data

MDPI

Data Descriptor

Cholec80-Boxes: Bounding Box Labelling Data for Surgical Tools in Cholecystectomy Images

Tamer Abdulbaki Alshirbaji [1,2], Nour Aldeen Jalal [2], Herag Arabian [1], Alberto Battistel [1,*], Paul David Docherty [1,3], Hisham ElMoaqet [4], Thomas Neumuth [2] and Knut Moeller [1]

1 Institute of Technical Medicine (ITeM), Furtwangen University, 78054 Villingen-Schwenningen, Germany; tamer.abdulbaki.alshirbaji@hs-furtwangen.de (T.A.A.); knut.moeller@hs-furtwangen.de (K.M.)
2 Innovation Center Computer Assisted Surgery (ICCAS), University of Leipzig, 04103 Leipzig, Germany
3 Department of Mechanical Engineering, University of Canterbury, Christchurch 8041, New Zealand
4 Department of Mechatronics Engineering, German Jordanian University, Amman 11180, Jordan; hisham.elmoaqet@gju.edu.jo
* Correspondence: alberto.battistel@hs-furtwangen.de

Abstract: Surgical data analysis is crucial for developing and integrating context-aware systems (CAS) in advanced operating rooms. Automatic detection of surgical tools is an essential component in CAS, as it enables the recognition of surgical activities and understanding the contextual status of the procedure. Acquiring surgical data is challenging due to ethical constraints and the complexity of establishing data recording infrastructures. For machine learning tasks, there is also the large burden of data labelling. Although a relatively large dataset, namely the Cholec80, is publicly available, it is limited to the binary label data corresponding to the surgical tool presence. In this work, 15,691 frames from five videos from the dataset have been labelled with bounding boxes for surgical tool localisation. These newly labelled data support future research in developing and evaluating object detection models, particularly in the laparoscopic image data analysis domain.

Dataset: https://doi.org/10.5281/zenodo.13170928.

Dataset License: The dataset is made available under the creative commons license CC-BY-NC-SA 4.0

Keywords: surgical tool detection; laparoscopic images; bounding box label

Academic Editor: Li-Yueh Hsu

Received: 14 November 2024
Revised: 17 December 2024
Accepted: 27 December 2024
Published: 8 January 2025

Citation: Abdulbaki Alshirbaji, T.; Jalal, N.A.; Arabian, H.; Battistel, A.; Docherty, P.D.; ElMoaqet, H.; Neumuth, T.; Moeller, K. Cholec80-Boxes: Bounding Box Labelling Data for Surgical Tools in Cholecystectomy Images. *Data* **2025**, *10*, 7. https://doi.org/10.3390/data10010007

1. Summary

Surgical workflow analysis is an active research field that aims to develop intelligent systems for assisting multidisciplinary teams inside the surgical theatre [1]. In addition to monitoring surgical processes and providing the surgical team with context-aware information, these systems can be used for other important purposes such as the optimisation of resources in operating rooms by predicting the remaining surgery time, generating reports and assessing surgical skills [1,2]. Surgical tool detection is key to recognising surgical activities and, consequently, for the development of such assistance systems. In laparoscopic surgeries, videos serve as a rich source of information [3]. Artificial intelligence (AI) offers an opportunity for surgical monitoring and thus has potential for model-based decision support. However, such approaches are constrained by the lack of labelled data [4]. The release of the Cholec80 dataset [5] allowed the development of AI methods, particularly deep learning approaches, to analyse laparoscopic videos. Nevertheless, the labels of the

Cholec80 dataset are limited to surgical phases and binary classes of surgical tools, thus preventing the investigation of object detection methodologies.

Weakly supervised and semi-supervised learning approaches have been proposed to address the absence of comprehensive data labels for surgical tool detection [6–8]. These deep learning methods require only a limited amount of labelled data, leveraging unlabelled or partially labelled data to improve performance and reduce the need for extensive manual labelling. Therefore, a portion of the Cholec80 dataset was labelled for the tool detection task.

The released data are valuable for developing approaches to detect surgical tools in laparoscopic images. Additionally, researchers can use these data to evaluate the performance and generalizability of object detection models. Moreover, they can also be used to augment and enhance the diversity of training data.

The published data were used in three publications, in 'Analysing attention convolutional neural network for surgical tool localisation: a feasibility study' [9], 'Surgical tool classification & localisation using attention and multi-feature fusion deep learning approach' [7] and 'Laparoscopic video analysis using temporal, attention, and multi-feature fusion based-approaches' [10]. In [9], a supervised learning approach, followed by a post-processing method using the class activation maps, was employed to locate the surgical tools in laparoscopic images. In the second publication [7], a weakly-supervised approach was implemented for surgical tool detection. Building on this approach, the latter publication [10] introduced a method for identifying and localising surgical tools jointly with recognising surgical phases. In the aforementioned publications, binary tool presence labels were solely used for training, while the released labelled data were employed to evaluate the performance of surgical tool detection.

The videos were obtained from a publicly available dataset, the Cholec80 dataset [5]. Bounding boxes were manually added to surgical tools in five laparoscopic videos. In addition to the bounding box published labels, a description of the image extraction and labelling protocol are available on an open-access repository to enable reproducibility.

In comparison to our work, other examples of labelling are available, e.g., the m2cai16-tool-locations dataset [11] and the Heidelberg Colorectal (HeiCo) dataset [12]. They offer location labels for surgical tools in laparoscopic images but contain a smaller sample number. The m2cai16-tool-locations dataset provides bounding-box labels for 2532 frames (≈16% of the size of our data) extracted from 10 cholecystectomy videos. The HeiCo dataset consists of labelled masks for 10,040 frames obtained from 30 videos of colorectal procedures, including rectal resection, proctocolectomy and sigmoid resection. Unlike our work, which focuses on cholecystectomy procedures, The HeiCo dataset has different surgical procedures. Additionally, the frames in the m2cai16-tool-locations and the HeiCo dataset were not extracted at a fixed rate, whereas in our work, the frames were consistently labelled at 1 fps, ensuring temporal uniformity and better applicability for video analysis tasks.

2. Data Description

The dataset consists of laparoscopic images and region of interest (ROI) labels for surgical tools present in the images. The images were extracted from five videos (videos 41–45) belonging to the Cholec80 dataset [5]. The images are stored as a .PNG format with a resolution of 854×480. Seven surgical tools are observed in the cholecystectomy videos. The surgical tools represented are grasper, bipolar, hook, scissors, clipper, irrigator and specimen bag, as shown in Figure 1.

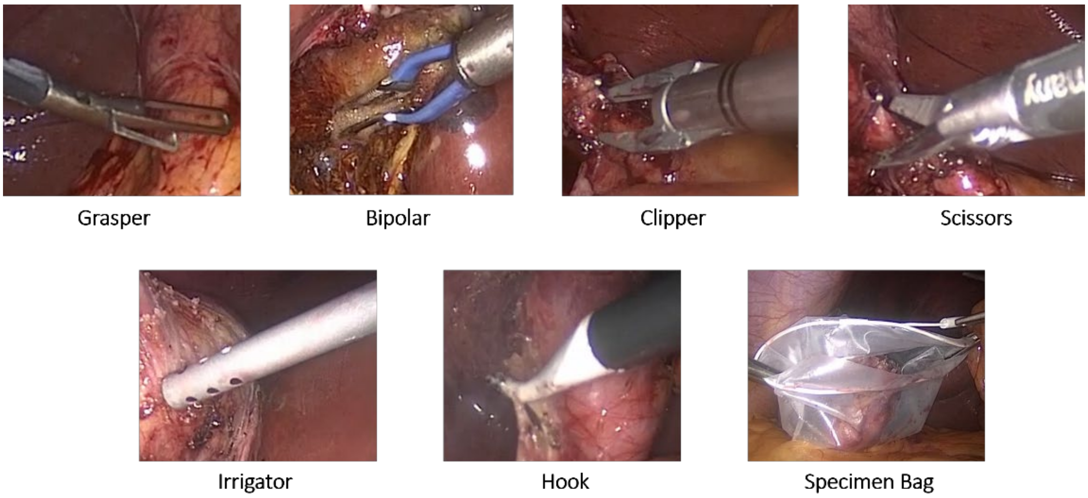

Figure 1. Surgical tools used in the Cholec80 dataset.

The annotated labels are stored in a CSV file. Each row in the file represents a bounding box label for a surgical tool in an image. The labelled data includes the following information:

- *Surgery_num*: the video number from which the image was extracted.
- *Dir*: the directory of the folder containing the image.
- *FrameName*: the image name, formatted as: *Video_SS_fffff.png*, where SS is the *Surgery_num* and fffff is the frame number.
- *NumBBox_inFrame*: the sequential order number of each bounding box within the image. For example, if there are two bounding boxes in an image, *NumBBox_inFrame* is 1 for the first bounding box and 2 for the second bounding box.
- *ToolName*: the surgical tool class, e.g., grasper.
- *BBox_X*: the x-coordinate of the top-left corner of the bounding box in the image.
- *BBox_Y*: the y-coordinate of the top-left corner of the bounding box in the image.
- *BBox_Width*: the width of the bounding box.
- *BBox_Height*: the height of the bounding box.

BBox_X, *BBox_Y*, *BBox_Width* and *BBox_Height* are in pixels. When multiple surgical tools are present in an image, each bounding box corresponding to a surgical tool is assigned a unique number, encoded in *NumBBox_inFrame*. Additionally, images that do not display any surgical tools are not labelled with a bounding box, resulting in a mismatch between the number of labels and the number of images, as shown in Figure 2.

Figure 2 illustrates the distribution of binary labels and bounding box labels for surgical tools in the images extracted from videos 41 to 45 of the Cholec80 dataset. These videos were selected from the second half of the dataset, which is primarily used for testing purposes. The total number of bounding boxes exceeds the number of binary tool presence labels over all images since up to three surgical tools can appear in a single image. Moreover, the labelling protocol of this work differs from that of the Cholec80 labelling protocol [5]. In the Cholec80 binary tool presence labels, a surgical tool was considered as present in the image if at least half of its tip was seen. In contrast, this work includes bounding box labels for any characteristic part of a surgical tool, such as the tip, that appears in the image. As a result, the number of bounding boxes (green bars) for a surgical tool is higher than the number of images belonging to that tool class (grey bars), as shown in Figure 2. The difference is most notable for the grasper, as multiple graspers were observed during certain parts of the procedures.

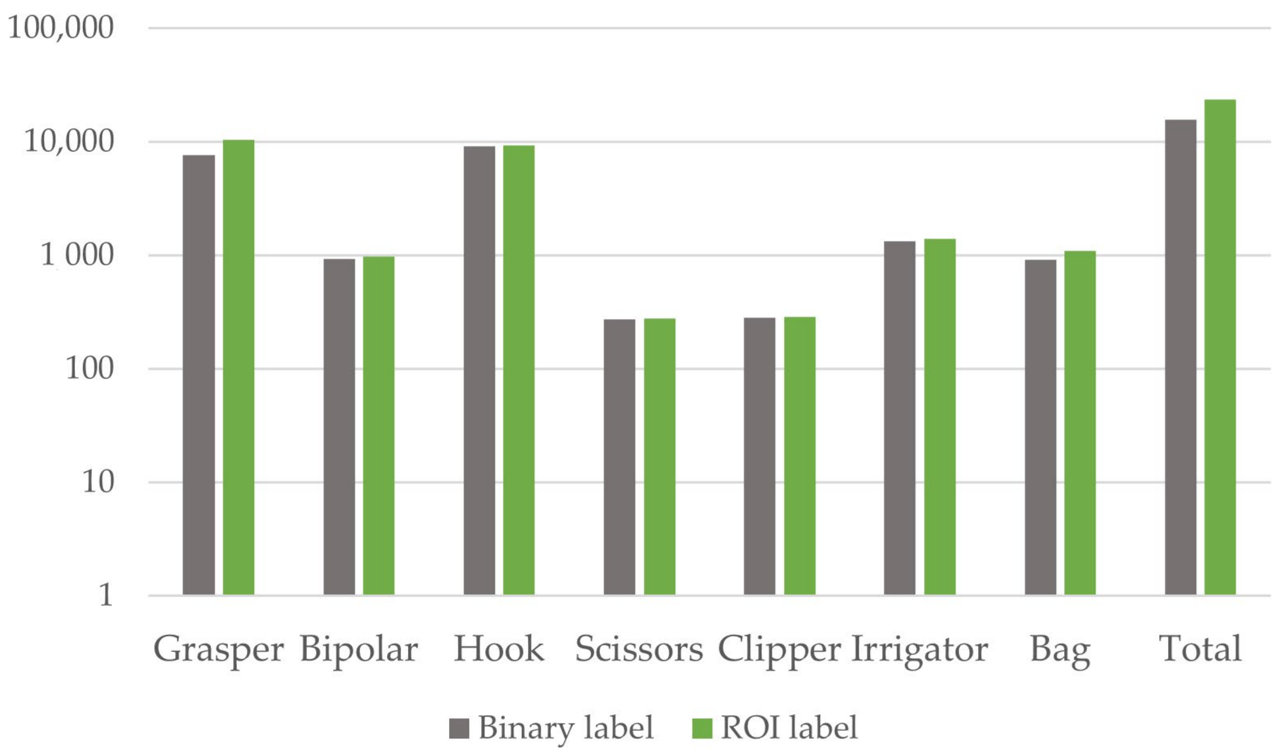

Figure 2. The absolute frequency of binary and bounding box labels for surgical tools in five videos (videos 41–45) from the Cholec80 dataset.

3. Methods

3.1. Data Collection

The videos used for the data labelling were obtained from the Cholec80 dataset [5], which includes recordings from laparoscopic cameras for 80 cholecystectomy procedures. The videos were acquired at 25 Hz and stored in MP4 format with a resolution of 854×480. Labels for surgical phases and surgical tool presence are provided in this dataset at rates of 25 Hz and 1 Hz, respectively. The labels for surgical tool presence are binary labels and annotated according to the seven types of surgical tools (grasper, bipolar, hook, scissors, clipper, irrigator and specimen bag) used during the procedures included in the Cholec80 dataset. The surgical tool types are presented in Figure 1.

Images were extracted from the videos at a rate of 1 Hz. The extracted images of each video were stored in a separate folder. The total number of the extracted images, the duration of each video and their respective mean over these videos are presented in Table 1. All seven surgical tools available in the complete Cholec80 dataset were used in the selected procedures (videos 41–45). However, each surgical tool was used at varying frequencies as shown in Figure 3, which demonstrates the number of tool occurrences in each of the selected five videos (41–45).

Table 1. Duration and number of images for the considered videos from the Cholec80 dataset.

Procedure	Duration (min)	Number of Extracted Images
41	51.72	3103
42	61.87	3712
43	39.37	2362
44	52.12	3127
45	56.45	3387
mean ± Std	52.31 ± 8.31	3138.2 ± 498.75

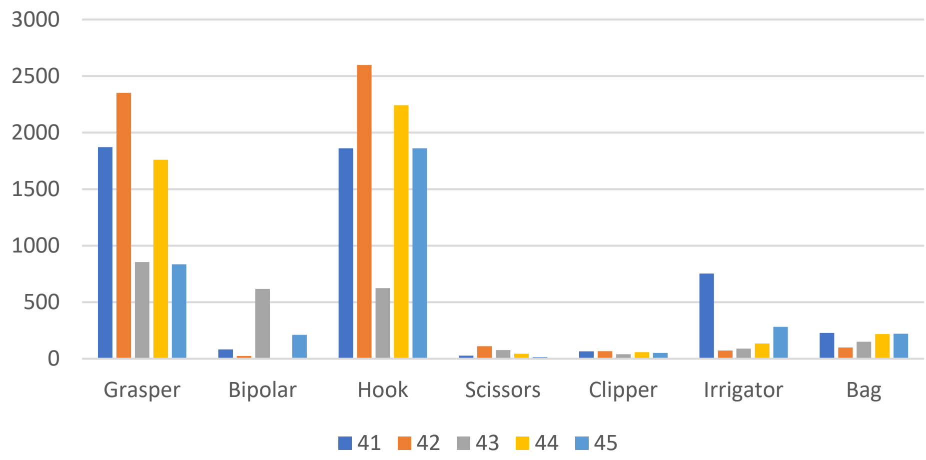

Figure 3. Occurrence of surgical tools in videos 41 to 45 from the Cholec80 dataset.

3.2. Labelling Protocol

Three engineers, with a background in biomedical technology, conducted the data labelling. Two of them manually labelled the extracted images by adding bounding boxes around the surgical tools, and the third engineer reviewed the bounding boxes. The labellers followed predefined rules to ensure consistency and worked on independent sections of the dataset. The following rules were applied:

- Videos were labelled for the localisation of surgical tools at a frequency of 1 Hz.
- A tool was considered present in the image if any part of its tip was visible. In cases where the tool type could not be identified from a single frame, labellers referred to frames before or after the target frame to assign the correct label class. However, information about the difficulty level of each instance was not included in the dataset. Furthermore, if the tip was not visible or was obscured by an anatomical structure, the tool was labelled as absent, even if its shaft was visible. Figure 4 illustrates this labelling rule, with three case examples: a visible tool tip (a), a partially visible tool tip (b) and a hidden tool tip in (c). Even though the tool shaft is visible in Figure 4c, a bounding box was not added as no part of the tool tip was visible in the scene.
- The smallest possible bounding box was added around the tool tip and the initial part of the tool shaft. This rule was applied only for surgical tools that have a distinct tip. Those surgical tools are grasper, bipolar, hook, scissors and clipper. Figure 5 presents an example for each of these surgical tools.
- The irrigator was labelled based on its shaft since the shaft was considered sufficient to identify this surgical tool. Figure 6 illustrates an example of labelling an irrigator. In Figure 6b, the irrigator tip is covered by anatomical structures and is not visible. However, this surgical tool can still be identified by the shaft.
- Since the specimen bag does not have a distinct tip or shaft, it was labelled by adding the smallest bounding box containing the entire specimen bag in the scene, without considering the strap. Figure 7 shows examples for labelling the specimen bag. The presented examples include a folded bag (a), an opened empty bag (b) and a closed bag filled with tissues (c).

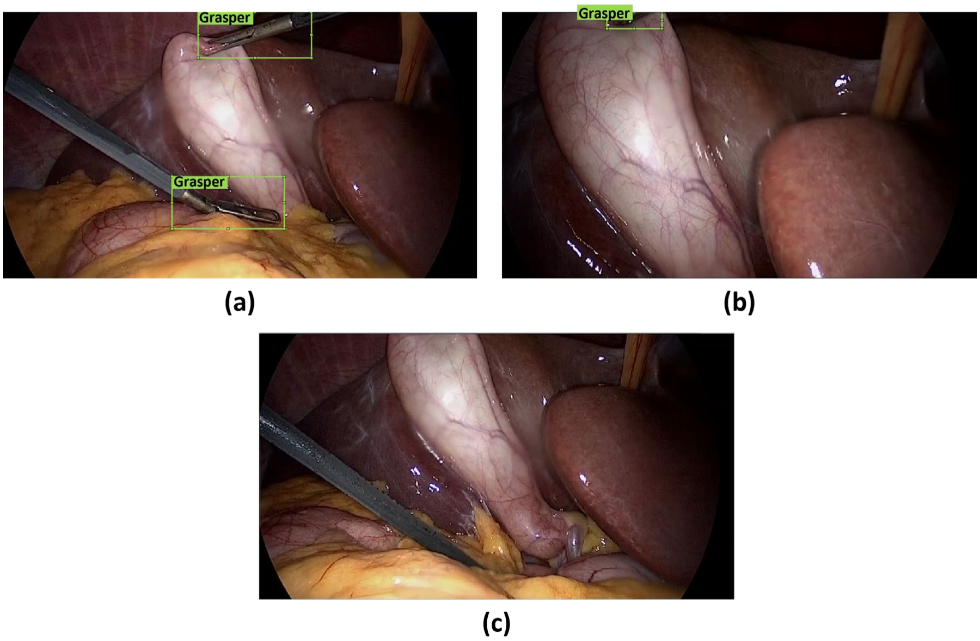

Figure 4. Illustrations of tool tip annotation: (**a**) grasper tip is visible, (**b**) grasper tip partially visible and (**c**) grasper tip is completely obscured by anatomical tissues.

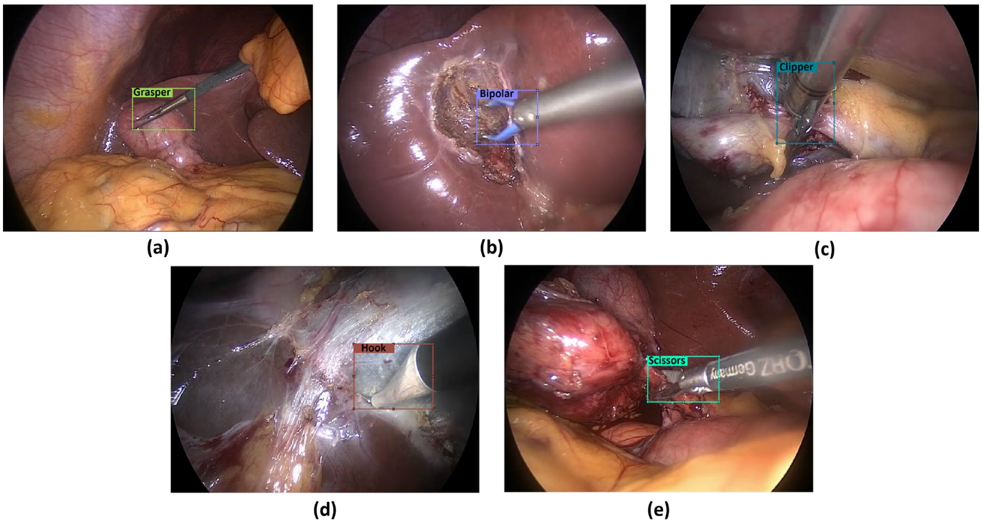

Figure 5. Examples for labelling (**a**) grasper, (**b**) bipolar, (**c**) clipper, (**d**) hook and (**e**) scissors.

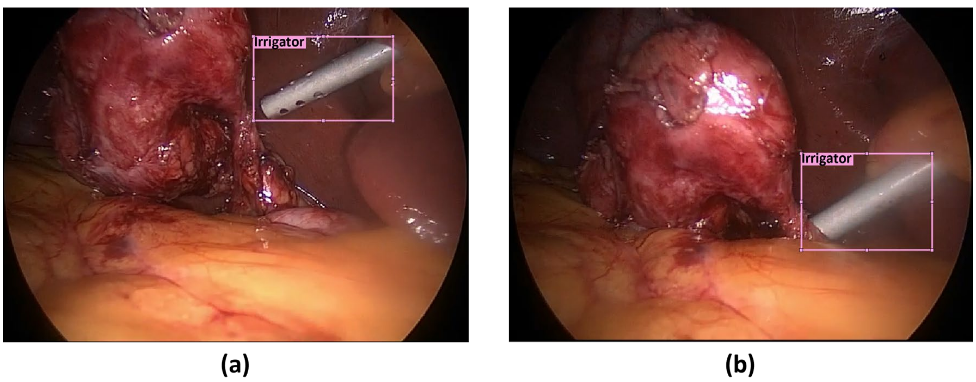

Figure 6. Two examples of labelling the irrigator: (**a**) irrigator tip is visible and (**b**) irrigator tip is obscured by anatomical tissues.

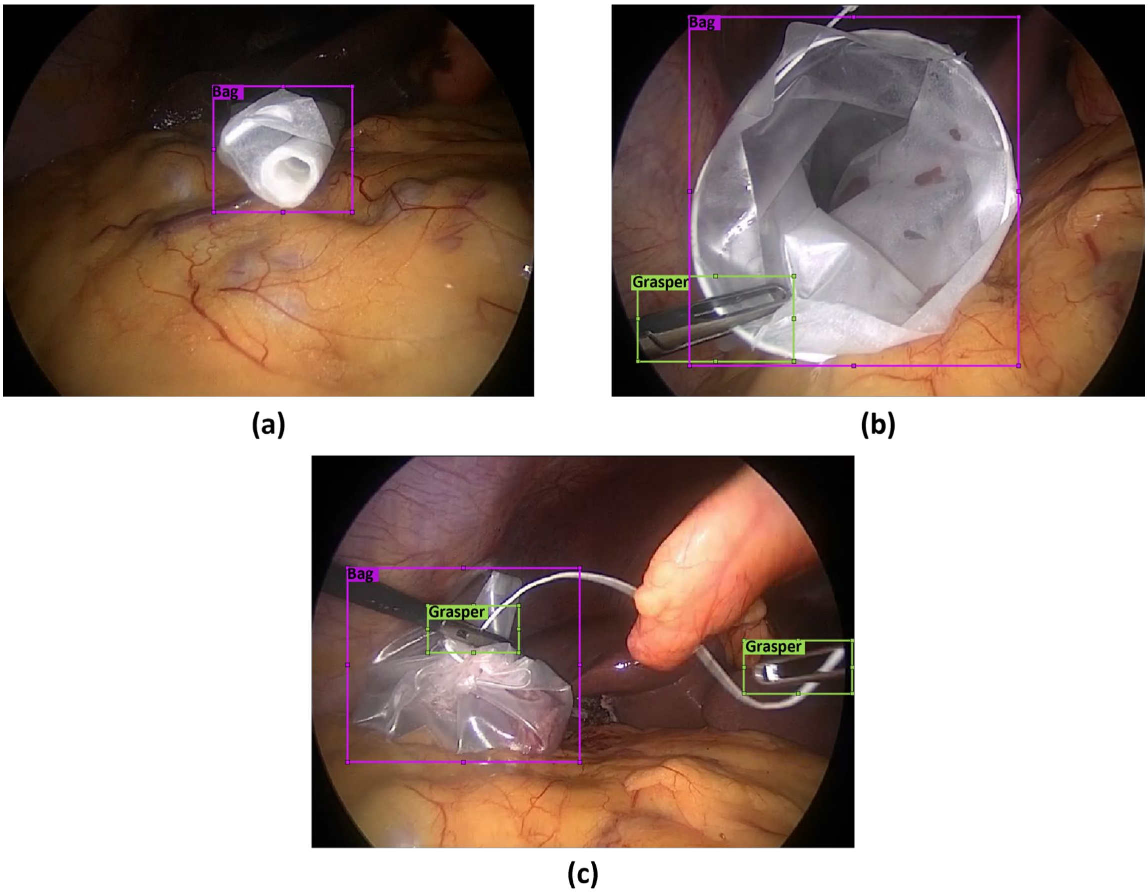

Figure 7. Examples of labelling the specimen bag: (**a**) a folded specimen bag, (**b**) an empty specimen bag and (**c**) a closed specimen bag filled with tissues.

3.3. Labelling Software and Process

The MATLAB Video Labeller toolbox (R2021a, The MathWorks, Natick, MA, USA) was employed to annotate the location of the surgical tools in the cholecystectomy videos. The labelling process with this toolbox involved the following steps:

1. Creation of a labelling session:
   A labelling session was created for each surgical procedure. The folder containing extracted images was imported into the labelling session to load them as a sequence of frames. Images in the imported sequence are ordered based on their names. Storing extracted images in *Video_SS_fffff.png* format (described in Section 2), ensures that images are imported in the correct order. The default MATLAB timestamp was used since the images were extracted at 1 Hz. Therefore, the timestamp, which is in seconds, ranges from 0 to the number of images minus one.
2. Labels definition:
   Seven labels were created, each corresponding to a certain surgical tool. These labels were transcribed according to the surgical tool names defined in the Cholec80 dataset to maintain consistency between classification and localisation labels. A label definition file was exported and stored for use in all labelling sessions, ensuring that the same label definitions were consistently applied across different videos and sessions.
3. Image labelling:
   Two labellers performed the labelling process on an image-by-image basis. The labellers navigated through the image sequence using the toolbox's interactive interface to draw bounding boxes around the surgical tools, adhering to the defined labelling protocol (see Section 3.2). For each surgical tool visible in the scene, a ROI label was

added. The rectangular ROI label was chosen for labelling the location of all surgical tool types.

4. Labelled data export:
   After completing the labelling process for a cholecystectomy video, the labelled data were exported and stored in a MAT file. The file includes information about the image sequence, timestamps in seconds and the labelled data. The labelled data are organised in a timetable variable, which is a table containing the ROI labels for the seven surgical tools associated with the timestamps. This table includes eight columns, one for the timestamps and the remaining seven for each surgical tool. Each row in the table represents the labelling data for all surgical tools in a single image corresponding to the associated timestamp.
   The labelling data for each tool type in the image are organised in a cell array to accommodate multiple ROI labels if multiple surgical tools of the same type are present. Columns of nonvisible tools contain empty cell arrays. Each ROI label is an array containing the starting position of the bounding box (x and y coordinates) and its width and height in pixels. In addition to the labelling data, the corresponding labelling session was saved to allow modifications or corrections if necessary.
5. Verification:
   Labels were reviewed by the third engineer independently to ensure consistency and accuracy. The review process was conducted using the MATLAB Video Labeller toolbox, which facilitates adjustment and verification of ROI labels. The engineer played the image sequence as a video using the toolbox′s play button. When inaccurate ROI labels (e.g., incorrect bounding box location or size) were identified, the issue was reported and discussed with the labellers to make the necessary modifications. Due to the nature of laparoscopic surgery, the precise bounding ROI can require a subjective choice to be made. For example, smoke may partially cover part of a tool [13], or interpretation of visual interference may be inconsistent across labellers. Hence, the quality of ROI labels for all instances cannot be guaranteed, and some noisy labels, such as ROI labels that do not perfectly fit the tool tip, may exist.
6. Data curation:
   The labelling data were organized and stored in a format that enables ease of use and ensures compatibility with various programming frameworks and software. To achieve this, the ROI label of each surgical tool was extracted from the MATLAB labelling data and stored as a table in a CSV file. Each row in this table represents an ROI label. Consequently, a single image may have multiple rows corresponding to the number of visible surgical tools, even if they belong to the same tool type. Additionally, image information (name, path and procedure number) is also included in the file. The content of this file is described in Section 2.

4. Conclusions

This work developed a dataset that included ROI indications from five videos in the Cholec80 data set. In total 15691 frames were labelled and cross-validated by 3 biomedical engineers. This dataset will enable researchers to develop AI models in tool localisation and recognition and benefit the scientific communities' goals for improving model-based surgical monitoring and assistance. Upon agreement of appropriate citation, the dataset is freely available to download at doi.org/10.5281/zenodo.13170928 under Creative Commons licence CC-BY-NC-SA 4.0.

Author Contributions: Conceptualization, T.A.A. and K.M.; methodology, T.A.A., N.A.J. and H.A.; software, T.A.A. and N.A.J.; validation, T.A.A., N.A.J., A.B. and H.E.; formal analysis, T.A.A., N.A.J., P.D.D., T.N. and K.M.; investigation, N.A.J., T.A.A. and H.A.; resources, K.M.; data curation, N.A.J.,

T.A.A. and H.A.; writing—original draft preparation, T.A.A.; writing—review and editing, N.A.J., H.A., A.B., H.E., T.N., P.D.D. and K.M.; visualization, T.A.A. and H.A.; supervision, P.D.D., T.N. and K.M.; project administration, K.M.; funding acquisition, K.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by German Federal Minister of Research and Education (BMBF), grant number CoHMed/PersonaMed-B 13FH5I09IA and the Deutscher Akademischer Austauschdienst (DAAD) under grant number AIDE-ASD FKZ 57656657.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original data presented in the study are openly available for academic research in Zenodo repository at doi.org/10.5281/zenodo.13170928 under Creative Commons licence CC-BY-NC-SA 4.0.

Acknowledgments: The authors thank the creators of the Cholec80 dataset for providing this valuable resource.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Maier-Hein, L.; Eisenmann, M.; Sarikaya, D.; März, K.; Collins, T.; Malpani, A.; Fallert, J.; Feussner, H.; Giannarou, S.; Mascagni, P.; et al. Surgical Data Science—From Concepts toward Clinical Translation. *Med. Image Anal.* **2022**, *76*, 102306. [CrossRef]
2. Abdulbaki Alshirbaji, T.; Jalal, N.A.; Docherty, P.D.; Neumuth, T.; Möller, K. Robustness of Convolutional Neural Networks for Surgical Tool Classification in Laparoscopic Videos from Multiple Sources and of Multiple Types: A Systematic Evaluation. *Electronics* **2022**, *11*, 2849. [CrossRef]
3. Alshirbaji, T.A.; Jalal, N.A.; Möller, K. Surgical Tool Classification in Laparoscopic Videos Using Convolutional Neural Network. *Curr. Dir. Biomed. Eng.* **2018**, *4*, 407–410. [CrossRef]
4. Nwoye, C.I.; Bose, R.; Elgohary, K.; Arboit, L.; Carlino, G.; Lavanchy, J.L.; Mascagni, P.; Padoy, N. Surgical Text-to-Image Generation. *arXiv* **2024**, arXiv:2407.09230.
5. Twinanda, A.P.; Shehata, S.; Mutter, D.; Marescaux, J.; de Mathelin, M.; Padoy, N. EndoNet: A Deep Architecture for Recognition Tasks on Laparoscopic Videos. *IEEE Trans. Med. Imaging* **2017**, *36*, 86–97. [CrossRef]
6. Zia, A.; Bhattacharyya, K.; Liu, X.; Berniker, M.; Wang, Z.; Nespolo, R.; Kondo, S.; Kasai, S.; Hirasawa, K.; Liu, B.; et al. Surgical Tool Classification and Localization: Results and Methods from the MICCAI 2022 SurgToolLoc Challenge. *arXiv* **2023**, arXiv:2305.07152.
7. Jalal, N.A.; Alshirbaji, T.A.; Docherty, P.D.; Arabian, H.; Neumuth, T.; Moeller, K. Surgical Tool Classification & Localisation Using Attention and Multi-Feature Fusion Deep Learning Approach. *IFAC-PapersOnLine* **2023**, *56*, 5626–5631. [CrossRef]
8. Abdulbaki Alshirbaji, T.; Jalal, N.A.; Docherty, P.D.; Neumuth, T.; Möller, K. A Deep Learning Spatial-Temporal Framework for Detecting Surgical Tools in Laparoscopic Videos. *Biomed. Signal Process. Control* **2021**, *68*, 102801. [CrossRef]
9. Jalal, N.A.; Arabian, H.; Alshirbaji, T.A.; Docherty, P.D.; Neumuth, T.; Moeller, K. Analysing Attention Convolutional Neural Network for Surgical Tool Localisation: A Feasibility Study. *Curr. Dir. Biomed. Eng.* **2022**, *8*, 548–551. [CrossRef]
10. Jalal, N.A.; Alshirbaji, T.A.; Docherty, P.D.; Arabian, H.; Laufer, B.; Krueger-Ziolek, S.; Neumuth, T.; Moeller, K. Laparoscopic Video Analysis Using Temporal, Attention, and Multi-Feature Fusion Based-Approaches. *Sensors* **2023**, *23*, 1958. [CrossRef]
11. Jin, A.; Yeung, S.; Jopling, J.; Krause, J.; Azagury, D.; Milstein, A.; Fei-Fei, L. Tool Detection and Operative Skill Assessment in Surgical Videos Using Region-Based Convolutional Neural Networks. In Proceedings of the 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Tahoe, NV, USA, 12–15 March 2018; pp. 691–699.
12. Maier-Hein, L.; Wagner, M.; Ross, T.; Reinke, A.; Bodenstedt, S.; Full, P.M.; Hempe, H.; Mindroc-Filimon, D.; Scholz, P.; Tran, T.N. Heidelberg Colorectal Data Set for Surgical Data Science in the Sensor Operating Room. *Sci. Data* **2021**, *8*, 101. [CrossRef]
13. Alshirbaji, T.A.; Jalal, N.A.; Mündermann, L.; Möller, K. Classifying Smoke in Laparoscopic Videos Using SVM. *Curr. Dir. Biomed. Eng.* **2017**, *3*, 191–194. [CrossRef]