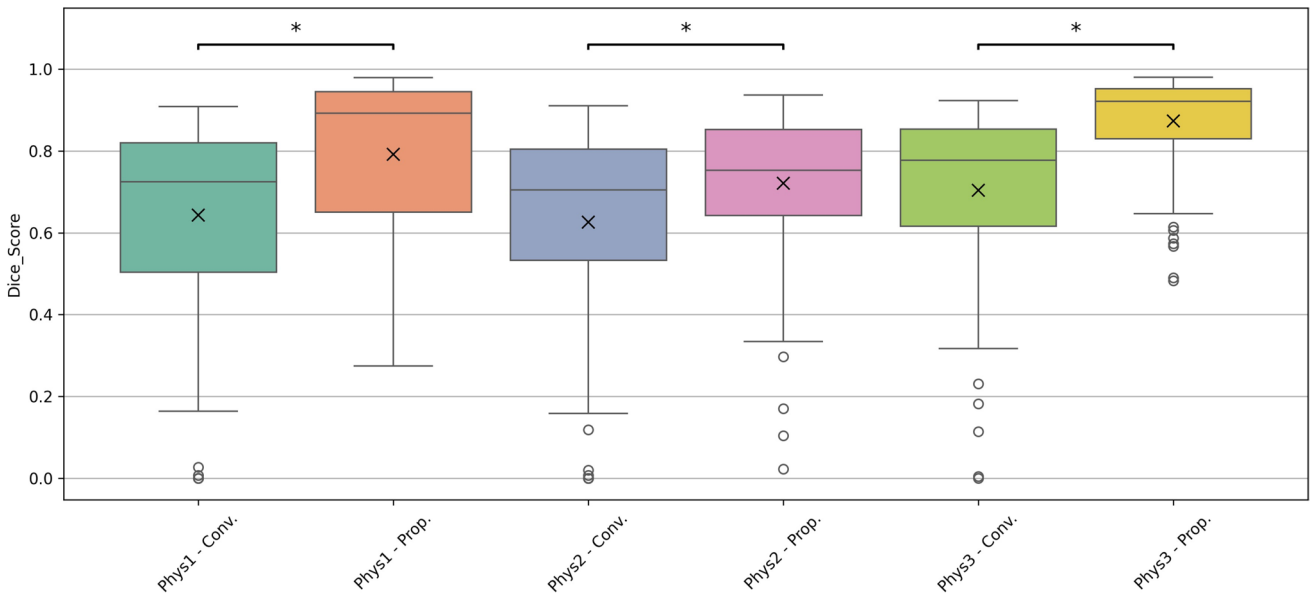

Fig. 6 Boxplot comparison of Dice scores between conventional predictions (Image B) and the proposed method (Image E) across four external evaluator. Asterisks denote significant improvements (Wilcoxon signed-rank test, $P < 0.001$ for all comparisons)

were significant across all evaluators regardless of bleeding intensity ($P < 0.001$), underscoring the method's robustness under both clean and visually compromised conditions.

These findings highlight the clinical applicability of the proposed approach in real-world surgical settings, particularly in patients with AC where bleeding is frequent and visibility is impaired. By maintaining high performance even under adverse visual conditions, the system may assist general surgeons in making safer intraoperative decisions.

No significant interaction was observed between bleeding severity and the magnitude of improvement, suggesting that the method remains stable even under bleeding-induced image quality degradation.

Subjective evaluation by expert surgeons

To further assess the perceived utility of the proposed method, three expert laparoscopic surgeons independently evaluated 20 surgical videos from the test dataset using a 5-point Likert scale (1 = strongly disagree to 5 = strongly agree). The evaluation items addressed improvements in scar region detection, mitigation of bleeding-related visual obstruction, clinical reliability of the results, and the usefulness of the output as intraoperative support. Additionally, a binary-choice item was included to determine whether the conventional or proposed method was more helpful for scar region recognition.

As presented in Table 2, all items except Question 4 achieved an average score of ≥ 4.1, reflecting a generally favorable impression of the proposed method. Notably, the positive response rate (defined as the proportion of ratings of 4 or 5) exceeded 75% for four of the five Likert scale items. By contrast, concern regarding oversegmentation (Question 4) received a relatively lower score of 3.40 ± 0.98 and a positive response rate of 61.7%, indicating more mixed evaluations for this aspect. For the binary-choice item (Question 6), the proposed method was preferred in 91.7% of the patients, demonstrating a clear advantage over the conventional method as a visual support tool for scar region identification.

Collectively, these findings support the clinical applicability of the proposed approach and highlight its potential to enhance intraoperative decision-making, particularly in

Table 1 Subgroup analysis of Dice coefficient improvements stratified by bleeding severity and external evaluator

	Group	N	Dice of (B)	Dice of (E)	⊿Dice	*P*-value
Evaluator 1	Low-bleeding	41	0.617	0.778	+0.161	1.95×10^{-7}
	High-bleeding	58	0.675	0.811	+0.136	3.16×10^{-8}
Evaluator 2	Low-bleeding	41	0.606	0.690	+0.084	4.45×10^{-4}
	High-bleeding	58	0.650	0.757	+0.107	3.41×10^{-6}
Evaluator 3	Low-bleeding	41	0.684	0.859	+0.174	1.69×10^{-7}
	High-bleeding	58	0.728	0.892	+0.163	2.44×10^{-10}

Table 2 Summary of subjective evaluation results

No.	Question	Mean score ± SD	Positive response (4–5) rate (%)
1	Do you perceive that the detection of scar regions has improved?	4.38 ± 0.76	86.7%
2	Do you perceive that the influence of bleeding-related visual obstruction has been mitigated?	4.23 ± 0.87	75.0%
3	Do you consider this method useful as an adjunct to intraoperative decision-making?	4.23 ± 0.59	91.7%
4	Were you concerned about the over-detection of scar regions (false positives)?	3.40 ± 0.98	61.7%
5	Do you consider the results to be clinically reliable?	4.18 ± 0.72	81.7%
6	Which video did you find more helpful for identifying scar regions?	Proposed: 91.7% / Conventional: 8.3%	

bleeding-prone surgical scenarios where conventional methods frequently underperform.

Discussion

This study demonstrated that our AI-assisted workflow, which integrates CycleGAN-based image translation with uncertainty-aware fusion, significantly enhances the detection of scarred tissue during LC, particularly under bleeding-compromised conditions. In scar probability maps (SPMs) generated by HyperSeg, boundary regions frequently exhibited higher probability values than the more homogeneous central areas. This pattern reflects the patch-wise prediction mechanism of HyperSeg, which prioritizes local texture changes and boundary features. As a result, central scar regions often appeared with relatively low scores. However, the uncertainty-aware fusion process compensated for these discrepancies, yielding a more coherent and consistent representation of the scarred tissue. Consequently, the proposed method achieved higher Dice coefficients compared with conventional segmentation across all evaluators and maintained robust performance under both high- and low-bleeding conditions. From a clinical perspective, the model design did not explicitly incorporate the anatomical relationship with Calot's triangle. Nevertheless, this landmark is crucial for safe dissection, and the proposed system may assist by highlighting scarred regions adjacent to the cystic duct and artery, which are often obscured by intraoperative bleeding.

Intraoperative scarring within Calot's triangle is a major risk factor for BDI, particularly in patients with acute or recurrent cholecystitis. Accurate recognition of scarring is therefore critical for planning bailout procedures such as subtotal cholecystectomy or conversion to open surgery [18]. Intraoperative bleeding often obscures the operative field, making reliable recognition difficult for both surgeons and AI-based systems. The proposed method addresses this challenge by generating pseudo non-bleeding representations to restore visibility, enabling more dependable identification of scarred areas that might otherwise be overlooked.

A key innovation of this approach lies in the use of unpaired image-to-image translation via CycleGAN, which enables the generation of clean images without requiring aligned image pairs. The incorporation of perceptual loss further preserved anatomical fidelity, ensuring that the translated images retained clinical interpretability. The subsequent uncertainty-aware fusion adaptively weighted outputs according to pixel-wise entropy, enhancing consistency and robustness under variable intraoperative conditions.

Effectiveness was demonstrated not only through objective performance metrics but also through subjective evaluation by expert surgeons. High Likert scale ratings for scar visibility and clinical usefulness suggest that the system can provide actionable intraoperative insights. All evaluators showed significant improvements in Dice coefficients ($P < 0.001$), with effect sizes ranging from moderate ($d = 0.62$) to large ($d = 0.92$). These differences may reflect variations in evaluation criteria, subjective judgment tendencies, or the relative difficulty of the evaluated patients. A tendency toward broader segmentation in some contexts may also have influenced the results, highlighting the inherent trade-off between sensitivity and specificity in safety–critical applications. Incorporating active learning or expert-in-the-loop corrections could further refine predictions and improve alignment with surgical intuition.

Although novice surgeons were not included in this evaluation, real-time visualization of scarred regions may enhance situational awareness and decision-making, potentially accelerating skill acquisition. However, incorrect AI indications could misguide surgical decisions. False positives may lead to the overestimation of anatomical changes, whereas false negatives could result in missed critical findings. These risks can be mitigated by standard intraoperative confirmation using irrigation, palpation with forceps, and integration with preoperative imaging. Such measures ensure that AI output complements rather than replaces surgical judgment.

Compared with prior studies that applied GAN-based methods for general feature enhancement or domain adaptation under smoke or lighting variations [19–21], the present study specifically addresses scarring—a clinically critical finding that directly influences intraoperative decision-making and patient safety. To our knowledge, this is among the first studies to combine generative translation with entropy-based fusion for this purpose, representing a meaningful advancement in surgical AI.

This study has some limitations. First, although the dataset was derived from real surgical procedures, its size was modest and limited to a single institution. Validation using multi-institutional and multi-device datasets will be essential to ensure robustness across diverse surgical environments. Second, subjective evaluation was conducted by only three expert surgeons; expanding the reviewer panel would help capture inter-rater variability and strengthen usability validation. Third, although the system achieved real-time inference (approximately 0.06 s per frame on current GPUs), integration into existing surgical platforms will require additional engineering and usability testing. In clinical practice, overlay displays under compromised visibility may obscure important visual cues, such as microbleeding. To address this concern, an operational setup is envisioned in which AI-generated instructional information is presented on a secondary monitor rather than overlaid on the primary endoscopic view. This configuration would allow surgeons to maintain direct visualization of the live operative field while consulting scarred-area information as needed. The primary role of the AI system is to provide alerts regarding the presence of

scarring within the surgical field, prompting reconsideration of dissection strategies irrespective of the surgeon's level of experience.

In the context of LC, three representative intraoperative scenarios are envisioned for the proposed system: (1) identification of scarred areas temporarily obscured by bleeding, enabling prompt recognition of potential hazards once hemostasis is restored; (2) differentiation between scarred and non-scarred tissue in severely inflamed operative fields, thereby facilitating safer dissection planes; and (3) confirmation of residual scarred areas after total or subtotal cholecystectomy, reducing the risk of inadvertent bile duct or vascular injury.

In the future, this framework may be extended beyond scarring detection to include visualization of other critical intraoperative findings, such as vascular structures, nerve pathways, and oncologic margins. Incorporating real-time feedback loops between the AI system and surgeons could foster adaptive learning and enhance intraoperative trust. For clinical implementation, careful consideration of workflow integration, as well as regulatory and ethical requirements, will be essential to ensure safe and effective deployment.

In summary, a clinically meaningful AI framework was developed and validated to enhance intraoperative detection of scarring, even in bleeding-compromised fields. Through the integration of generative translation and uncertainty-aware fusion, the proposed method addresses a practical unmet need in LC and demonstrates potential for broader incorporation into intraoperative guidance systems aimed at improving surgical safety.

Conclusion

This study introduced a hybrid AI framework designed to support the recognition of scar tissue during LC under bleeding-contaminated conditions. By integrating unpaired image-to-image translation using CycleGAN with an uncertainty-aware fusion strategy, the system enhanced scar detection performance while preserving anatomical interpretability. Quantitative analyses demonstrated consistent improvements across all evaluators, and subjective assessments confirmed the clinical relevance and usability of the approach.

The proposed framework addresses a critical gap in surgical image analysis, namely, the reliable identification of scar tissue under compromised visibility, and demonstrates feasibility for real-time intraoperative support. With further validation and seamless integration into surgical workflows, this approach has the potential to strengthen operative decision-making, reduce the risk of BDI, and improve outcomes in complex cholecystectomy cases.

Acknowledgements We would like to thank Dr. Koji Asai of the Toho University Medical Center Ohashi Hospital, Dr. Yasutoshi Mori of the University of Occupational and Environmental Health, Japan, and Dr. Hajime Fujishima of Fujishima Hospital, who served as external evaluators of this study. We would also like to thank Editage (www.editage.jp) for providing English language editing assistance.

Author contributions Tatsushi Tokuyasu contributed toward conceptualization, methodology, design, data curation, formal analysis, software development, drafting, writing—original draft, and writing-review and editing. Subal Ikeda contributed toward design, data curation, software development, and formal analysis. Hiroki Orimoto contributed toward conceptualization, design, data curation, and writing—original draft. Teijiro Hirashita contributed toward conceptualization, design, data curation, writing—original draft, and writing—review and editing. Yuichi Endo contributed toward conceptualization, design, data curation, clinical validation, and writing—review and editing. Masafumi Inomata contributed toward funding acquisition, project administration, supervision, and writing—review and editing. All authors have given final approval for the version to be published.

Funding This study was partially supported by collaborative research funding provided by Olympus Corporation.

Declarations

Disclosures Drs. Tatsushi Tokuyasu, Subaru Ikeda, Hiroki Orimoto, Teijiro Hirashita, Yuichi Endo, and Masafumi Inomata declare that they have no conflicts of interest or financial relationships to disclose.

References

1. Soper NJ, Stockmann PT, Dunnegan DL, Ashley SW (1992) Laparoscopic cholecystectomy. The new "gold standard"? Arch Surg 127:917–921. https://doi.org/10.1001/archsurg.1992.01420080051008
2. Strasberg SM, Hertl M, Soper NJ (1995) An analysis of the problem of biliary injury during laparoscopic cholecystectomy. J Am Coll Surg 180:101–125
3. Melton GB, Lillemoe KD, Cameron JL, Sauter PA, Coleman J, Yeo CJ (2002) Major bile duct injuries associated with laparoscopic cholecystectomy: effect of surgical repair on quality of life. Ann Surg 235:888–895. https://doi.org/10.1097/00000658-200206000-00018
4. Wakabayashi G, Iwashita Y, Hibi T, Takada T, Strasberg SM, Asbun HJ, Endo I, Umezawa A, Asai K, Suzuki K, Mori Y, Okamoto K, Pitt HA, Han HS, Hwang TL, Yoon YS, Yoon DS, Choi IS, Huang WSW, Giménez ME, Garden OJ, Gouma DJ, Belli G, Dervenis C, Jagannath P, Chan ACW, Lau WY, Liu KH, Su CH, Misawa T, Nakamura M, Horiguchi A, Tagaya N, Fujioka S, Higuchi R, Shikata S, Noguchi Y, Ukai T, Yokoe M, Cherqui D, Honda G, Sugioka A, de Santibañes E, Supe AN, Tokumura H, Kimura T, Yoshida M, Mayumi T, Kitano S, Inomata M, Hirata K, Sumiyama Y, Inui K, Yamamoto M (2018) Tokyo guidelines 2018: surgical management of acute cholecystitis: safe steps in laparoscopic cholecystectomy for acute cholecystitis (with videos). J Hepatobiliary Pancreat Sci 25:73–86. https://doi.org/10.1002/jhbp.517
5. Strasberg SM, Brunt LM (2017) The critical view of safety: why it is not the only method of ductal identification within the standard of care in laparoscopic cholecystectomy. Ann Surg 265:464–465. https://doi.org/10.1097/SLA.0000000000002054

6. Way LW, Stewart L, Gantert W, Liu K, Lee CM, Whang K, Hunter JG (2003) Causes and prevention of laparoscopic bile duct injuries: analysis of 252 cases from a human factors and cognitive psychology perspective. Ann Surg 237:460–469. https://doi.org/10.1097/01.SLA.0000060680.92690.E9
7. Saguil A, Wayne D, DeDonato C, Galvin S (2018) A systematic review of bile duct injury prevention in laparoscopic cholecystectomy. Int J Surg 54:325–336. https://doi.org/10.1016/j.ijsu.2018.11.006
8. Tokuyasu T, Iwashita Y, Matsunobu Y, Kamiyama T, Ishikake M, Sakaguchi S, Ebe K, Tada K, Endo Y, Etoh T, Nakashima M, Inomata M (2021) Development of an artificial intelligence system using deep learning to indicate anatomical landmarks during laparoscopic cholecystectomy. Surg Endosc 35:1651–1658. https://doi.org/10.1007/s00464-020-07548-x
9. Aoyama Y, Matsunobu Y, Etoh T, Suzuki K, Fujita S, Aiba T, Fujishima H, Empuku S, Kono Y, Endo Y, Ueda Y, Shiroshita H, Kamiyama T, Sugita T, Morishima K, Ebe K, Tokuyasu T, Inomata M (2024) Artificial intelligence for surgical safety during laparoscopic gastrectomy for gastric cancer: indication of anatomical landmarks related to postoperative pancreatic fistula using deep learning. Surg Endosc 38:5601–5612. https://doi.org/10.1007/s00464-024-11117-x
10. Shinozuka K, Turuda S, Fujinaga A, Nakanuma H, Kawamura M, Matsunobu Y, Tanaka Y, Kamiyama T, Ebe K, Endo Y, Etoh T, Inomata M, Tokuyasu T (2022) Artificial intelligence software available for medical devices: surgical phase recognition in laparoscopic cholecystectomy. Surg Endosc 36:7444–7452. https://doi.org/10.1007/s00464-022-09160-7
11. Turan M, Vedula SS, Dornberger K, Scheikl T, árová R, Kwon OD, Stauder S, Mirota DJ, Neumuth T, Vrtovec T, Maier-Hein K, Kashyap D, Kranzfelder M (2021) Context-aware surgical phase recognition using hybrid deep models. Int J Comput Assist Radiol Surg 16:1145–1154
12. Madani A, Namazi B, Altieri MS, Hashimoto DA, Rivera AM, Pucher PH, Navarrete-Welton A, Sankaranarayanan G, Brunt LM, Okrainec A, Alseidi A (2022) Artificial intelligence for intraoperative guidance: using semantic segmentation to identify surgical anatomy during laparoscopic cholecystectomy. Ann Surg 276:363–369. https://doi.org/10.1097/SLA.0000000000004594
13. Mascagni P, Vardazaryan A, Alapatt D, Urade T, Emre T, Fiorillo C, Pessaux P, Mutter D, Marescaux J, Costamagna G, Dallemagne B, Padoy N (2022) Artificial intelligence for surgical safety: automatic assessment of the critical view of safety in laparoscopic cholecystectomy using deep learning. Ann Surg 275:955–961. https://doi.org/10.1097/SLA.0000000000004351
14. Endo Y, Tokuyasu T, Mori Y, Asai K, Umezawa A, Kawamura M, Fujinaga A, Ejima A, Kimura M, Inomata M (2023) Impact of AI system on recognition for anatomical landmarks related to reducing bile duct injury during laparoscopic cholecystectomy. Surg Endosc 37:5752–5759. https://doi.org/10.1007/s00464-023-10224-5
15. Orimoto H, Hirashita T, Ikeda S, Amano S, Kawamura M, Kawano Y, Takayama H, Masuda T, Endo Y, Matsunobu Y, Shinozuka K, Tokuyasu T, Inomata M (2025) Development of an artificial intelligence system to indicate intraoperative findings of scarring in laparoscopic cholecystectomy for cholecystitis. Surg Endosc 39:1379–1387. https://doi.org/10.1007/s00464-024-11514-2
16. Liu Y, Xue Y, Yang X, Zhang J, Li H, Chen Q (2020) Segmentation of bleeding regions in laparoscopic surgery videos using deep learning. Comput Biol Med 122:103785
17. Iwashita Y, Hibi T, Ohyama T, Honda G, Yoshida M, Miura F, Takada T, Han HS, Hwang TL, Shinya S, Suzuki K, Umezawa A, Yoon YS, Choi IS, Huang WS, Chen KH, Watanabe M, Abe Y, Misawa T, Nagakawa Y, Yoon DS, Jang JY, Yu HC, Ahn KS, Kim SC, Song IS, Kim JH, Yun SS, Choi SH, Jan YY, Shan YS, Ker CG, Chan DC, Wu CC, Lee KT, Toyota N, Higuchi R, Nakamura Y, Mizuguchi Y, Takeda Y, Ito M, Norimizu S, Yamada S, Matsumura N, Shindoh J, Sunagawa H, Gocho T, Hasegawa H, Rikiyama T, Sata N, Kano N, Kitano S, Tokumura H, Yamashita Y, Watanabe G, Nakagawa K, Kimura T, Yamakawa T, Wakabayashi G, Mori R, Endo I, Miyazaki M, Yamamoto M (2017) An opportunity in difficulty: Japan-Korea-Taiwan expert Delphi consensus on surgical difficulty during laparoscopic cholecystectomy. J Hepatobiliary Pancreat Sci 24:191–198. https://doi.org/10.1002/jhbp.440
18. Asai K, Iwashita Y, Ohyama T, Endo I, Hibi T, Umezawa A, Suzuki K, Watanabe M, Kurata M, Mori Y, Higashida M, Kumamoto Y, Shindoh J, Yoshida M, Honda G, Misawa T, Abe Y, Nagakawa Y, Toyota N, Yamada S, Norimizu S, Matsumura N, Sata N, Sunagawa H, Ito M, Takeda Y, Nakamura Y, Rikiyama T, Higuchi R, Gocho T, Homma Y, Hirashita T, Kanemoto H, Nozawa M, Watanabe Y, Kohga A, Yazawa T, Tajima H, Nakahira S, Asaoka T, Yoshioka R, Fukuzawa J, Fujioka S, Hata T, Haruta H, Asano Y, Nomura R, Matsumoto J, Kameyama N, Miyoshi A, Urakami H, Seyama Y, Morikawa T, Kawano Y, Ikoma H, Kim DH, Takada T, Yamamoto M (2022) Application of a novel surgical difficulty grading system during laparoscopic cholecystectomy. J Hepatobiliary Pancreat Sci 29:758–767. https://doi.org/10.1002/jhbp.1068
19. Okamoto T, Ohnishi T, Kawahira H, Dergachyava O, Jannin P, Haneishi H (2019) Real-time identification of blood regions for hemostasis support in laparoscopic surgery. Signal Image Video Process 13:405–412. https://doi.org/10.1007/s11760-018-1369-7
20. Zhu JY, Park T, Isola P, Efros AA (2017) Unpaired image-to-image translation using cycle-consistent adversarial networks. Proc IEEE Int Conf Comput Vis. https://doi.org/10.1109/ICCV.2017.244
21. Alapatt D, Murali A, Satyanaik S, Wang X, Mascagni P, Padoy N (2022) Domain adaptation in surgical video analysis using style transfer. IEEE Trans Med Imaging 41:1170–1181
22. Zhang Z, Rosa B, Nageotte F (2021) Surgical tool segmentation using generative adversarial networks with unpaired training data. IEEE Robot Autom Lett 6:6266–6273. https://doi.org/10.1109/LRA.2021.3092302
23. Sahu M, Mukhopadhyay A, Zachow S (2021) Simulation-to-real domain adaptation with teacher–student learning for endoscopic instrument segmentation. Int J Comput Assist Radiol Surg 16:849–859. https://doi.org/10.1007/s11548-021-02383-4
24. Nirkin Y, Wolf L, Hassner T (2021) HyperSeg: patch-wise hypernetwork for real-time semantic segmentation. In: Proceedings of the IEEE computer society conference on computer vision and pattern recognition. IEEE Computer Society. https://doi.org/10.1109/cvpr46437.2021.00405
25. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, Uszkoreit J, Houlsby N (2021) An image is worth 16×16 words: transformers for image recognition at scale. Proc Int Conf Learn Represent (ICLR). Preprint at https://arxiv.org/abs/2010.11929
26. Johnson J, Alahi A, Fei-Fei L (2016) Perceptual losses for real-time style transfer and super-resolution. In: Leibe B, Matas J, Sebe N, Welling M (eds) European conference on computer vision. Springer International Publishing, Cham, pp 694–711
27. Gal Y, Ghahramani Z (2016) Dropout as a Bayesian approximation: representing model uncertainty in deep learning. Preprint at https://arxiv.org/abs/1506.02142
28. Kohl SA, Romera-Paredes B, Meyer C, De Fauw J, Ledsam JR, Maier-Hein KH, Eslami S, Jimenez Rezende D, Ronneberger O (2018) A probabilistic U-Net for segmentation of ambiguous

images. Adv Neural Inf Process Syst 31:6965–6975. https://doi.org/10.48550/arXiv.1806.05034