Automated Assessment of Critical View of Safety in Laparoscopic Cholecystectomy

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

Himanshu Gupta
Department of Computer Science
Stony Brook University
Stony Brook, USA
hgupta@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

Georgios Georgakis
Department of Surgery
Stony Brook University Hospital
Stony Brook, USA
georgios.georgakis@stonybrookmedicine.edu

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

***Abstract*—Cholecystectomy (gallbladder removal) is one of the most common procedures in the US, with more than 1.2M procedures annually. Compared with classical open cholecystectomy, laparoscopic cholecystectomy (LC) is associated with significantly shorter recovery period, and hence is the preferred method. However, LC is also associated with an increase in bile duct injuries (BDIs), resulting in significant morbidity and mortality. The primary cause of BDIs from LCs is misidentification of the cystic duct with the bile duct. Critical view of safety (CVS) is the most effective of safety protocols, which is said to be achieved during the surgery if certain criteria are met. However, due to suboptimal understanding and implementation of CVS, the BDI rates have remained stable over the last three decades.**

In this paper, we develop deep-learning techniques to automate the assessment of CVS in LCs. An innovative aspect of our research is on developing specialized learning techniques by incorporating domain knowledge to compensate for the limited training data available in practice. In particular, our CVS assessment process involves a fusion of two segmentation maps followed by an estimation of a certain region of interest based on anatomical structures close to the gallbladder, and then finally determination of each of the three CVS criteria via rule-based assessment of structural information. We achieved a gain of over 11.8% in mIoU on relevant classes with our two-stream semantic segmentation approach when compared to a single-model baseline, and 1.84% in mIoU with our proposed Sobel loss function when compared to a Transformer-based baseline model. For CVS criteria, we achieved up to 16% improvement and, for the overall CVS assessment, we achieved 5% improvement in balanced accuracy compared to DeepCVS under the same experiment settings.

***Index Terms*—Laparoscopic Cholecystectomy, Critical View of Safety, Deep Learning**

I. Introduction

Cholecystectomy is one of the most common surgical procedures in the US, done to remove an inflamed or infected gallbladder. Majority of cholecystectomy procedures are now done as laparoscopic cholecystectomy (LC), as they are associated with shorter recovery times. However, LCs are also associated with an increased number of bile duct injuries (BDIs), which occur due to limited field of vision. BDIs resulting from LCs may lead to serious complications which can even endanger the patient's life and safety [1], [2], while driving up the medical litigation [3] and healthcare costs to over a billion dollars in the US alone [4]. A safety protocol, termed as critical view of safety (CVS), has been developed and widely embraced over the years, with the goal of minimizing misidentification of ducts and thus reduce incidence of BDIs. In spite of many evidences of the effectiveness of CVS protocol, the incidence of BDIs has not decreased over the past decades; the main reason for this stems from the insufficient implementation and understanding of CVS criteria by the surgeons [5]. Thus, automation of the CVS attainment in LC surgeries can potentially reduce incidence of BDIs in LCs.

Vision. Our long-term vision is to develop a AI-driven surgical aid that will prevent BDIs by a combination of real-time CVS assessment during LC, enforcement of related safety processes (e.g., identifying and guiding surgeons to bailout strategies [6]), and training of surgeons via video reviews to improve their understanding of CVS and LC surgeries. As a step towards the above vision, in this paper, we focus

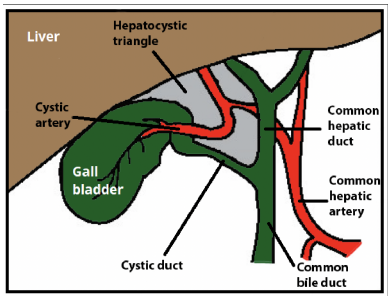

Fig. 1. Anatomy of hepatocystic triangle. [8]

on developing a technique to assess CVS based on its three criteria; such a technique can be used to raise alerts in real-time (i.e., while LC surgery is in progress) if an attempt is made to clamp or cut any structure before a true CVS has been attained and thus, prevent BDIs. The key challenge in CVS assessment from learning techniques is the lack of sufficient training data (at most a few hundred LC surgery videos) as well as the intrinsic difficulties in CVS assessment, such as the cluttered texture and occlusion among organs. Our approach addresses these challenges by proposing a fusion approach followed by incorporation of clinical domain knowledge. In particular, our approach involves estimating a region of interest based on anatomical structures around the gallbladder, and rule-based assessment of CVS criteria. We demonstrate that such an approach has a great potential in accurate detection of CVS by showing an advantage in performance on both individual CVS criteria and overall CVS classification when compared to CNN-based DeepCVS [7] as baseline.

II. Background

In this section, we provide general background and related work.

Laparoscopic Cholecystectomy (LC). Gallbladder is a small organ underneath the liver that concentrates and stores bile fluid. Inflammation and infection of the gallbladder may necessitate surgical removal of the gallbladder, which is done via LC, a minimally invasive procedure associated with quick recovery time. LC, performed through four small incisions, uses a camera and surgical tools to remove the gallbladder. Removal of gallbladder essentially entails exposing (by removing the fat and fibrous tissues) and cutting the only two structures that connect it to the body: the cystic duct (CD) and the cystic artery.

BDI Risks of LCs. The most feared adverse event of LC is bile duct injury (BDI), which occurs in thousands of cases in the US annually [2]. BDIs largely result from misidentification of the common bile duct as the cystic duct [9], due to the increased complexity of LC procedures and limited field of vision. BDIs due to LCs may lead to serious complications and even endanger the patient's life and safety [1], [2]. Overall, BDIs frequently result in a 3-fold increase in the 1-year mortality rate [10], while driving up the medical litigation [3] and healthcare costs to over a billion dollars in the US alone [4], [11], [12].

The Critical View of Safety (CVS) Technique. Over the past few decades, surgeons have expended considerable effort in developing safe ways for identification of the cystic duct [13], 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 [6], [14]. CVS is said to be achieved if the following three criteria are met:[1]

- C1: All fibrous and adipose tissues cleared within the hepatocystic triangle (see Fig. 1).
- 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 [15].

Impact and Limitation of CVS. The promise of CVS spurred several studies [16], [17] on its effectiveness in the LC procedure, which provide strong evidence of the value of CVS as a means of unambiguously identifying biliary structures in LC. However, despite the evidence of the efficacy of CVS in reducing mis-identification of CD, BDI rates over the last 3 decades have remained stable at 0.36%–1.5% [10]. The primary reasons for this status quo are: insufficient or inadequate implementation of CVS [18], and weak understanding of CVS among many surgeons [5], [19]. Sometimes, overconfidence (partly due to the low incidence of BDIs) with LC also plays a part [5], [17], [20], [21]. Thus, automated assessment of CVS criteria has the potential to reduce BDIs, especially with the advances and contributions of computer vision in medical image analysis over the recent years.

Related Work. There have been two very-recent works on assessment of CVS. In particular, Mascagni et al. [7] utilizes the semantic segmentation results of DeepLabV3+ [22] and predicts binary labels of CVS criteria and overall CVS achievement from a compactly-designed CNN. More recently, Murali et al. [23] proposed incorporating graph neural networks (GNNs) to encode the latent scene graph in LC video frames, and shows improved performance over DeepCVS. However, these methods do not involve domain knowledge on CVS criteria and thus their results could not be easily analyzed or explained. In another related work, Madani et al. [24] proposed using CNN-based semantic segmentation methods to identify safe and dangerous zones of dissections, which could serve as an important intermediary stage for CVS assessment.

III. Methodology

Key Challenges in Automated CVS Assessment. Since the BDI incidence rate in LCs is extremely low (0.36% to 1.5%) [10], a CVS detection technique must necessarily have

[1]CVS is a reworking of the open cholecystectomy protocol wherein the gallbladder is detached from the cystic plate (liver bed) so that it is attached to the body by only the two cystic structures which can then be clipped. In laparoscopic surgery, as complete separation of the gallbladder from the cystic plate makes clipping of the structures difficult, we require that only the lower part of the gallbladder be separated [9].

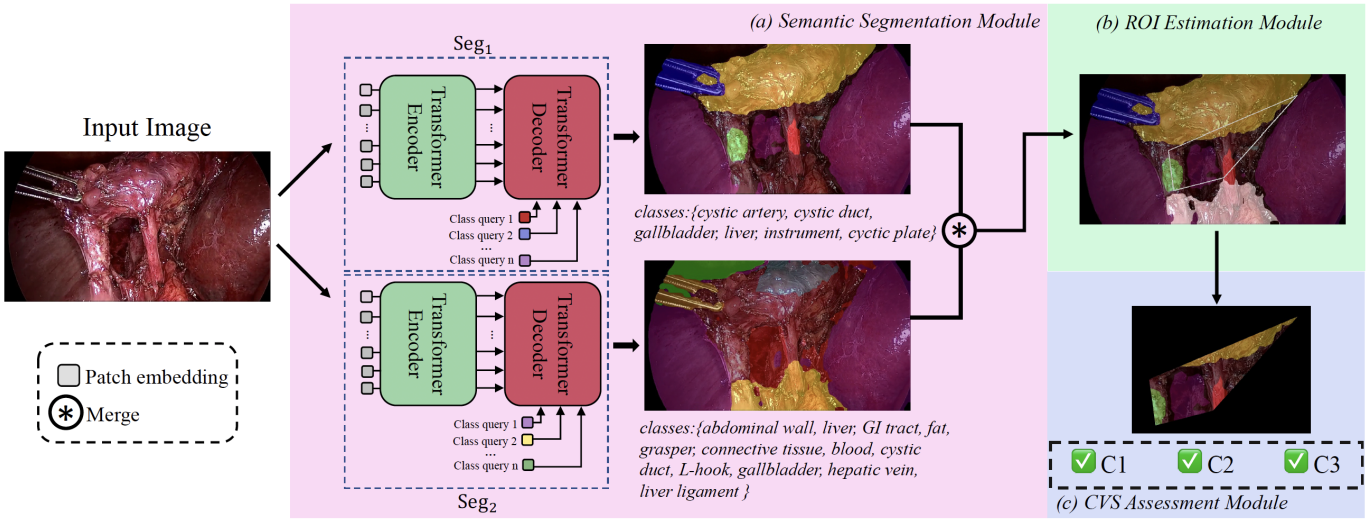

Fig. 2. Overall pipeline of our approach. The input frame is first segmented by two Transformer-based models. The segmentation maps are then merged for ROI estimation. Finally, CVS conditions are evaluated based on ROI and segmentation maps.

very high accuracy (e.g., 90% or more) to lower this BDI rate even further. Due to limited training data available,[2] such a high accuracy is infeasible by direct application of machine-learning techniques, as seen in some of the prior works. One approach to achieve such accuracy would be to integrate extensive clinical/domain knowledge, as incorporating such knowledge has been shown to boost the accuracy of ML algorithms (e.g., [25]–[27]). However, leveraging clinical domain knowledge in ML models can be quite challenging.

Method Pipeline and Key Contributions. Our approach tackles the aforementioned challenges by incorporating domain knowledge with limited training data. In particular, our approach's pipeline is as follows (see Fig. 2). First, to address the imbalance of classes in available datasets, we segment each image frame by using two Transformer-based models trained on separate semantic segmentation datasets; relevant classes from these two segmentation maps are then appropriately fused. Then, we use structural anatomic knowledge of the gallbladder and surrounding structures to estimate the region of interest (ROI), which is used to efficiently assess the CVS conditions. Finally, we assess each of the three CVS conditions based on their structural definitions, and then the overall CVS as a conjunction of the three CVS conditions. Overall, our main contributions include:

1. Introducing a *two-stream approach for semantic segmentation* to address the issue of class imbalance.
2. Proposing a novel *Sobel loss function* to reduce artifacts and over-segmentation around edges.
3. *Integration of clinical domain knowledge:* Developing a rule-based approach for estimating ROIs and assessing CVS conditions in LC videos based on domain knowledge.

[2]One can realistically expect to curate a few hundred or at most a few thousand LC surgical videos; by contrast, highly accurate ML models tend to use millions of training samples.

A. Semantic Segmentation

Two-stream Segmentation and Fusion. For segmentation of LC frames, we wish to use the publicly available *CholecSeg8K* dataset which includes 8,080 frames annotated with related classes. However, the *CholecSeg8K* dataset is missing two important classes, viz., *cystic plate* and *cystic artery*, and has low number of pixels in *cystic duct* class; all of these three classes are crucial to our approach (in particular, in estimation of the region of interest, discussed in the next section). To compensate for the above shortcomings, we created the *CholecSeg170* dataset which includes annotations for cystic plate and cystic artery, and much higher proportion of *cystic duct* pixels. We believe that training two separate segmentation models over the above two datasets separately should yield better performance, especially on the important classes *cystic duct* and *cystic artery*, than training a single segmentation model over the union of the above datasets; our intuition is confirmed in our evaluation results (see Section. IV-B).

Thus, the first segmentation model $\mathbf{Seg_1}$ is trained on the *CholecSeg170* dataset, while the second model $\mathbf{Seg_2}$ is trained

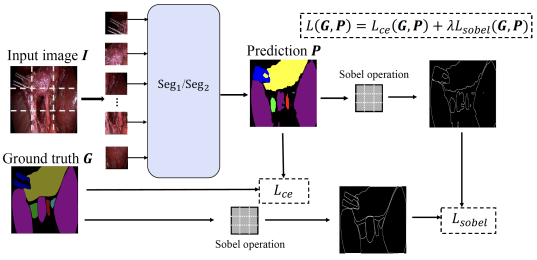

Fig. 3. Our proposed Sobel loss. It reduces artifacts and over-segmentation around edges by penalizing the difference between edge maps derived from segmentation maps.

on the *CholecSeg8K* dataset. We use $\mathbf{Seg_1}$ for segmentation of 6 classes: *cystic artery, cystic duct, gallbladder, liver, instrument, cystic plate*, while $\mathbf{Seg_2}$ is used for segmentation of only the *fat* class. For an input image $\mathbf{I}$, let $P_1 = \mathbf{Seg_1}(\mathbf{I})$, $P_2 = \mathbf{Seg_2}(\mathbf{I})$. Then, the merged segmentation map is constructed by $P_{merged} = P_1 \oplus \mathbf{Fat}(P_2)$, where $\mathbf{Fat}$ denotes creating a mask of the *fat* class.

Sobel Loss Function. We use the Transformer-based Segmenter [28] model as the baseline for our semantic segmentation method. When evaluating the segmentation results, we observed that the edges between different anatomical classes are not clearly separated, causing artifacts and over-segmentation (see Section. IV-B). To address this issue, we propose adding an edge-based constraint to the loss function. Specifically, we use the Sobel operator to generate class-agnostic edge information from the segmentation maps, and then apply Smooth *L1* Loss [29] between the ground truth and predicted edges.

The Sobel operator uses of two 3×3 convolutional filters to calculate the approximations of the derivatives both vertically and horizontally. Given input image $\mathbf{I}$, we calculate the gradient of the image *Sobel*($\mathbf{I}$) as: *Sobel*($\mathbf{I}$) $= \sqrt{G_x^2 + G_y^2}$, where

$$G_x = \begin{bmatrix} 2 & 0 & -2 \\ 4 & 0 & -4 \\ 2 & 0 & -2 \end{bmatrix} * \mathbf{I}, \qquad G_y = \begin{bmatrix} 2 & 4 & 2 \\ 0 & 0 & 0 \\ -2 & -4 & -2 \end{bmatrix} * \mathbf{I}, \quad (1)$$

G_x, G_y are the two images containing horizontal and vertical derivatives respectively, and $*$ denotes the 2-D convolution operation. Given ground truth segmentation map G and predicted segmentation map P, we define our Sobel loss function as:

$$L_{Sobel}(G, P) = smooth_{L_1}(Sobel(G) - Sobel(P)) \quad (2)$$

where $smooth_{L_1}$ is the Smooth *L1* Loss. Finally, our training objective is defined as

$$L(G, P) = L_{ce}(G, P) + \lambda L_{Sobel}(G, P) \quad (3)$$

where L_{ce} is the cross-entropy loss, and λ is a hyperparameter. The segmentation model pipeline is shown in Fig. 3.

B. Region of Interest (RoI) Estimation

In LC procedures, the assessment of CVS is mainly based on a specific region where the surgeon dissects tissue to expose cystic duct, cystic artery, and the cystic plate, and thereby creating the CVS. In LC terminology, this region is referred to as the *hepatocystic triangle*. In most surgeries, the triangle is never fully visible since the surgeons usually only dissect to the point where cystic duct and cystic artery are sufficiently exposed while the common hepatic duct and common bile duct remain hidden. Thus, in the LC surgery frames, we observe that only a part (in shape of a quadrilateral) of the hepatoycstic triangle is visible. Hence, our region of interest (ROI) is of a quadrilateral shape with four sides.

The **ROI quadrilateral** (see Fig. 4) is defined by anatomical structures around the gallbladder observed in the LC surgery videos. Thus, we develop a clinically-motivated rule-based method to determine the ROI, rather than applying standard learning techniques as is typically done. In particular, the ROI quadrilateral is formed by four points in an LC surgery image: (A) Cystic duct's end that is connected to the gallbladder; (B) Other end of the (visible) cystic duct; (C) Intersection point between the liver edge and a line drawn from point B to the outline of the largest cluster of *fat* class; (D) the point connecting the gallbladder to the liver. Note that the determination of point (C) is done to exclude the main cluster of fat tissue from the ROI—we use the condition of such a quadrilateral being devoid of any fat tissue as the sub-condition for the C1 criteria of CVS.

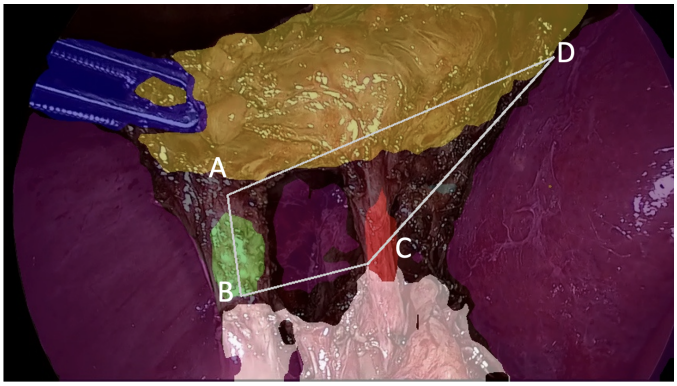

Fig. 4. ROI Quadrilateral.

In a segmented frame, we estimate the above defined four points as follows. First, we estimate points A and B as follows (see Fig. 5) . We perform principal component analysis (PCA) on the main cluster $\mathbf{C_{duct}}$ of *cystic duct* pixels, as detected by the first segmentation model $\mathbf{Seg_1}$. Let the two primary components obtained from PCA be $\mathbf{X_1}$ and $\mathbf{X_2}$, with $\mathbf{X_1}$ being the one with a higher angle (almost perpendicular) to the gallbladder edge. Next, we create a line segment by starting from the centroid of the cluster $\mathbf{C_{duct}}$ and extending in both directions along $\mathbf{X_1}$ till the outline of the cluster is reached; let the endpoints of this line segment be p_1 and p_2, with p_1 being the point closer to the gallbladder. We define A to the point between p_1 and its nearest neighbour on the gallbladder edge, and B as p_2. To estimate the point C, we start with the line connecting A and B, and rotate it clockwise till it intersects with the main cluster of *fat* tissue; the intersection point is assigned to be point C. Finally, we estimate the point D as follows. Since the segmentation maps usually do not yield a unique point where the gallbladder and liver edges intersect, we choose a pair of points, one from each edge, that has the minimal Euclidean distance between them; for this, we use a modified KD-Tree Nearest Neighbour algorithm [30]. The point D is defined as the midpoint between these two points.

C. CVS Assessment

Given the semantic segmentation maps and the ROI quadrilateral in an image frame, we develop a rule-based method to determine attainment of each of the three CVS criteria and thus the CVS. Recall the three CVS conditions from Section. II. For

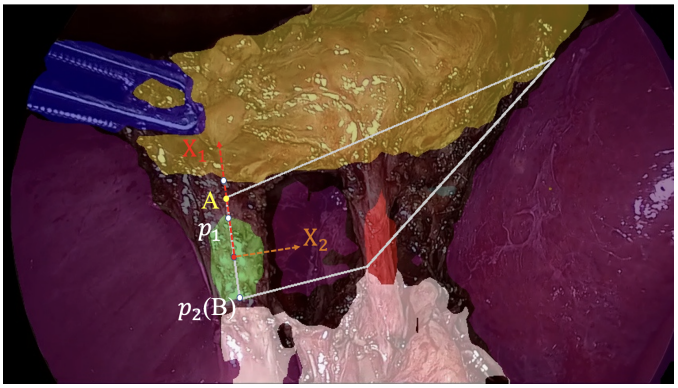

Fig. 5. Estimation of point **A** and **B** in our ROI estimation method. We first identify the two main components of the cystic duct $\mathbf{X_1}$,$\mathbf{X_2}$ cluster using PCA. Then we extend $\mathbf{X_1}$ in both directions from the centroid of the cluster to find p_1 and p_2. Finally, we define the mid-point between p_1 and its nearest neighbour on the gallbladder edge as **A**, and p_2 as **B**.

C1, which is to check for fat or fibrous tissue in the hepatocystic triangle (and thus, the ROI quadrilateral), we determine attainment of C1 condition based on following two conditions: (a) No *fat* pixels in the ROI; (b) The size of the cluster of *liver* pixels in the ROI is more than a certain threshold T_{liver}. Note that the *fat* and *liver* classes are determined by $\mathbf{Seg_2}$ and $\mathbf{Seg_1}$ segmentation maps respectively. If both the above conditions are satisfied, we consider C1 condition to be satisfied. For **C2**, if the size of the cluster of *cystic plate* pixels in the ROI surpasses a certain threshold T_{cp}, it is considered satisfied. For **C3**, if exactly one cluster of *cystic duct* pixels and one cluster of *cystic artery* pixels are detected by $\mathbf{Seg_1}$ in the ROI, it is considered satisfied. We empirically set $T_{liver} = 100$ and $T_{cp} = 100$ to eliminate some of the noisy predictions.

IV. Results

In this section, we introduce the datasets we used for development and evaluation of our techniques and the results of our method.

A. Datasets

The combined *Cholec80* [31]and *m2cai16-workflow* [32] dataset consists of 117 videos after excluding duplicate cases [24]. We use the 17 videos from the *CholecSeg8K* dataset as the development set and the remaining 100 as the evaluation set. The development set consists of two separate semantic segmentation datasets, namely *CholecSeg8K* and *CholecSeg170*. The evaluation set, named *CVS6K*, consists of 6,000 frames with only binary CVS annotations.

CholecSeg8K. The *CholecSeg8K* dataset is a publicly available semantic segmentation dataset based on the *Cholec80* dataset. In total, 8,080 frames were collected from 17 videos in the *Cholec80* dataset, and 13 different semantic classes (including background) were annotated. Most relevant classes in LC are annotated, such as *liver, fat, gallbladder* and *cystic duct*. However, *CholecSeg8K* is highly unbalanced in class distribution, and some crucial classes for assessing CVS, such as *cystic plate* and *cystic artery*, are absent from the dataset.

CholecSeg170. To address the limitations of *CholecSeg8K*, we collected 170 frames from the same 17 videos to form a separate semantic segmentation dataset, which we call the *CholecSeg170* dataset. For each video, 10 frames are manually selected close to the *ClippingCutting* stage as defined in *Cholec80*, where most anatomical structures necessary for evaluating CVS are visible. The selected frames are annotated with the following 7 semantic classes: {*cystic artery, cystic duct, gallbladder, instrument, liver, cystic plate, background*}. Additionally, ground truth CVS conditions are labeled for each frame.The 170 frames are divided into 140 frames for training and 30 frames for validation.

CVS6K. The 100 videos which are not included in the semantic segmentation datasets are used to construct the CVS evaluation set. We first sample a one minute clip at 1fps from each video, all of which near the *ClippingCutting* stage of the videos, when CVS conditions can be clearly evaluated in most frames. For each frame, we assign three binary labels corresponding to the three criteria of CVS as suggested by SAGES [6]. If and only if all three criteria are satisfied in a frame do we consider CVS achieved in that frame. The proportions of positive examples on the dataset is shown in Fig. 6. All annotations on the CVS evaluation dataset are verified independently by two experienced oncology surgeons (co-authors).

B. Semantic Segmentation

We start by evaluating the effectiveness of our two-stream segmentation approach by computing the IoU metric on each relevant class in TABLE I. We observe that the two-stream approach improves the IoU by 11.85% on average, and the improvements are especially significant on low-frequency classes like *cystic duct* (18.55%), *cystic artery* (44.84%), and *cystic plate* (14.84%). We also assess the enhancement resulting from the proposed Sobel loss on the validation set of *CholecSeg170* in TABLE II. We see that the Sobel loss function resulted in 1.84% improvement in mIoU and 1.8% improvement in Dice

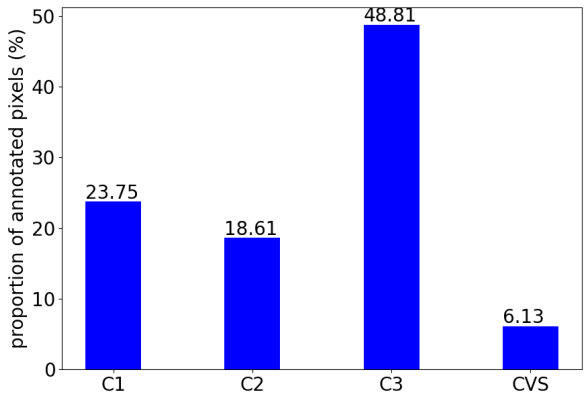

Fig. 6. Proportion of positive examples in *CVS6K*.

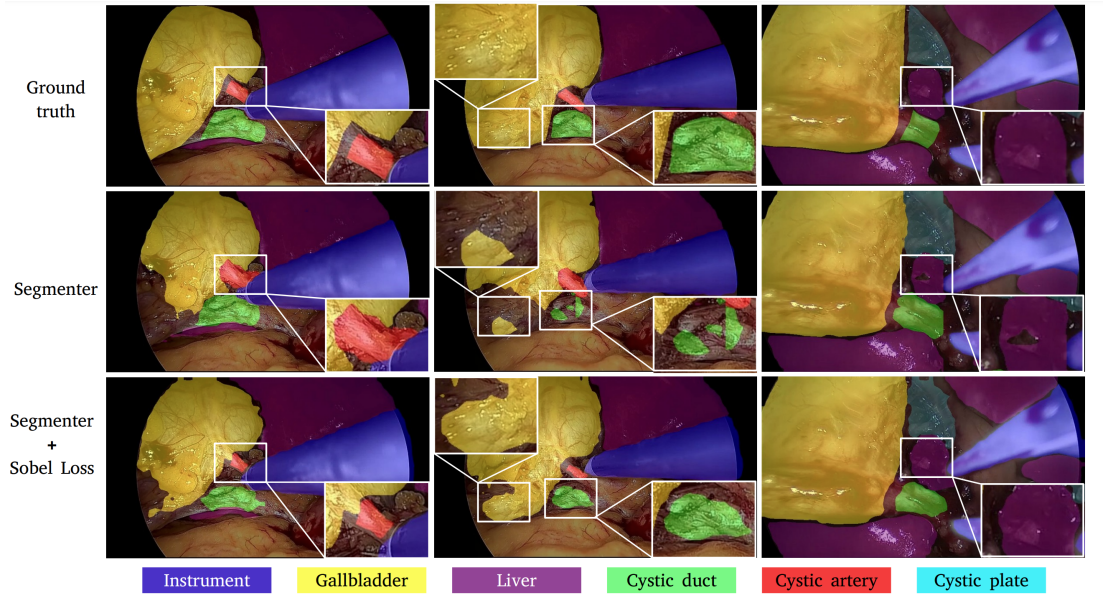

Fig. 7. Qualitative results. Our proposed Sobel Loss reduced over-segmentation of cystic artery in column 1, and improved on the artifacts/fragmented segmentations of gallbladder, cystic duct, and liver (columns 2, 3).

TABLE I
COMPARISON OF ONE VS. TWO-MODEL SEGMENTATION APPROACHES, IN TERMS OF IOU.

Approach	Gallbladder	Liver	Cystic Duct	Cystic artery	Cystic plate	Instrument
Single model	0.8964	**0.9244**	0.4978	0.0	0.4229	**0.8989**
Two-stream	**0.9139**	0.8913	**0.6833**	**0.4484**	**0.5713**	0.8433

TABLE II
COMPARISON OF SOBEL LOSS BASED SEGMENTATION AND THE BASELINE METHOD.

Model/Metric	mIoU	Acc.	Dice
Baseline	0.7270	**0.9372**	0.8247
Baseline+Sobel loss	**0.7454**	0.9323	**0.8427**

score compared to Segmenter baseline. We used $\lambda = 1$ when deploying Sobel loss.

We also evaluated **qualitative results** in Fig. 7. We see that our proposed Sobel loss penalizes noisy predictions around edges, leading to more inter-class separation and thereby creating more defined edges on anatomical structures and organs. Additionally, it also reduces noisy patches often observed from the baseline model.

C. CVS Conditions and CVS Assessment

We present the accuracy (Acc.), balanced accuracy (Bacc.), Positive Predictive Value (PPV) and Negative Predictive Value (NPP) on the independent *CVS6K* dataset in TABLE III. For the baseline approach, we re-implemented DeepCVS according to the descriptions in [7], with slight modification to fit our experiment settings, and for the purpose of fair comparison. In particular, we trained two separate DeepLabV3+ semantic segmentation models on *CholecSeg170* and *CholecSeg8K* datasets. The segmentation maps are fused the same way as described in Section.III-A. The CNN for classification of CVS conditions are implemented according to [7] except for the first layer. As may be observed in TABLE III, our rule-based method significantly outperforms the baseline model on both independent CVS criteria and overall CVS assessment, and shows more consistent performance among different CVS conditions.

TABLE III
RESULTS OF CVS ASSESSMENT COMPARED TO DEEPCVS.

	C1		C2		C3		CVS	
DeepCVS	Acc.	0.72	Acc.	0.39	Acc.	0.54	Acc.	0.92
	Bacc.	0.48	Bacc.	0.49	Bacc.	0.53	Bacc.	0.49
	PPV	0.14	PPV	0.18	PPV	0.53	PPV	NaN[3]
	NPV	0.75	NPV	0.80	NPV	0.54	NPV	0.93
Ours	Acc.	**0.76**	Acc.	**0.79**	Acc.	**0.69**	Acc.	0.92
	Bacc.	**0.57**	Bacc.	**0.65**	Bacc.	**0.69**	Bacc.	**0.54**
	PPV	**0.49**	PPV.	**0.43**	PPV	**0.72**	PPV	**0.23**
	NPV	**0.79**	NPV	**0.86**	NPV	**0.67**	NPV	**0.94**

V. CONCLUSION

In this work, we have addressed a critical unmet clinical need, viz, assessing CVS in LC procedures to help minimize incidence of BDIs. We developed a 3-step pipeline, which addresses the issues of class imbalance and artifacts in semantic segmentation, while also incorporates domain

[3]PPV is undefined in this case since all frames are predicted as negative in CVS.

knowledge for more accurate CVS assessment. The results show great promise in future applications in computer-assisted LC procedures. However, one limitation of our approach is that it heavily relies on the quality of the segmentation results and does not include a reasonable fail-safe mechanism when segmentation models produce undesirable results. To address this challenge, we aim to develop methods that take advantage of segmentation-failure detection techniques in our future work.

ACKNOWLEDGMENT

We would like to acknowledge Twinanda et al. [31] and Hong et al. [33] for making their datasets publicly available to the research community.

Research reported in this publication was supported by National Science Foundation (NSF) under award numbers FET-2106447, CNS-2128187, 2153056, 2125147, 2113485, 2006655 and National Institutes of Health (NIH) under award numbers R01EY030085, R01HD097188, 1R21CA258493-01A1. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NSF and the NIH.

REFERENCES

[1] L. Barbier, R. Souche, K. Slim, and P. Ah-Soune, "Long-term consequences of bile duct injury after cholecystectomy," *Journal of visceral surgery*, vol. 151, no. 4, pp. 269–279, 2014.

[2] N. de'Angelis, F. Catena, R. Memeo, F. Coccolini, A. Martínez-Pérez, O. M. Romeo, B. De Simone, S. Di Saverio, R. Brustia, R. Rhaiem *et al.*, "2020 wses guidelines for the detection and management of bile duct injury during cholecystectomy," *World Journal of Emergency Surgery*, vol. 16, no. 1, pp. 1–27, 2021.

[3] B. Alkhaffaf and B. Decadt, "15 years of litigation following laparoscopic cholecystectomy in england," *Annals of surgery*, vol. 251, no. 4, pp. 682–685, 2010.

[4] G. Berci, J. Hunter, L. Morgenstern, M. Arregui, M. Brunt, B. Carroll, M. Edye, D. Fermelia, G. Ferzli, F. Greene *et al.*, "Laparoscopic cholecystectomy: first, do no harm; second, take care of bile duct stones," pp. 1051–1054, 2013.

[5] S. C. Daly, D. J. Deziel, X. Li, M. Thaqi, K. W. Millikan, J. A. Myers, S. Bonomo, and M. B. Luu, "Current practices in biliary surgery: Do we practice what we teach?" *Surgical endoscopy*, vol. 30, no. 8, pp. 3345–3350, 2016.

[6] "The SAGES safe cholecystectomy program - strategies for minimizing bile duct injuries," https://www.sages.org/safe-cholecystectomy-program/, published: Oct 15,2021.

[7] P. Mascagni, A. Vardazaryan, D. Alapatt, T. Urade, T. Emre, C. Fiorillo, P. Pessaux, D. Mutter, J. Marescaux, G. Costamagna *et al.*, "Artificial intelligence for surgical safety: automatic assessment of the critical view of safety in laparoscopic cholecystectomy using deep learning," *Annals of surgery*, vol. 275, no. 5, pp. 955–961, 2022.

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

[9] S. M. Strasberg and M. L. Brunt, "Rationale and use of the critical view of safety in laparoscopic cholecystectomy," *Journal of the American College of Surgeons*, vol. 211, no. 1, pp. 132–138, 2010.

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

[11] G. B. Melton, K. D. Lillemoe, J. L. Cameron, P. A. Sauter, J. Coleman, and C. J. Yeo, "Major bile duct injuries associated with laparoscopic cholecystectomy: effect of surgical repair on quality of life," *Annals of surgery*, vol. 235, no. 6, p. 888, 2002.

[12] S. M. Strasberg and L. M. Brunt, "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*, vol. 265, no. 3, pp. 464–465, 2017.

[13] J. Kaczynski and J. Hilton, "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*, vol. 7, no. 1, pp. 42–45, 2015.

[14] N. Vettoretto, C. Saronni, A. Harbi, L. Balestra, L. Taglietti, and M. Giovanetti, "Critical view of safety during laparoscopic cholecystectomy," *JSLS: Journal of the Society of Laparoendoscopic Surgeons*, vol. 15, no. 3, p. 322, 2011.

[15] S. M. Strasberg, "A perspective on the critical view of safety in laparoscopic cholecystectomy," *Annals of Laparoscopic and Endoscopic Surgery*, vol. 2, 2017.

[16] S. Yegiyants and J. C. Collins, "Operative strategy can reduce the incidence of major bile duct injury in laparoscopic cholecystectomy," *The American Surgeon*, vol. 74, no. 10, pp. 985–987, 2008.

[17] M. Nijssen, J. Schreinemakers, Z. Meyer, G. Van Der Schelling, R. Crolla, and A. Rijken, "Complications after laparoscopic cholecystectomy: a video evaluation study of whether the critical view of safety was reached," *World journal of surgery*, vol. 39, no. 7, pp. 1798–1803, 2015.

[18] L. W. Way, L. Stewart, W. Gantert, K. Liu, C. M. Lee, K. Whang, and J. G. Hunter, "Causes and prevention of laparoscopic bile duct injuries: analysis of 252 cases from a human factors and cognitive psychology perspective," *Annals of surgery*, vol. 237, no. 4, p. 460, 2003.

[19] C. B. Chen, F. Palazzo, S. M. Doane, J. M. Winter, H. Lavu, K. A. Chojnacki, E. L. Rosato, C. J. Yeo, and M. J. Pucci, "Increasing resident utilization and recognition of the critical view of safety during laparoscopic cholecystectomy: a pilot study from an academic medical center," *Surgical endoscopy*, vol. 31, no. 4, pp. 1627–1635, 2017.

[20] A. Rawlings, S. E. Hodgett, B. D. Matthews, S. M. Strasberg, M. Quasebarth, and L. M. Brunt, "Single-incision laparoscopic cholecystectomy: initial experience with critical view of safety dissection and routine intraoperative cholangiography," *Journal of the American College of Surgeons*, vol. 211, no. 1, pp. 1–7, 2010.

[21] D. Stefanidis, N. Chintalapudi, B. Anderson-Montoya, B. Oommen, D. Tobben, and M. Pimentel, "How often do surgeons obtain the critical view of safety during laparoscopic cholecystectomy?" *Surgical endoscopy*, vol. 31, no. 1, pp. 142–146, 2017.

[22] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 801–818.

[23] A. Murali, D. Alapatt, P. Mascagni, A. Vardazaryan, A. Garcia, N. Okamoto, D. Mutter, and N. Padoy, "Latent graph representations for critical view of safety assessment," *arXiv preprint arXiv:2212.04155*, 2022.

[24] A. Madani, B. Namazi, M. S. Altieri, D. A. Hashimoto, A. M. Rivera, P. H. Pucher, A. Navarrete-Welton, G. Sankaranarayanan, L. M. Brunt, A. Okrainec *et al.*, "Artificial intelligence for intraoperative guidance: using semantic segmentation to identify surgical anatomy during laparoscopic cholecystectomy," *Annals of surgery*, 2022.

[25] X. Xie, J. Niu, X. Liu, Z. Chen, S. Tang, and S. Yu, "A survey on incorporating domain knowledge into deep learning for medical image analysis," *Medical Image Analysis*, vol. 69, p. 101985, 2021.

[26] C. Pape, A. Matskevych, A. Wolny, J. Hennies, G. Mizzon, M. Louveaux, J. Musser, A. Maizel, D. Arendt, and A. Kreshuk, "Leveraging domain knowledge to improve microscopy image segmentation with lifted multicuts," *Frontiers in Computer Science*, p. 6, 2019.

[27] A. Konwer, X. Xu, J. Bae, C. Chen, and P. Prasanna, "Temporal context matters: Enhancing single image prediction with disease progression representations," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 18 824–18 835.

[28] R. Strudel, R. Garcia, I. Laptev, and C. Schmid, "Segmenter: Transformer for semantic segmentation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 7262–7272.

[29] R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1440–1448.

[30] S. Maneewongvatana and D. M. Mount, "Analysis of approximate nearest neighbor searching with clustered point sets," *arXiv preprint cs/9901013*, 1999.

[31] A. P. Twinanda, S. Shehata, D. Mutter, J. Marescaux, M. De Mathelin, and N. Padoy, "Endonet: a deep architecture for recognition tasks on laparoscopic videos," *IEEE transactions on medical imaging*, vol. 36, no. 1, pp. 86–97, 2016.

[32] R. Stauder, D. Ostler, M. Kranzfelder, S. Koller, H. Feußner, and N. Navab, "The tum lapchole dataset for the m2cai 2016 workflow challenge," *arXiv preprint arXiv:1610.09278*, 2016.

[33] W.-Y. Hong, C.-L. Kao, Y.-H. Kuo, J.-R. Wang, W.-L. Chang, and C.-S. Shih, "Cholecseg8k: a semantic segmentation dataset for laparoscopic cholecystectomy based on cholec80," *arXiv preprint arXiv:2012.12453*, 2020.