Critical View of Safety Assessment in Laparoscopic Cholecystectomy via Segment Anything Model

Yunfan Li
Department of Computer Science
Stony Brook University
Stony Brook, USA
yunfli@cs.stonybrook.edu

Haibin Ling
Department of Computer Science
Stony Brook University
Stony Brook, USA
hling@cs.stonybrook.edu

IV Ramakrishnan
Department of Computer Science
Stony Brook University
Stony Brook, USA
ram@cs.stonybrook.edu

Prateek Prasanna
Department of Biomedical Informatics
Stony Brook University
Stony Brook, USA
prateek.prasanna@stonybrook.edu

Aaron Sasson
Department of Surgery
Stony Brook University Hospital
Stony Brook, USA
aaron.sasson@stonybrookmedicine.edu

Himanshu Gupta
Department of Computer Science
Stony Brook University
Stony Brook, USA
hgupta@cs.stonybrook.edu

***Abstract*—Laparoscopic Cholecystectomy (LC) is a minimally invasive surgery for the removal of diseased gallbladders. Compared to traditional open cholecystectomy, LC procedures is associated with significantly shorter recovery period, but has an increased chance of bile duct injuries (BDIs). Critical view of safety (CVS) is an important validation method and safety protocol which has a set of conditions that can be visually identified during LC surgeries. In this paper, we approach the problem of automated CVS prediction by combining state-of-the-art object detection methods and prompting the Segment Anything model to achieve more accurate localization of anatomical structures and classification of CVS conditions. When evaluated on our dataset of 5,750 frames with CVS annotations, our method achieved competitive results on frame-level CVS condition prediction, and around 20% improvement on video-level CVS assessment compared to previous SoTA LG-CVS.**

***Index Terms*—Laparoscopic Cholecystectomy, Critical View of Safety, Object Detection, Segment Anything Model**

I. Introduction

Recent advances in surgical video analysis have demonstrated significant potential for deep learning methods to gain important insights into surgical procedures and provide intraoperative guidance to improve patient safety. In particular, laparoscopic cholecystectomy (LC) has received increased interest in recent years as it is one of the most popular procedures performed each year. Today, 92% of cholecystectomies are performed laparoscopically, as LCs are associated with shorter hospital stays and quicker convalescence periods as compared to open procedures. However, the downside of LCs is the resulting 3-fold increase in the incidence of bile duct injuries (BDIs) [14]–[16]. A safety protocol termed Critical View of Safety (CVS) has been developed and widely adopted over the years, with the goal of minimizing the misidentification of ducts and thus reducing the incidence of BDIs. Despite strong evidence of the effectiveness of CVS, the incidence of BDIs has not decreased over the past decades, with the main reason being the insufficient implementation and understanding of CVS criteria by surgeons [17]. In recent years, machine learning (ML) based methods have been proposed to automatically detect the attainment of CVS in LC surgeries [1], [4], [5], [11], providing additional verification for surgeons, which can potentially reduce the incidence of BDIs.

In this paper, we developed a technique to assess CVS based on its three criteria. In particular, we propose to formulate the CVS assessment problem into a combination of object detection and semantic segmentation problems, and combine state-of-the-art object detection methods with the power of the Segment Anything Model (SAM) for accurate identification of anatomical structures and assessment of CVS criteria in LC images. We demonstrated the effectiveness of our method by comparing it against prior classification-based method *LG-CVS* and showed better results in both frame-level and video-level performance in CVS assessment.

II. Background

Laparoscopic Cholecystectomy. The gallbladder is a small organ underneath the liver that concentrates and stores bile, a fluid made by the liver to help digestion. Inflammation and infection of the gallbladder can cause sharp and constant abdominal pain and may necessitate surgical removal of the gallbladder, which is most often done via Laparoscopic cholecystectomy (LC), a minimally invasive procedure. LC, performed through four small incisions, uses a camera to visualize the insides of the abdomen and surgical tools to remove the gallbladder. Removal of the gallbladder essentially entails exposing (by removing the fat and fibrous tissues) and cutting the only two structures that connect it to the rest of the body: the cystic duct (CD) and the cystic artery.

<u>The Risks of LC.</u> Compared to cholecystectomy, LC being minimally invasive is associated with low rate of complications, shorter hospital stays, and more comfortable postoperative periods. The main side-effect of LCs is a higher increase in the incidence of bile duct surgeries (BDIs), which happens mainly due to misidentification of the two structures, viz.,

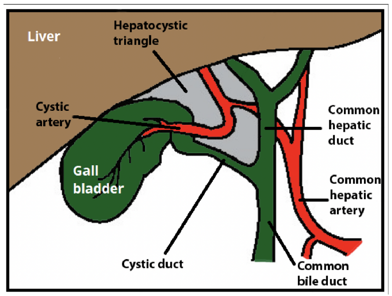

Fig. 1. Anatomy of the hepatocystic triangle. [18]

cystic duct and cystic artery, that are to be cut to remove the gallbladder. Unfortunately, BDIs due to LCs can lead to significant morbidity [21], endangering patient's life and safety, and increasing medical litigation and healthcare costs as discussed below.

BDI Complications. In general, BDIs resulting from LCs may also lead to biliary peritonitis, bile duct stenosis, obstructive jaundice, biliary tract infection, and even more serious complications such as biliary cirrhosis, portal hypertension, liver atrophy, etc., which can even endanger the patient's life and safety [22], [23]. Overall, BDIs frequently result in a 3-fold increase in the 1-year mortality rate [24] and a reduction in quality of life, while driving up the medical litigation [25] and healthcare costs to over a billion dollars in the US alone [26]–[28]. Consequently, the Society of American Gastrointestinal and Endoscopic Surgeons (SAGES) has raised awareness of the safety of LCs through its SAGES safe cholecystectomy program [29], underscoring the fundamental importance of minimizing BDIs in LCs.

Critical View of Safety. Over the past few decades, surgeons have expended considerable effort in developing safe ways for identification of the cystic duct [8], of which the Critical View of Safety (CVS) technique is considered to be the most effective at target identification and hence is widely embraced in LC procedures [9]. CVS is said to be achieved if the following three criteria are met:

- C1: All fibrous and adipose tissues cleared within the hepatocystic triangle.
- C2: Separation of the lower one-third of the gallbladder from the cystic plate (liver bed).
- C3: Two and only two structures are seen to enter the gallbladder [10].

The promise of CVS spurred several studies on its effectiveness in the LCs, which provides strong evidence of the value of CVS as a means of unambiguously identifying biliary structures in LC. There is now widespread consensus on the high utility of CVS, so much so that all major guidelines for safe LC strongly recommend achieving CVS before the division of the cystic duct and artery [9].

Limitations of CVS. Despite many years since the introduction of CVS in the practice of LCs, the incidence rate of BDIs has not reduced. Given that more than a million LCs are performed annually, this amounts to several thousands of BDI cases every year. This persistence in the BDI rate has been attributed to: (i) weak adherence to the CVS criteria [19], (ii) inadequate understanding of CVS [17], [20], and (iii) insufficient confidence about CVS attainment, e.g., video analyses have shown a vast discrepancy between self-reported and formal CVS achievement [17].

Computer-Aided CVS. Techniques that can lead to a reduction of BDIs in LCs are of great value. As outlined above, there is strong evidence to suggest that better implementation and understanding of CVS holds great promise in reducing BDIs. Thus, developing computational techniques that automatically assess CVS during LCs and raise alerts have a great potential in reducing BDIs in LCs and thus, have a great clinical significance. This is the focus of our work.

Related Work. Automated identification of Critical View of Safety using deep learning methods has received increased interest in recent years. Mascagni et al. [4] first proposed using a segmentation model to identify hepatocystic anatomy and a classification model to predict binary CVS labels. Li et al. [1] proposed combining the segmentation model with domain knowledge and translating CVS conditions into algorithmic constraints to produce predictions for CVS conditions. Later, Murali et al. [5] proposed incorporating Graph Neural Networks (GNNs) with object detection models to explicitly encode semantic information to improve anatomy-driven reasoning.

III. **Method**

In this section, we describe our approach for automated assessment of the Critical View of Safety by prompting the Segment Anything Model (SAM) for semantic segmentation. We start by describing our high-level approach for CVS assessment on both frame and video levels. Then we will describe in detail our approach for assessing each CVS condition at the frame level, followed by a method for generating video-level CVS results.

High-level Approach. Our high-level approach consists of the following components.

Frame-level Assessment. For each frame in an LC video, we utilize two strategies to assess CVS conditions:

1. For C1 and C2, we formulate the problem into an object detection problem, and train a *Co-DETR* model to perform the task;
2. For C3, we first crop the images using a pre-trained *Co-DETR* model and then finetune the SAM model to perform semantic segmentation of cystic duct (CD) and cystic artery (CA) on the cropped images. Finally, we generate C3 results based on the segmentation results.

Video-level Assessment. We generate video-level CVS results by combining frame-level CVS condition results within each video with a thresholding mechanism. Video-level results are ultimately more relevant because we evaluate CVS conditions for each case instead of individual video frames.

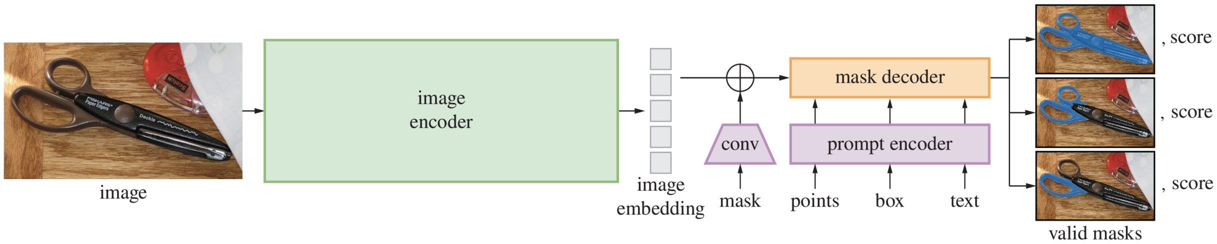

Fig. 2. The SAM model architecture.

We start by discussing the frame-level assessment of C3, which utilizes the SAM model.

Frame-level Assessment of C3 Criterion. C3 requires that two and only two structures be seen entering the gallbladder, namely the cystic duct (CD) and the cystic artery (CA). In our previous work [1], we relied on the *Segmenter* model to perform semantic segmentation of the two classes and use the segmentation results to assess the C3 criterion. However, segmentation models can sometimes generate noisy predictions in image parts that are far away from CD and CA, resulting in erroneous C3 results. Thus, we propose to utilize object detection as guidance for the segmentation model to improve both segmentation performance and C3 assessment. In particular, we follow the steps below to generate frame-level C3 predictions. See Fig 3.

1) Generate mask prompt, and detection boxes for cropping.
2) Finetune SAM model.
3) Assess the C3 criterion based on the SAM output.

We explain each of these below. We first give a brief introduction to the Segment Anything Model (SAM).

Segment Anything Model (SAM). The Segment Anything Model (SAM) [6] is a recent advancement in the field of computer vision and image segmentation, representing a significant leap in the ability of AI models to process and understand visual data. SAM is built upon the principles of Masked AutoEncoders (MAEs) [12] and is designed to segment a wide range of objects in images, regardless of whether they were included in the training dataset. Since its introduction, the SAM model has been adapted to perform a variety of downstream tasks and demonstrated impressive results [2], [7]. As shown in Figure 2, the SAM architecture consists of a Vision Transformer (ViT) [13] as the image encoder, a set of prompt encoders to deal with both sparse (points, boxes, text) and dense (masks) prompts, and a mask decoder to map the image embedding, prompt embeddings, and an output token to a mask.

Step 1. Generate mask prompt and detection boxes. The SAM model takes several forms of prompts to help with segmentation, i.e. points, boxes, text, and masks. Instead of prompting the SAM model with human input, we propose to automatically generate these prompts with pre-trained models. Through preliminary experiments, we found that mask prompts are the most effective at improving the segmentation capability of SAM on LC images. To generate mask prompts, we finetuned a pre-trained *Segmenter* model from [1] to generate "coarse" segmentation masks. In addition to the mask prompts, we finetuned a *Co-DETR* object detection model to detect cystic duct and cystic artery to assist in narrowing down the region where these two structures appear.

Step 2. Finetune SAM model. We adopt the finetuning strategy presented in [2] by freezing the image encoder in the SAM architecture and only finetuning the prompt encoder, mask decoder, and the LoRA [3] layers. We experimented with several variants and present two which improved segmentation results on CD and CA.

- *Mask prompt.* For this variant, we simply feed the predicted segmentation masks from the pre-trained *Segmenter* model as a dense prompt into the SAM model.
- *Mask prompt with cropping.* For this variant, we first crop the input image and the mask prompt with the predicted bounding boxes predicted by the pre-trained *Co-DETR* model. Then we use the cropped image as the input image and the cropped mask as prompt. If the detection model had no predictions, we revert to the mask prompt variant.

Step 3. Assess C3 criterion based on the SAM output. We generate frame-level C3 results based on the segmentation results of the SAM model. If one cluster of cystic artery and one cluster of cystic duct are segmented from the SAM model, it is considered satisfied. If the model produced more than one segmentation cluster for either of the two classes, then C3 is not considered satisfied.

Frame-level Assessment of C2 Criterion. C2 requires the separation of the lower one-third of the gallbladder from the liver, exposing the cystic plate. As it is difficult to obtain accurate pixel-level annotations of cystic plates, we formulate the assessment of C2 into an object detection problem. Specifically, we train the model to detect the region where the dissected cystic plate should be and classify the region as "1" if the cystic plate is sufficiently dissected according to C2 standard, "0" if the region does not contain sufficiently dissected cystic plate. We train the *Co-DETR* model in the same manner as in C3 assessment.

Frame-level assessment of C1 Criterion. C1 requires the hepatocystic triangle to be clear of fibrous and adipose tissues

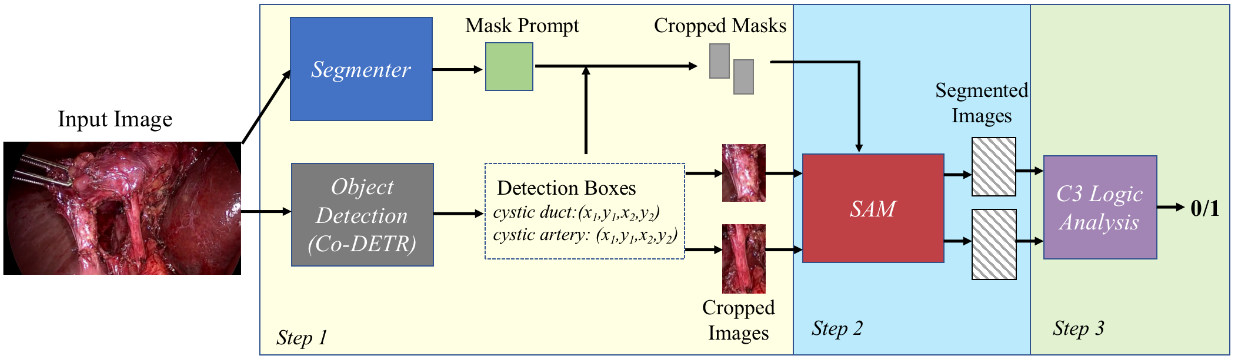

Fig. 3. Frame-level assessment of C3 pipeline.

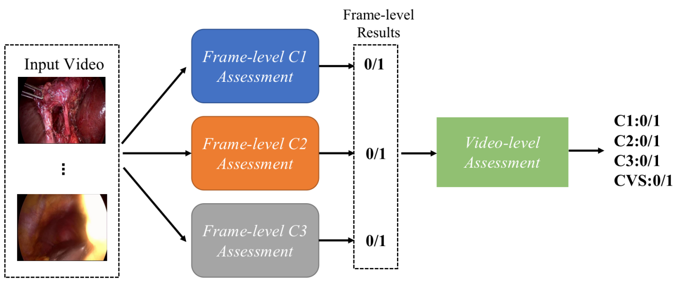

Fig. 4. The overall pipeline of our method.

to show an unimpeded view. Similar to C2, as pixel-level annotations for fat and fibrous tissues are difficult to obtain, we formulate the problem into an object detection problem. In particular, we train the *Co-DETR* model to detect the hepatocystic triangle and classify it as “1” when the triangle satisfies C1 criteria, and “0” if the triangle does not.

Video-level Assessment of CVS. Video-level predictions are generated based on predictions from individual frames via thresholding. For instance, if the number of frames identified as C1 positive is larger than a certain threshold t_1, then the video level C1 is considered satisfied. The same mechanism applies for C2 and C3. If all three conditions are satisfied at the video level, then CVS is considered satisfied for the video. Individual conditions don’t have to be satisfied simultaneously in a certain frame.

IV. Results

Dataset and Annotations. We combined *Cholec80* and *m2cai16-workflow* datasets and collected a total of 5750 frames from 107 videos. Among these videos, 80 videos are randomly selected as the training set and the remaining 27 videos as the validation set. The resulting split has 4162 frames for training and 1588 frames for validation. Each video frame was annotated with several key semantic classes, as well as bounding box annotations for C1 and C2. Specifically, the following classes are annotated with segmentation

TABLE I
SEGMENTATION RESULTS ON CD AND CA (CLASS IOU).

Method	CA	CD
Segmenter	0.1797	0.4263
SAM	0.2517	0.4137
SAM+MP	0.2687	0.4624
SAM+MP+crop	**0.3260**	**0.4657**

masks: *cystic artery (CA)*, *cystic duct (CD)*, *gallbladder*, liver, instrument. We annotate the bounding boxes for C1 and C2 in the following manner: For C1, we place the bounding box over the entire hepatocystic triangle if they are visible, and only label the box as 1 (C1 achieved) if 1) the entirety of the triangle should be visible, and 2) the triangle region should be sufficiently dissected, to the degree which we can certainly confirm and/or exclude the presence of >2 structures. For C2, we place the bounding box over the visible part of the cystic plate, or the region where the cystic plate should be after being dissected. We only label the box as 1 (C2 achieved) if a visible and significant portion of the cystic plate (roughly 1/3 according to CVS), is enough to clearly separate the gallbladder from the liver.

Segmentation Results on CD and CA. Table I shows the segmentation performance of our prompting strategies on CD and CA measured in class IoU. When finetuned on our dataset, the base SAM model outperforms *Segmenter* on the CA class but did not show improvement on the CD class. After implementing our `Mask prompt` and `Mask prompt with cropping`, we can see significant improvement in both classes.

Frame-level Results. We present the frame-level accuracy (Acc.), specificity (Spec.), sensitivity (Sens.), balanced accuracy (Bacc.) and precision (Prec.) for each CVS condition in Table II. For comparison, we trained two variants of the *LG-CVS* model on our dataset, with two different detector backbones. Our method performed better than the two variants of *LG-CVS* in terms of balanced accuracy (Bacc.) on both C1 and C3, and achieved competitive results on C2. It is worth noting that for specificity, which is important because a low amount of false positive is preferred in CVS assessment, our method consistently achieved over 0.9 performance.

Video-level Results. We present video-level results in Table III. For each of the methods, we do a parameter search and chose the thresholds that maximizes balanced accuracy (Bacc.). Compared to the two variants of the *LG-CVS* model, our method consistently performed better in Bacc. It is also worth noting that for video-level CVS results, both variants of the *LG-CVS* model ran into a break-down scenario where no positive samples were predicted, while our method still gave reasonable results.

V. Conclusion

In this paper, we proposed to address the problem of automated assessment of CVS by formulating it into object detection and semantic segmentation problems, and utilized state-of-the-art object detection method and the SAM model to achieve superior results on both frame and video level compared to previous state-of-the-art method *LG-CVS*. In our future work, we will further explore the potential of prompting large Vision-Language Models (VLMs) for surgical scene understanding and CVS assessment.

References

[1] Li Y, Gupta H, Ling H, Ramakrishnan IV, Prasanna P, Georgakis G, Sasson A. Automated Assessment of Critical View of Safety in Laparoscopic Cholecystectomy. In2023 IEEE 11th International Conference on Healthcare Informatics (ICHI) 2023 Jun 26 (pp. 330-337). IEEE.

[2] Zhang K, Liu D. Customized segment anything model for medical image segmentation. arXiv preprint arXiv:2304.13785. 2023 Apr 26.

[3] Hu EJ, Shen Y, Wallis P, Allen-Zhu Z, Li Y, Wang S, Wang L, Chen W. Lora: Low-rank adaptation of large language models. arXiv preprint arXiv:2106.09685. 2021 Jun 17.

[4] Mascagni P, Vardazaryan A, Alapatt D, Urade T, Emre T, Fiorillo C, Pessaux P, Mutter D, Marescaux J, Costamagna G, Dallemagne B. Artificial intelligence for surgical safety: automatic assessment of the critical view of safety in laparoscopic cholecystectomy using deep learning. Annals of surgery. 2022 May 15;275(5):955-61.

[5] Murali A, Alapatt D, Mascagni P, Vardazaryan A, Garcia A, Okamoto N, Mutter D, Padoy N. Latent graph representations for critical view of safety assessment. IEEE Transactions on Medical Imaging. 2023 Nov 16.

[6] Kirillov A, Mintun E, Ravi N, Mao H, Rolland C, Gustafson L, Xiao T, Whitehead S, Berg AC, Lo WY, Dollár P. Segment anything. arXiv preprint arXiv:2304.02643. 2023 Apr 5.

[7] Ma J, He Y, Li F, Han L, You C, Wang B. Segment anything in medical images. Nature Communications. 2024 Jan 22;15(1):654.

[8] Kaczynski J, Hilton J. A gallbladder with the "hidden cystic duct": A brief overview of various surgical techniques of the Calot's triangle dissection. Interventional Medicine and Applied Science. 2015 Mar 1;7(1):42-5.

[9] Vettoretto N, Saronni C, Harbi A, Balestra L, Taglietti L, Giovanetti M. Critical view of safety during laparoscopic cholecystectomy. JSLS: Journal of the Society of Laparoendoscopic Surgeons. 2011 Jul;15(3):322.

[10] Strasberg SM. A perspective on the critical view of safety in laparoscopic cholecystectomy. Annals of Laparoscopic and Endoscopic Surgery. 2017 May 10;2(5).

[11] Murali A, Alapatt D, Mascagni P, Vardazaryan A, Garcia A, Okamoto N, Mutter D, Padoy N. Encoding Surgical Videos as Latent Spatiotemporal Graphs for Object and Anatomy-Driven Reasoning. InInternational Conference on Medical Image Computing and Computer-Assisted Intervention 2023 Oct 1 (pp. 647-657). Cham: Springer Nature Switzerland.

[12] He K, Chen X, Xie S, Li Y, Dollár P, Girshick R. Masked autoencoders are scalable vision learners. InProceedings of the IEEE/CVF conference on computer vision and pattern recognition 2022 (pp. 16000-16009).

[13] Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, Uszkoreit J. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929. 2020 Oct 22.

[14] Buddingh KT, Weersma RK, Savenije RA, van Dam GM, Nieuwenhuijs VB. Lower rate of major bile duct injury and increased intraoperative management of common bile duct stones after implementation of routine intraoperative cholangiography. Journal of the American College of Surgeons. 2011 Aug 1;213(2):267-74.

[15] Halbert C, Pagkratis S, Yang J, Meng Z, Altieri MS, Parikh P, Pryor A, Talamini M, Telem DA. Beyond the learning curve: incidence of bile duct injuries following laparoscopic cholecystectomy normalize to open in the modern era. Surgical endoscopy. 2016 Jun;30:2239-43.

[16] Pucher PH, Brunt LM, Davies N, Linsk A, Munshi A, Rodriguez HA, Fingerhut A, Fanelli RD, Asbun H, Aggarwal R, SAGES Safe Cholecystectomy Task Force. Outcome trends and safety measures after 30 years of laparoscopic cholecystectomy: a systematic review and pooled data analysis. Surgical endoscopy. 2018 May;32:2175-83.

[17] Daly SC, Deziel DJ, Li X, Thaqi M, Millikan KW, Myers JA, Bonomo S, Luu MB. Current practices in biliary surgery: do we practice what we teach?. Surgical endoscopy. 2016 Aug;30:3345-50.

[18] "Calot's triangle." https://teachmeanatomy.info/abdomen/areas/calots-triangle, accessed: Jan 8,2023.

[19] Way LW, Stewart L, Gantert W, Liu K, Lee CM, Whang K, Hunter JG. Causes and prevention of laparoscopic bile duct injuries: analysis of 252 cases from a human factors and cognitive psychology perspective. Annals of surgery. 2003 Apr;237(4):460.

[20] Chen CB, Palazzo F, Doane SM, Winter JM, Lavu H, Chojnacki KA, Rosato EL, Yeo CJ, Pucci MJ. Increasing resident utilization and recognition of the critical view of safety during laparoscopic cholecystectomy: a pilot study from an academic medical center. Surgical endoscopy. 2017 Apr;31:1627-35.

[21] Gordon-Weeks A, Samarendra H, de Bono J, Soonawalla Z, Silva M. Surgeons opinions of legal practice in bile duct injury following cholecystectomy. HPB. 2017 Aug 1;19(8):721-6.

[22] Barbier L, Souche R, Slim K, Ah-Soune P. Long-term consequences of bile duct injury after cholecystectomy. Journal of visceral surgery. 2014 Sep 1;151(4):269-79.

[23] de'Angelis N, Catena F, Memeo R, Coccolini F, Martínez-Pérez A, Romeo OM, De Simone B, Di Saverio S, Brustia R, Rhaiem R, Piardi T. 2020 WSES guidelines for the detection and management of bile duct injury during cholecystectomy. World journal of emergency surgery. 2021 Dec;16(1):1-27.

[24] Törnqvist B, Strömberg C, Persson G, Nilsson M. Effect of intended intraoperative cholangiography and early detection of bile duct injury on survival after cholecystectomy: population based cohort study. Bmj. 2012 Oct 11;345.

[25] Alkhaffaf B, Decadt B. 15 years of litigation following laparoscopic cholecystectomy in England. Annals of surgery. 2010 Apr 1;251(4):682-5.

[26] Berci G, Hunter J, Morgenstern L, Arregui M, Brunt M, Carroll B, Edye M, Fermelia D, Ferzli G, Greene F, Petelin J. Laparoscopic cholecystectomy: first, do no harm; second, take care of bile duct stones. Surgical endoscopy. 2013 Apr;27:1051-4.

[27] Melton GB, Lillemoe KD, Cameron JL, Sauter PA, Coleman J, Yeo CJ. Major bile duct injuries associated with laparoscopic cholecystectomy: effect of surgical repair on quality of life. Annals of surgery. 2002 Jun;235(6):888.

[28] Strasberg SM, Brunt LM. The critical view of safety: why it is not the only method of ductal identification within the standard of care in laparoscopic cholecystectomy. Annals of Surgery. 2017 Mar 1;265(3):464-5.

[29] Brunt LM. SAGES Safe Cholecystectomy Program. Strategies for Minimizing Bile Duct Injuries: Adopting a Universal Culture of Safety in Cholecystectomy.

TABLE II
FRAME-LEVEL CVS RESULTS

	Approach	Acc.	Spec.	Sens.	Bacc.	Prec.
C1	**LG-CVS (Faster-RCNN)**	0.8533	0.9162	**0.1591**	0.5376	0.1468
	LG-CVS (Mask-RCNN)	0.8356	0.8984	0.1439	0.5211	0.1138
	Ours	**0.9193**	**0.9938**	0.0984	**0.5461**	**0.4229**
C2	**LG-CVS (Faster-RCNN)**	0.6870	**0.9566**	0.0704	0.5135	0.4146
	LG-CVS (Mask-RCNN)	0.6045	0.5620	**0.7019**	**0.6319**	0.4119
	Ours	**0.7059**	0.9031	0.2546	0.5789	**0.5347**
C3	**LG-CVS (Faster-RCNN)**	0.6650	0.9315	0.0984	0.5150	0.4032
	LG-CVS (Mask-RCNN)	0.6877	0.6130	**0.8465**	0.7297	0.5071
	Ours	**0.8671**	**0.9838**	0.5919	**0.7879**	**0.9395**

TABLE III
VIDEO-LEVEL CVS RESULTS

	Approach	Acc.	Spec.	Sens.	Bacc.	Prec.
C1	**LG-CVS (Faster-RCNN)**	0.4444	0.4737	**0.375**	0.4243	0.2308
	LG-CVS (Mask-RCNN)	0.5926	0.8421	0.0	0.4211	0.0
	Ours	**0.7407**	**0.8947**	**0.375**	**0.6348**	**0.6**
C2	**LG-CVS (Faster-RCN)**	0.4815	**1.0**	0.0	**0.5**	0.0
	LG-CVS (Mask-RCNN)	0.1852	0.4167	0.0	0.2083	0.0
	Ours	**0.5185**	0.3333	**0.6667**	**0.5**	**0.5556**
C3	**LG-CVS (Faster-RCN)**	0.3704	**1.0**	0.0	0.5	0.0
	LG-CVS (Mask-RCNN)	**0.7037**	0.3	**0.9412**	0.6206	0.0
	Ours	0.6667	**1.0**	0.5714	**0.7857**	**1.0**
CVS	**LG-CVS (Faster-RCN)**	**0.8889**	1.0	0.0	0.5	0.0
	LG-CVS (Mask-RCNN)	0.7778	**0.913**	0.0	0.4565	0.0
	Ours	0.8148	0.9091	**0.4**	**0.6545**	**0.5**