

CholecSeg8k: A Semantic Segmentation Dataset for Laparoscopic Cholecystectomy Based on Cholec80

W.-Y. Hong, C.-L. Kao, Y.-H. Kuo, J.-R. Wang, W.-L. Chang and C.-S. Shih

Graduate Institute of Networking and Multimedia
 Department of Computer Science and Information Engineering
 National Taiwan University, Taipei, Taiwan, 106

cshih@csie.ntu.edu.tw

Abstract—Computer-assisted surgery has been developed to enhance surgery correctness and safety. However, the researchers and engineers suffer from limited annotated data to develop and train better algorithms. Consequently, the development of fundamental algorithms such as Simultaneous Localization and Mapping (SLAM) are limited. This article elaborates the efforts of preparing the dataset for semantic segmentation, which is the foundation of many computer-assisted surgery mechanisms. Based on the Cholec80 dataset [3], we extracted 8,080 laparoscopic cholecystectomy image frames from 17 video clips in Cholec80 and annotated the images. The dataset is named CholecSeg8K and its total size is 3GB. Each of these images are annotated at pixel-level for thirteen classes, which are commonly founded in laparoscopic cholecystectomy surgery. CholecSeg8k is released under the license CC BY-NC-SA 4.0.

I. INTRODUCTION

Endoscopy is a common procedure for detection, diagnosis, and treatment on organs, which are difficult to examine without surgery, such as esophagus, stomach, and colon. In a clinical setting, an endoscopist navigates an endoscope by controlling the handle while watching the output video simultaneously recorded on an external monitor. However, the outcome of endoscopy procedures are highly dependent on the operator’s training and proficiency level. Unsteady hand-control motion and organ motion could significantly affect the outcomes of image analysis and operations.

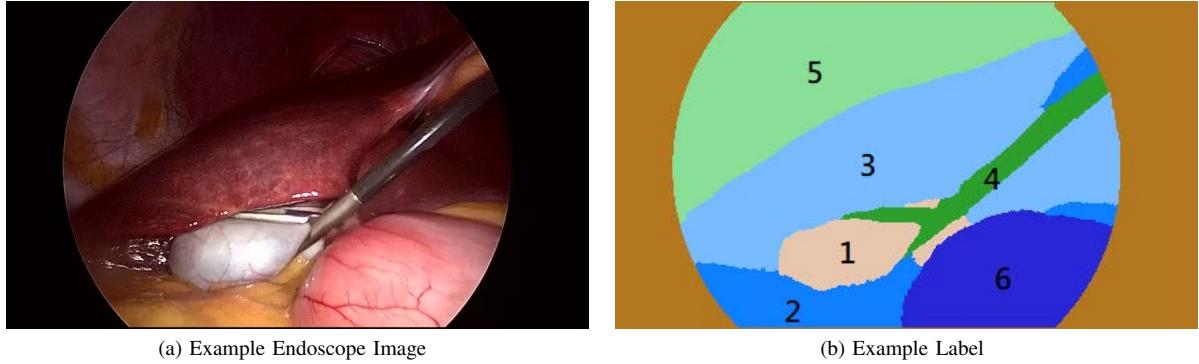
Many computer-assisted systems have been developed to facilitate surgeon conducting endoscopic surgeries by providing guidance and additional context-specific information to the surgeon during the operation [5]. Several of such systems use magnetic or radio-based external sensors to estimate and detect the endoscope location within the patient’s body. However, these are prone to errors and it is not practical to achieve sub-centimeter localization accuracy. For example, the system proposed

by Hu *et al.* [2] requires 16 sensors to achieve 5.6 mm localization accuracy.

Simultaneous Localization and Mapping (SLAM) is an image-based approach for localizing the camera at pixel level accuracy and requires no infrastructure. This method yields a real-time 3D map by comparing sensed data and reference data, and provides an estimate of the camera location in the 3D space. Fortunately, the reference data can be collected in advance. For instance, 256 beam LiDAR and GPS-RTK are used to collect high-definition point cloud maps in advance. During the runtime, the others can use 16 beams LiDAR to collect point cloud data and localize itself at 5 centimeter accuracy. To apply SLAM to endoscope navigation, an accurate assessment of semantic segmentation on images are inevitable.

One of key requirements for accurate SLAM is to use the correct reference data to compare with. In an outdoor environment, one can use GPS to obtain location at large and query for the corresponding reference data, indexed by GPS location. However, this method is not practical for endoscopy. One possible approach is to identify the organs shown in the images and, then, uses the results to query for the reference data. The first part can be conducted by either semantic segmentation, which labels object class for each pixel, or object detection, which labels object class for a bound box. Both semantic segmentation and object detection require well-labelled image data to train the prediction networks, which are *not* widely available for endoscope images.

We construct an open semantic segmentation endoscopic dataset, which is available to medical and computer vision communities. The researchers can eliminate the efforts on training the prediction models, such as FCN, PSPNet, U-Net, and other machine learning approaches, e.g., Deeplab series. This dataset consists



(a) Example Endoscope Image

(b) Example Label

Figure 1: Example of Semantic Segmentation Label of Endoscope Image

of in total 8,080 frames extracted from 17 video clips in Cholec80 dataset [3, 6]. The annotation has thirteen classes. Figure 1 shows one example of the image and annotations. Figure 1(a) is the raw image data from Cholec80 dataset and Figure 1(b) is the color mask to represent the class of the pixels in the image. The details of the annotation will be presented later.

II. DESCRIPTION OF CHOLECSEG8K DATASET

CholecSeg8K dataset uses the endoscopic images from Cholec80 [3], which is published by CAMMA (Computational Analysis and Modeling of Medical Activities) research group, as the base. The research group cooperated with the University Hospital of Strasbourg, IHU Strasbourg, and IRCAD to construct the dataset. Cholec80 contains 80 videos of cholecystectomy surgeries performed by 13 surgeons. Each video in Cholec80 dataset recorded the operations at 25 fps and has the annotation for instruments and operation phases. Our work selected a subset of the highly related videos from Cholec80 dataset and annotated semantic segmentation masks on extracted frames in the selected videos.

Data in CholecSeg8K Dataset are grouped into a two-level directory for better organization and accessibility. Figure 2 shows part of the directory for elaboration. Each directory on the first level collect the data of the video clips extract from Cholec80 and is named by the filename of the video clips. As examples, `video01` and `video09` in Figure 2 are two different video clips, i.e., 01 and 09, in Cholec80 dataset. Each directory on the secondary level tree stores the data for 80 images from the video clip and is named by the video filename and the frame index of the first image in the selected video clip. As examples, directory `video12_15750` stores the data from the 15750-th frame to the 15829-th frame from video file `video12`. Each secondary level directory stores the raw image data, annotation, and color masks for 80 frames. There are a total of 101 directories and the total number of frames is 8,080. The resolution

of each image is 854 pixels \times 480 pixels. The total size of the data set is 3GB.

The reason for not annotating all the frames is to skip the operations not related to semantic segmentation. For example, preparation phase and ending phase are not annotated.

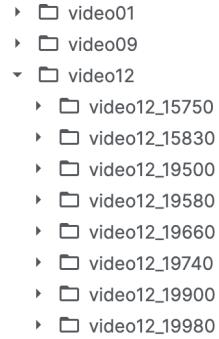


Figure 2: Example of Directory Tree

The number of annotation classes in this dataset is 13, including black background, abdominal wall, liver, gastrointestinal tract, fat, grasper, connective tissue, blood, cystic duct, L-hook electrocautery (Instrument), gallbladder, hepatic vein, and liver ligament. Table I shows the corresponding class names of the class ID in the dataset.

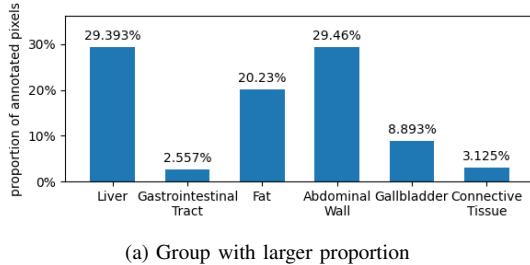
While annotating the pixels, the annotation classes are defined to aim on cholecystectomy surgeries, which is the targeted operations in Cholec80 dataset. Specifically, the annotation classes aim to recognize liver and gallbladder. Other annotation classes, which are not closely related to the surgeries, are defined for broader coverage. In particular, two classes have broader coverage. They are *gastrointestinal tract* and *liver ligament*. The gastrointestinal tract class includes stomach, small intestine, and nearby tissues. The liver ligament class includes coronary ligament, triangular ligament, falciform ligament, liga-

Class ID	Class Name
Class 0	Black Background
Class 1	Abdominal Wall
Class 2	Liver
Class 3	Gastrointestinal Tract
Class 4	Fat
Class 5	Grasper
Class 6	Connective Tissue
Class 7	Blood
Class 8	Cystic Duct
Class 9	L-hook Electrocautery
Class 10	Gallbladder
Class 11	Hepatic Vein
Class 12	Liver Ligament

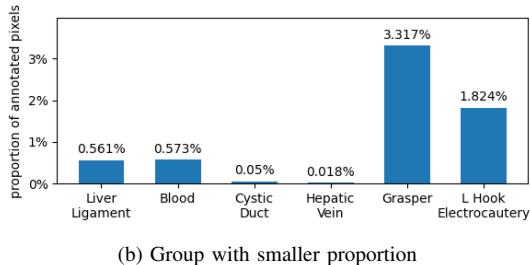
Table I: Class numbers and their corresponding class names

mentum teres (hepatis), ligamentum venosum, and lesser omentum.

In this dataset, not all 13 classes appear in every frame at the same time. The ratios of annotated pixels of 13 classes are shown in Figure 3. As one may notice, the classes are not well balanced, which may lead to poor training results if using without cautions. Figure 3(a)



(a) Group with larger proportion



(b) Group with smaller proportion

Figure 3: Proportion of annotated pixels (y-axis) per class (x-axis)

shows the annotated ratio of the classes whose ratio of annotated pixels are greater than 1%. Among them, the liver class has almost 30% annotated pixels and gallbladder class has 10% annotated pixels among total annotated pixels. Abdominal wall class, which often appears in the background of the image, also has 30% annotated pixels among total annotated pixels. The gastrointestinal tract tissues, including stomach and small intestine, are located next to the liver and gallbladder and, hence, are frequently viewable. Its annotated ratio is greater than 1% as well. Figure 3(b) shows the annotated

ratio of the classes whose ratio of annotated pixels are less than 1%. The L hook electrocautery and grasper classes are two frequently used tools in cholecystectomy surgeries. However, their physical sizes are relatively small, compared to other objects in the images. Hence, the ratio of their annotated pixels are less than 1%.

In the dataset, each frame has three corresponding masks, including one color mask, one annotation mask used by annotation tool, and one watershed mask. The annotated classes for different objects are presented in both color mask and watershed mask. The color mask is annotated for visualization purpose, as shown in Figure 1(b) as an example. The watershed mask is presented for the sake of programming, where each annotated pixel has the same class ID value for three color channels. PixelAnnotationTool [1] is used to annotate the images and generate three masks in PNG format. The annotation tool allows the users to annotate the pixels for pre-defined classes and generate the color mask and the watershed mask.

Figure 4 shows the example masks and raw images for video01_00080 folder. Each item in the dataset

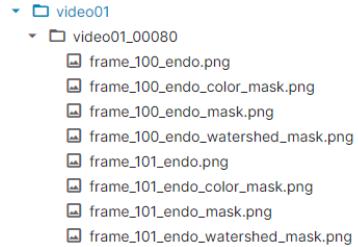


Figure 4: Example of annotation masks and raw images

consists of four files: raw image, color mask, annotation mask, and watershed mask. In CholecSeg8K dataset, four files are all stored in PNG format.

III. EXAMPLES OF ANNOTATED DATA

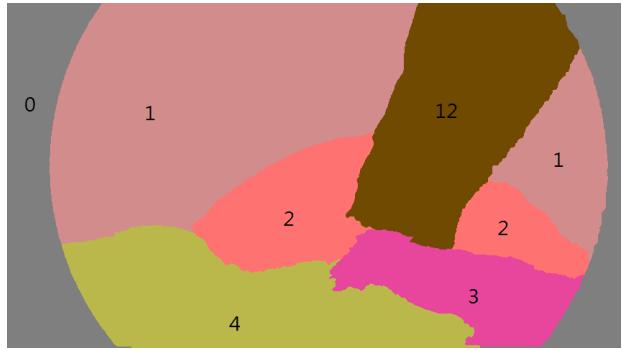
This section presents several pairs of the extracted images and color masks, i.e., annotations of class IDs, from the dataset as examples to demonstrate the dataset. Figure 5, 6, and 7 are three representative examples. The left-hand side of the figures show the raw images extracted from Cholec80 dataset and the right-hand side of the figures show the color mask of the raw images in CholecSec8K dataset.

Figure 5 is a simple case and shows the gallbladder (class ID 10), ligament (class ID 12), and fat (class ID 4) in endoscope image. Part of the liver (class ID is 2) is also viewable. The tissues annotated as ID 12 is the round ligament of liver. As mentioned earlier, we do not specifically annotate different ligament of liver in this dataset and use only one class for all liver ligaments.

Figure 6 shows a more complicated case which has 11 classes and instruments in the image. The image contains



(a) Example Endoscope Image of Gallbladder



(b) Example Color Mask of Gallbladder and nearby tissues

Figure 5: Example of Semantic Segmentation Annotation of Gallbladder Endoscope Image

the liver (class ID 2) in the center and gallbladder (class ID 10) at top central of the image. Two instruments, grasper (class ID 5) and L-Hook electrocautery (class ID 10), are also viewable and annotated. Due to the low brightness in the image, the annotation at the edge of the FoV remains non-trivial for human experts. For example, in this figure, the top edge of grasper (class ID 5) is not trivial to identify. Such ambiguous cases are left as it is because the dataset aims for segmenting human organs.

Figure 7 shows the liver and gallbladder from a different angle. This example also contains the pixels of 11 classes.

IV. ACCESS TO THE DATASET

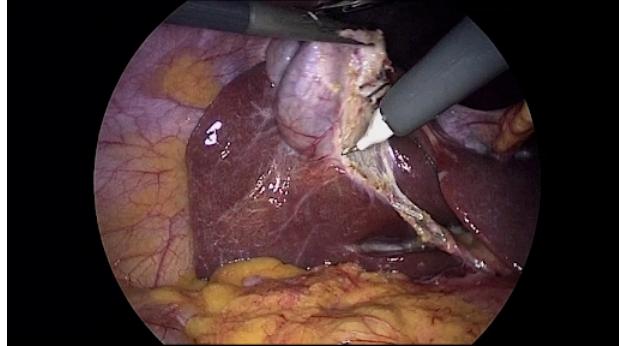
The CholecSeg8k dataset is released under the license CC BY-NC-SA 4.0 and freely available on the Kaggle website [4]. Figure 8 shows the screenshot of the CholecSec8K data-set page on www.kaggle.com. The instructions on using the dataset is available on the dataset page.

ACKNOWLEDGEMENT

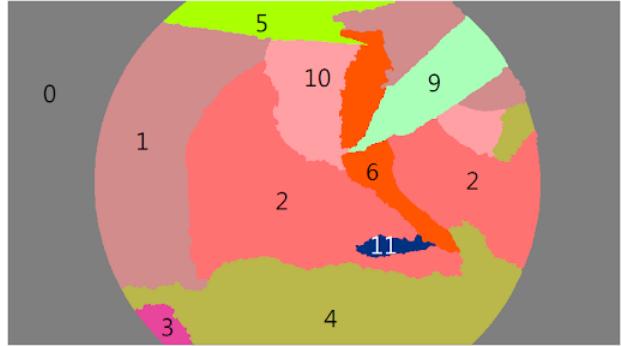
This research was supported in part by the Ministry of Science and Technology of Taiwan (MOST 108-2218-E-002-020).

REFERENCES

- [1] Amaury Bréhéret. *Pixel Annotation Tool*. <https://github.com/abreheret/PixelAnnotationTool>. last access Oct.2020. 2017.
- [2] Chao Hu, M. Q. -. Meng, and M. Mandal. “Efficient Linear Algorithm for Magnetic Localization and Orientation in Capsule Endoscopy”. In: *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*. Jan. 2005, pp. 7143–7146. DOI: 10.1109/IEMBS.2005.1616154.
- [3] *Cholec80 dataset*. last access Oct. 2020. Research Group CAMMA. URL: <http://camma.u-strasbg.fr/> datasets.
- [4] *CholecSeg8k dataset*. last access Dec. 2020. NEWS Lab, Dept. of Computer Science and Information Engineering, National Taiwan University. URL: <http://www.kaggle.com/newslab/cholecseg8k>.
- [5] Bernd Münzer, Klaus Schöffmann, and László Böszörmenyi. “Content-Based Processing and Analysis of Endoscopic Images and Videos: A Survey”. In: *Multimedia Tools Appl.* 77.1 (Jan. 2018), pp. 1323–1362. ISSN: 1380-7501. DOI: 10.1007/s11042-016-4219-z. URL: <https://doi.org/10.1007/s11042-016-4219-z>.
- [6] Andru Twinanda et al. “EndoNet: A Deep Architecture for Recognition Tasks on Laparoscopic Videos”. In: *IEEE Transactions on Medical Imaging* 36 (Feb. 2016). DOI: 10.1109/TMI.2016.2593957.



(a) Example Endoscope Image of Liver and Gallbladder

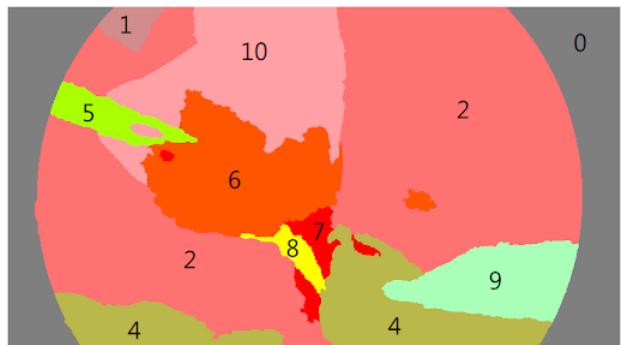


(b) Example Color Mask of Liver and Gallbladder

Figure 6: Example of Semantic Segmentation Annotation for Liver and Gallbladder Endoscope Image



(a) Example Endoscope Image of Liver and Gallbladder



(b) Example Color mask of Liver and Gallbladder

Figure 7: Example of Semantic Segmentation Annotation of Liver and Gallbladder in Endoscope Image

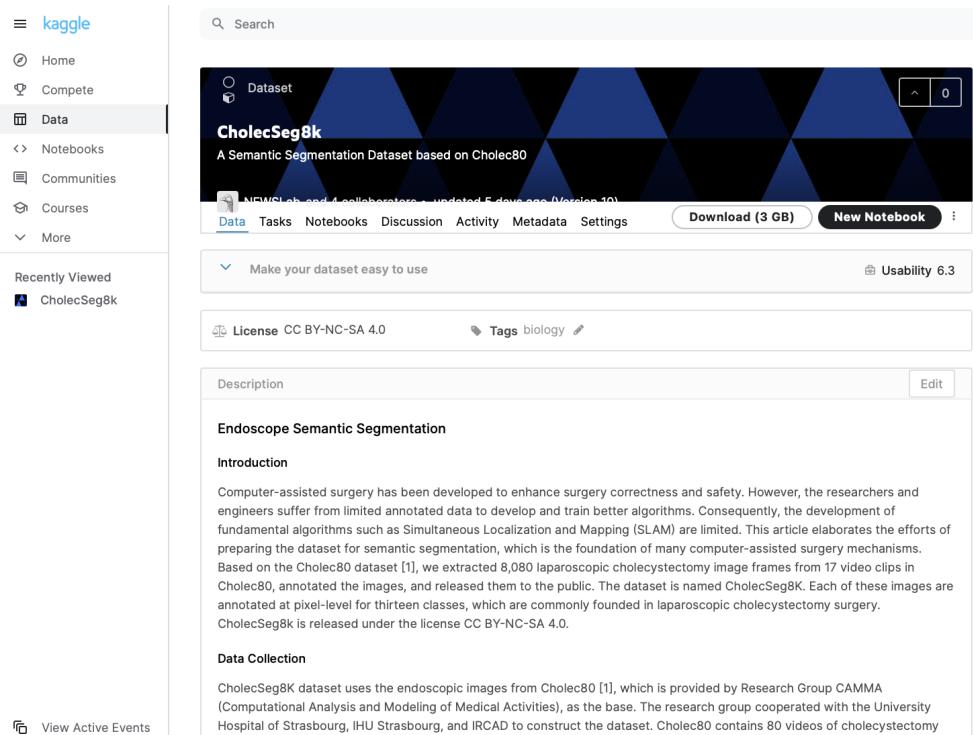


Figure 8: Screenshot of CholecSeg8K dataset page on www.kaggle.com