GCPM-SL-20: a pixel-level segmentation dataset with location and extent labels for peritoneal metastasis during staging laparoscopy for gastric cancer

Shiyun Peng[1,2*], Yixi Li[1,3*], Ying Zhang[3*], Jiawei Chen[3*], Zhixian Yan[3*], Hongyu Chen[3], Yuxuan Wang[4], Mianrun Hu[2], Chunqi Liang[2], Cheng Peng[2], Jiang Yu[1†], Yanfeng Hu[1†], Hong Wang[5†], Hao Chen[1†] , Longfei Gou[1†]

[1] Department of General Surgery & Guangdong Provincial Key Laboratory of Precision Medicine for Gastrointestinal Tumor, Nanfang Hospital, Southern Medical University, Guangzhou, China.

[2] The Second School of Clinical Medicine, Southern Medical University, Guangzhou, China.

[3] The First School of Clinical Medicine, Southern Medical University, Guangzhou, China.

[4] School of Biomedical Engineering, Southern Medical University, Guangzhou, China

[5] Key Laboratory of Biomedical Information Engineering of Ministry of Education, School of Life Science and Technology, Xi'an Jiaotong University, Xi'an, Shaanxi, China

[*] **These authors contributed equally:** Shiyun Peng, Yixi Li, Ying Zhang, Jiawei Chen, Zhixian Yan

[†] **Corresponding authors:** Longfei Gou (calvin_smu@163.com), Hao Chen (chenhao.05@163.com), Hong Wang (hongwang01@xjtu.edu.cn), Yanfeng Hu (yfenghu@qq.com), Jiang Yu (balbc@163.com)

Abstract

Peritoneal metastasis (PM) is the most common distant metastatic pattern and leading cause of mortality in advanced gastric cancer (GC), but tiny or occult PM is often missed by conventional preoperative imaging. Staging laparoscopy (SL) is recommended by multiple international guidelines as the gold standard for the diagnosis of PM, yet intraoperative recognition still depends heavily on surgeons' experience. Computer vision (CV) may assist PM recognition during SL, yet public surgical datasets mostly focus on normal anatomy and instruments, with little metastatic lesion data. Here, we present the GCPM-SL-20 dataset, the first publicly available pixel-level dataset, containing 1920 high-resolution frames from 20 SL procedures in patients with GC: 1013 positive frames with PM lesion segmentation masks and 907 negative frames without lesions. All frames were labeled by peritoneal location, while positive frames were labeled according to metastatic extent. Five-fold cross-validation with eight representative and popular segmentation models showed that Swin Transformer achieved the highest Dice coefficient of 0.55. Overall, the GCPM-SL-20 dataset uniquely integrates pixel-level PM lesion masks, peritoneal location and metastatic extent labels, and visually similar negative frames, providing essential information for developing CV models for accurate intraoperative PM recognition during SL for GC.

Background & Summary

Gastric cancer (GC) is the fifth most prevalent malignancy and the third leading cause of cancer-related mortality worldwide[1]. Peritoneal metastasis (PM) represents the most common pattern of distant metastasis in GC, occurring in approximately 14% of patients at initial presentation[2, 3]. Once PM is confirmed, the disease is immediately classified as stage IV GC, resulting in a complete shift of the treatment strategy towards predominantly systemic therapy[4]. If PM goes undetected, treatment is delayed or suboptimal, and the 5-year survival rate drastically declines to below 8%[5]. Consequently, the accurate diagnosis of PM is critical for guiding appropriate clinical management[6]. Currently, contrast-enhanced computed tomography (CT) serves as the primary diagnostic method for clinical staging of GC, while positron emission tomography (PET) is typically

utilized when distant metastasis is suspected[7, 8]. However, these diagnostic methods demonstrate limited performance in detecting PM preoperatively and are particularly inadequate for identifying tiny or occult metastatic lesions, yielding a sensitivity of merely 11%-48%, which frequently leads to missed diagnoses[9, 10]. To overcome these diagnostic limitations, staging laparoscopy (SL) serves as a minimally invasive and cost-effective procedure for accurate GC staging, as recommended by several international treatment guidelines[8, 11, 12]. SL achieves a direct, comprehensive inspection of the abdominal cavity, enabling the detection of occult PM that typically is undetectable by conventional preoperative imaging. Given its diagnostic superiority, SL plays a pivotal role in the evaluation of PM[11, 13].

However, the diagnostic accuracy of SL remains highly dependent on the surgeons' clinical experience, with reported missed diagnosis rate for intraoperative identification of PM ranging from 5.1% to 15.4%[14, 15, 16, 17]. As SL is typically performed by junior surgeons under the supervision of senior surgeons, tiny or occult metastatic lesions are difficult to identify. Furthermore, PM lesions exhibit significant variability in their size, shape, and color[18], making intraoperative detection particularly demanding and requiring substantial surgical experience. These PM lesions are highly prone to being missed due to visual fatigue, low tissue contrast, restricted visibility, and uneven illumination within the surgical cavity. The omission of PM ultimately leads to incorrect tumor staging and potentially unnecessary gastrectomies[18, 19], which hinders the selection of the optimal intraoperative treatment strategy and results in poorer patient survival outcomes. Consequently, high-quality lesion-annotated laparoscopic image datasets are urgently needed to characterize the intraoperative visual features of PM lesions and provide reusable resources for surgical education. They may also support the development of real-time intraoperative assistance tools, thereby improving PM lesion detection accuracy during SL.

Surgical Data Science (SDS) is an interdisciplinary field that leverages artificial intelligence (AI) and data analytics to systematically analyze and model surgical data, aiming to enhance the quality of surgical care[20, 21]. Specifically, within the intraoperative setting, SDS utilizes computer vision (CV) to analyze surgical videos, providing context-aware assistance through applications such as intraoperative diagnosis[22], surgical navigation[23] and skill assessment[24]. However, the successful development of SDS applications relies fundamentally on the availability of high-quality surgical datasets. Currently, publicly available surgical datasets such as SurgAI[25], Heidelberg colorectal dataset[26], Cataract-1K Dataset[27], and SLAM dataset[28], mainly focusing on anatomical structures, surgical instruments, phases, and actions, but generally lack annotations for intraoperative lesions, particularly metastatic tumor lesions. This gap reflects the difficulty of identifying metastatic lesions in surgical videos and generating reliable pixel-level annotations, highlighting the clinical and research value of a public laparoscopic dataset for PM during SL in patients with GC.

Despite the absence of publicly available surgical image datasets of lesions, recent studies have increasingly investigated AI-driven intraoperative diagnosis, particularly for the recognition and evaluation of PM during SL[29]. Specifically, existing studies have explored a variety of CV tasks for PM lesion analysis, including lesion object detection[30], pixel-level segmentation[29], benign-malignant classification[22], prognosis prediction[31], and lesion size estimation[32]. Noteworthy, we developed a segmentation-based real-time AI system for PM lesion recognition during SL in patients with GC[29]. A comprehensive summary of current research dedicated to intraoperative PM evaluation is provided in Table 1. However, the datasets used in these pioneering studies have not been publicly released, which remains a critical bottleneck for algorithmic reproducibility and the broader application of SDS in SL.

Our previous construction and public release of the Laparoscopic Gastric Cancer Key Vascular Anatomy Dataset (LapGC-KVAD-30) established a foundation for developing high-quality surgical image datasets in laparoscopic GC surgery[33]. As summarized in Table 1, existing AI studies for PM identification and evaluation have involved both gynecological and digestive-system malignancies. However, AI models developed for gynecological cancers may not be directly transferable to GC because PM in gynecological malignancies can differ from that in digestive-system malignancies in metastatic distribution and laparoscopic appearance[34]. Building on our prior dataset development experience and addressing the lack of publicly available surgical image datasets for PM in GC, we further developed the Gastric Cancer Peritoneal Metastasis Staging Laparoscopy (GCPM-SL-20) dataset,

also referred to as Southern Medical University - Nanfang Hospital - Gastric Cancer - 02 Dataset (SMU-NF-GC-02 Dataset). Derived from 20 complete SL videos from patients with GC, the dataset comprises 1920 expert-validated, manually annotated high-resolution frames, including 1013 positive frames depicting pathologically confirmed PM lesions and 907 negative frames without PM lesions. As the first open-source dataset for PM identification during SL, the GCPM-SL-20 dataset features comprehensive, expert-validated annotations that combine pixel-level lesion segmentation with structured labels, systematically characterizing both the metastatic extent and peritoneal location. Figure 1 provides an overview of the dataset construction and annotation workflow. Aimed at providing a foundational resource for the development of AI-driven intraoperative diagnosis and assistance systems, the GCPM-SL-20 dataset enables surgeons and researchers to gain a deeper understanding of PM characteristics during SL. Moreover, given the shared lesion features of PM across digestive system malignancies, including GC, colorectal, hepatic, pancreatic, and biliary tract cancers, this dataset may also inform AI development for intraoperative PM recognition during SL of these malignancies. This dataset can contribute to refining intraoperative staging and real-time lesion identification, ultimately enhancing the accuracy of SL and potentially improving clinical outcomes for GC patients.

Methods

Data collection

This dataset comprises 20 complete videos of SL performed at the initial presentation of patients with advanced GC between April 2017 and September 2022 at Nanfang Hospital, Southern Medical University. None of the enrolled patients had received perioperative radiotherapy or chemotherapy. All laparoscopic videos were captured with KARL STORZ imaging systems and saved as MPG4 format with a resolution of 1920×1080 pixels. All metastatic lesions were confirmed by SL and pathological biopsy. To ensure patient confidentiality, all identifiable patient information (e.g., names, ages, and surgery dates) was anonymized during dataset construction. To avoid using patient-related identifiers, each video was given a simple numeric ID (e.g., 1, 2, 3). This study was approved by the Medical Ethics Committee of Nanfang Hospital, Southern Medical University (Approval Number: NFEC-2023-392). Given the retrospective, observational design of the study and the use of anonymized patient data, the committee waived the requirement for individual patient consent.

Frame extraction

All videos included in this dataset were curated to exclude peritoneal cytology, lesion excision biopsy, and closure of the abdominal wall incision; only clips depicting the intra-abdominal exploration of various peritoneal locations were retained. Figure 2a displays the durations of the original and edited videos included in the dataset. Video frames were manually extracted from the edited videos using the FFmpeg framework (www.ffmpeg.org) by three expert surgeons (H.C., Y.H., and J.Y.).

The dataset consists of both positive and negative frames. Positive frames were defined as frames displaying the peritoneal surface with at least one visible metastatic lesion. To ensure data quality, the selection of positive frames adhered to the following criteria: (1) metastatic lesions must be located on the peritoneum; (2) the intraoperative observation time of the lesions must be sufficient (at least 5 seconds) to allow frame extraction from multiple angles and distances; (3) frames are clear without excessive glare, surgical smoke or severe blurring; (4) the lesions shown in the frames was pathologically confirmed as PM lesions by biopsy. Negative frames were defined as frames that capture peritoneal areas without visible metastatic lesions and maintain adequate visual clarity, consistent with the third criterion applied to positive frames. These negative frames were included to represent visually similar non-metastatic regions, including normal anatomical structures, inflammatory changes, and reflections, thereby helping models distinguish PM lesions from common intraoperative mimics and reduce false-positive detections during intraoperative application. Figures 2b and 2c summarize the distributions of positive and negative frames across the dataset.

Annotation

The annotation group consisted of three expert surgeons (H.C., Y.H., and J.Y.), each with experience in more

than 600 laparoscopic GC surgeries and serving as a core member of CLASS group (Chinese Laparoscopic Gastrointestinal Surgery Study Group), as well as six medical annotators, including one novice surgeon (L.G.) and five medical students (S.P., Y.L., Y.Z., J.C., and Z.Y.). The novice surgeon had participated in over 20 laparoscopic GC surgeries. The medical students had received undergraduate training in medicine, including courses in *anatomy* and *surgery*, and had reviewed more than 50 laparoscopic GC surgery videos (including SL procedures). Consequently, all annotators were highly familiar with the visual characteristics of PM lesions under laparoscopy.

Prior to the annotation process, all annotators underwent systematic and strict training, which included reviewing a substantial number of SL videos of patients with GC, observing live laparoscopic surgeries, identifying PM lesions, and receiving instruction in the use of the LabelMe annotation tool[35]. We chose to use the polygon annotation approach to segment PM lesions. Compared with point annotation or bounding-box annotation, polygon annotation enables more precise delineation of the true shape and irregular boundaries of PM lesions, while reducing the inclusion of surrounding normal peritoneal tissue and other false-positive regions. Ultimately, a detailed annotation protocol was established to define the annotation standards and minimize annotation inconsistency. The annotation protocol is outlined as follows: (1) only annotate clearly visible PM lesions. (2) when multiple PM lesions appear within the same frame, annotators should outline the boundary of each lesion separately. (3) for extensive PM lesions, outline the entire lesion boundary and then separately outline the boundary of the normal peritoneal tissue within the lesion; during segmentation mask generation, exclude the normal tissue to ensure accurate annotation of the lesion boundaries. (4) for extensive PM lesions, when metastases and normal peritoneum within the entire metastatic region are similar in appearance and cannot be reliably distinguished, annotate the entire region together.

During the annotation process, the medical annotators performed the initial pixel-level segmentation annotations. Expert surgeons subsequently reviewed the resulting annotations to ensure strict adherence to the annotation protocol. Annotations that did not meet the required standards were sent back to the medical annotators for re-annotation until full compliance was achieved. The entire annotation process lasted seven months from September 2025 to March 2026. PM lesions with different peritoneal locations and metastatic extents exhibit substantial variations in morphology and color, which contributed to the complexity of the annotation process. Annotation of each frame typically required 3 to 5 minutes, whereas frames containing extensive PM lesions could require more than 10 minutes to complete. In addition, to minimize the influence of lighting conditions and other visual factors in the surgical videos on lesion identification, annotators spent approximately one additional hour reviewing the original video to ensure annotation consistency across consecutive frames.

Finally, expert surgeons performed a thorough check to prevent the misidentification of PM lesions, requiring approximately 2 minutes per frame. During this quality-control process, about 30% of frames required re-annotation and subsequent re-review. As a result, the entire workflow, from frame extraction to final annotation, for each surgical video required approximately 20 hours, making the construction of the dataset both time-consuming and labor-intensive. Overall, the development of the complete dataset required approximately 400 hours, representing a substantial and highly resource-intensive undertaking. Sample annotation frames illustrating PM across various peritoneal locations and metastatic extents are presented in Figure 3.

Furthermore, metastatic lesions were categorized into different classes based on the peritoneal location and the metastatic extent. The peritoneal cavity was delineated into three primary regions: diaphragmatic peritoneum, lateral peritoneum, and pelvic peritoneum. The peritoneum located along the diaphragm and the falciform ligament was defined as the “diaphragmatic peritoneum”. The parietal peritoneum on the flanks near the jejunum, ileum, ascending colon, and descending colon was defined as the “lateral peritoneum”. The peritoneum within the pelvic cavity was defined as the “pelvic peritoneum”. Based on our surgical practice, we divided the metastatic extent into three types: single, multiple, and extensive[29]. A frame containing only one lesion was defined as “single metastasis”. A frame containing multiple, separate lesions was defined as “multiple metastasis”. A frame containing large, confluent lesions was defined as “extensive metastasis”. Expert surgeons manually assigned all labels to the 1013 positive frames. The 907 negative frames were also assigned peritoneal location labels using

the same three-region scheme to indicate the imaged peritoneal region, but metastatic extent labels were not applicable because no visible PM lesion was present.

Data records

The GCPM-SL-20 dataset is stored at Science Data Bank (https://www.scidb.cn/s/bINr6v)[36]. This dataset consists of 20 folders and one CSV file. Each folder was assigned the same unique ID as its corresponding surgical video, ranging from “1” to “20”. The detailed structure is described in Figure 1c and below:

Folders: Each folder contains a “Positive frames” subfolder, which stores the original positive frames extracted from surgical videos in PNG format together with their corresponding masks of metastatic lesions, also in PNG format. The original positive frame filenames consist of the surgery index number and frame sequence number; for example, “1_001.png” denotes the first positive frame extracted from the surgical video of ID 1. The corresponding mask files are named using the same filename with a “_mask” suffix; for instance, the mask corresponding to “1_001.png” is named “1_001_mask.png”. In addition, folders numbered 2, 4, 5, 6, 9, 10, 11, 13, 14, and 15 contain an additional “Negative frames” subfolder, which stores negative frames extracted from the corresponding surgical videos in PNG format. These negative frame filenames include a “_neg” suffix; for example, “2_001_neg.png” denotes the first negative frame extracted from the surgical video of ID 2.

CSV file: The CSV file named “label_summary.csv” provides a detailed record for each frame. This file documents the filename and peritoneal location label for all 1920 frames, as well as the metastatic extent label for the 1013 positive frames.

Technical Validation

Dataset characteristics

The GCPM-SL-20 dataset contains a total of 1920 frames, including 907 negative frames and 1013 positive frames. Figure 2 shows the distribution of positive and negative frames across the videos. Figure 4 illustrates the distribution of different peritoneal locations and metastatic extents across all frames and videos, comprehensively characterizing the patterns of PM lesions. For peritoneal locations (Figure 4a), the diaphragmatic peritoneum, lateral peritoneum, and pelvic peritoneum accounted for 52.61% (533 frames), 27.05% (274 frames), and 20.34% (206 frames) of the positive frames, respectively. For the negative frames, the diaphragmatic peritoneum, lateral peritoneum, and pelvic peritoneum accounted for 46.42% (421 frames), 46.97% (426 frames), and 6.61% (60 frames), respectively. For metastatic extents (Figure 4b), single, multiple, and extensive metastases accounted for 17.37% (176 frames), 68.22% (691 frames), and 14.41% (146 frames) of the positive frames, respectively. Additionally, Figure 4c shows the combined distribution of peritoneal locations and metastatic extents. Furthermore, Figures 4d and 4e show the distribution of peritoneal locations and metastatic extents across the 20 cases. Overall, this detailed characterization of PM heterogeneity in clinical practice provides a solid foundation for developing and evaluating reliable AI models.

Intra- and inter-annotator consistency of PM lesion segmentation

This dataset was annotated by six medical annotators and then reviewed and verified by expert surgeons. To assess the consistency and quality of the annotation, three types of annotator consistency were evaluated: intra-annotator consistency (same annotator at different times), inter-annotator consistency (medical annotator vs. medical annotator), and inter-annotator consistency (medical annotators vs. the expert surgeon). To calculate the three kinds of consistencies, we utilized two classical metrics: Dice coefficient and the normalized 95% Hausdorff Distance (95% HD_{norm}):

- Dice coefficient (Figure 5a) is a region-based score measuring the overlap degree between different segmentations within the range of [0, 1]. A value of 0 denotes no overlap, whereas a value of 1 signifies complete overlap.
- The 95% HD_{norm} (Figure 5b) is a normalized variant of the standard 95% Hausdorff Distance (HD) used to evaluate the dissimilarity between different segmentation masks. The 95% HD calculates the directed

distances from points in one set to their closest counterparts in another set, using the 95th percentile to reduce the influence of outliers. In this study, this metric is normalized by the image diagonal length to obtain the 95% HD_{norm}. It restricts the metric value to a range of 0 to 1, allowing for a consistent evaluation of dissimilarity across different cases. Notably, a value of 0 represents the best result, indicating complete overlap between different segmentation masks, whereas larger values indicate greater dissimilarity between different segmentation masks. A value of 1 indicates that the 95% HD is equal to the image diagonal length, representing the upper limit of the normalized dissimilarity measure.

To assess annotator consistency, a subset of 50 positive frames was randomly selected from the GCPM-SL-20 dataset. This subset included at least one frame from each video, covering all kinds of metastatic extents and peritoneal locations. The detailed results of intra- and inter-annotator consistency are displayed in Table 2.

- For intra-annotator consistency, six medical annotators were required to annotate the 50 positive frames twice. Each annotation session was completed within one week, and the interval between the two sessions was at least two weeks. The Dice coefficients for different metastatic extents and peritoneal locations ranged from 0.90 (±0.09) to 0.94 (±0.05), indicating the high reproducibility of our labels and the reliability of the annotation process. Significantly, 95% HD_{norm} values between 0.03 (±0.07) and 0.06 (±0.12) also firmly supported the intra-annotator reliability of our dataset.
- For inter-annotator consistency (medical annotators), the segmentation masks provided by all annotators for the 50 frames were compared. The Dice coefficients for different metastatic extents and peritoneal locations ranged from 0.82 (±0.09) to 0.92 (±0.04), indicating high inter-annotator consistency. Correspondingly, the 95% HD_{norm} values, ranging from 0.04 (±0.08) to 0.06 (±0.14), further demonstrated the satisfactory consistency of the dataset.
- For inter-annotator consistency (medical annotators vs. the expert surgeon), the segmentations provided by the medical annotators were compared with those of the expert surgeon across the 50 frames. The mean Dice coefficient of all frames was 0.78 (±0.20). For different metastatic extents, the Dice coefficients ranged from 0.63 (±0.37) to 0.85 (±0.12); for different peritoneal locations, the Dice coefficients ranged from 0.76 (±0.23) to 0.83 (±0.13). The annotation consistency for extensive metastases was only 0.63, below the ideal threshold of 0.75 and lower than that for single and multiple metastases, indicating less satisfactory inter-annotator agreement. This was primarily because extensive metastases required the delineation of large regions containing multiple confluent lesions, rather than outlining individual lesion boundaries as in the single and multiple metastases. This increased subjectivity made accurate annotation more challenging, highlighting the necessity of the expert review process in this study. It also underscores the substantial clinical and research value of the publicly available PM surgical imaging dataset containing extensive metastatic cases.

In conclusion, the analysis of annotator consistency for the selected subset demonstrated a high level of agreement, which provides a solid foundation for the reproducible segmentation and accurate annotation of metastatic lesions.

Comparison of deep learning models for PM lesion segmentation

For the comparison of deep learning models, we applied a five-fold cross-validation strategy on 1920 frames corresponding to 20 cases. In each fold, 16 cases (80%) were assigned to the training set, while the remaining 4 cases (20%) were used for validation. To prevent overfitting induced by data leakage, each fold ensured that frames from the same patient are exclusively allocated to either the training set or the validation set. Figure 6 shows the five-fold cross-validation strategy and the label distribution in the training and validation sets for each fold. To evaluate the segmentation performance, we used four evaluation metrics: the Dice coefficient, Intersection over Union (IoU), Recall, and the 95% HD_{norm}. For negative frames with empty ground-truth masks, a prediction was considered correct only when no lesion pixels were predicted; false-positive lesion predictions on negative frames were counted as segmentation errors. The specific formulas and visual illustrations for these metrics are provided in Figure 5.

In order to provide a broad performance baseline, we selected CNN-based architectures including nnU-Netv2[37], Pyramid Scene Parsing Network (PSPNet)[38], DeepLabV3+[39], SegNet[40]; Transformer-based architectures including SEgmentation Transformer (SETR)[41], SegFormer[42] and Swin Transformer[43]; zero-shot Segment Anything Model (SAM)[44]. The architectures of selected models are presented in Figure 7. To prevent overfitting, we expanded the dataset size by five times using random data augmentations including geometric transformations and photometric adjustments. All models except for SAM were trained on an NVIDIA A30 GPU for 20 epochs with the input size as 480×480 pixels. We used cross-entropy loss and Adam optimizer during training process. The batch size and the learning rate were set to 48 and 0.001, respectively. The best checkpoint was saved after training for evaluation. Specifically, SAM was evaluated in a zero-shot setting without any training or parameter updates; for each positive frame, one point prompt was randomly sampled from the ground-truth PM lesion region to guide lesion segmentation.

The five-fold cross-validation evaluation results on the GCPM-SL-20 dataset are summarized in Table 3. Figure 8 presents the visual results of the PM lesion segmentation. Overall, Transformer-based models consistently outperformed CNN-based models across most metrics. Swin Transformer achieves the best overall segmentation performance, with the highest Dice coefficient (0.55), IoU (0.42), and the lowest 95% HD_{norm} (0.19). Among CNN-based architectures, the nnU-Netv2 model demonstrates the strongest performance, achieving a Dice coefficient of 0.49, an IoU of 0.38, and a Recall of 0.47. Notably, despite exhibiting strong zero-shot generalization capability on natural image datasets, SAM demonstrates suboptimal performance on this specific dataset (Dice coefficient of 0.28, IoU of 0.21), even with a relatively high Recall of 0.60. Tables 4 and 5 provide detailed performance comparisons across varying metastatic extents and peritoneal locations. Among all extents, cases with extensive metastases achieve the best Dice coefficient for all models except nnU-Netv2 and SAM. This is because larger and continuous lesions provide stronger contrasts against surrounding tissues. Among lesions with different metastatic extents, multiple metastases show the least favorable performance, likely because these lesions are small and scattered, making it difficult for the model to identify every lesion and leading to missed detections. Regarding peritoneal location, lesions on the lateral peritoneum demonstrate the best segmentation performance across nearly all models. Their Dice coefficients and IoU values are typically 0.1 to 0.2 higher than those of pelvic peritoneum lesions. The pelvic peritoneum remains the most challenging location, yielding the lowest evaluation metrics. This may be explained by the location of the pelvic peritoneum in a dependent and relatively flat anatomical region, where physiologic peritoneal fluid and intraoperative irrigation fluid tend to accumulate under gravity. Under laparoscopic illumination, the light source often strikes this fluid-covered surface nearly perpendicularly, generating specular reflections and producing a bright whitish appearance. These reflective areas may mimic PM lesions and increase the risk of false-positive interpretation by the models.

Overall, this comparison provides a reliable performance baseline for the GCPM-SL-20 dataset. We find that the Swin Transformer achieved better performance than the other models in this task, although substantial room for improvement remains. This further indicates the need to develop more effective AI models tailored to this specific clinical scenario. Furthermore, the results clearly show that lesion extent and location significantly affect model accuracy, emphasizing the complex nature of this PM lesion segmentation challenge. Therefore, constructing and publicly releasing a segmentation dataset of PM lesions with varying metastatic extents and peritoneal locations is of substantial importance for the development of AI-assisted intraoperative identification technologies for metastatic lesions during SL for GC.

Advantages and limitations of the dataset

Currently, publicly available surgical datasets mainly focus on anatomical structures, instruments, phases, and actions in laparoscopic or robotic surgery[33], while open-source datasets with intraoperative metastatic lesion annotations remain scarce, particularly for GC-related PM during SL. Although previous studies have explored AI-assisted PM detection, segmentation, classification, and size estimation, their datasets have not been publicly released[22, 29, 30, 32]. This GCPM-SL-20 dataset addresses this gap by providing, to our knowledge, the first open-source pixel-level segmentation dataset for PM identification during SL in patients with GC. The dataset includes 1920 high-resolution frames from 20 complete SL videos, comprising 1013 pathologically confirmed positive

frames with PM lesion masks and 907 negative frames. The main advantage of this dataset lies in its integration of pixel-level lesion segmentation masks, metastatic extent labels, peritoneal location labels, and visually similar negative frames, which together support SDS research, surgical education, benchmark development, and the development of AI-assisted intraoperative PM diagnosis systems.

However, several limitations should be acknowledged. First, frames with severe blur, smoke, or glare were excluded as PM lesions could not accurately identified and annotated under these conditions, which may introduce selection bias. Second, the distribution of peritoneal locations and metastatic extents is imbalanced, although this partly reflects real clinical heterogeneity. Third, all videos were collected from a single center using the same laparoscopic system, which may limit generalizability. Future work should incorporate multicenter and multi-device data, as well as data augmentation[45] or class-balanced learning strategies[46], to improve dataset diversity and model robustness.

Data Availability

The GCPM-SL-20 dataset is publicly available at Science Data Bank (https://www.scidb.cn/s/bINr6v). The repository contains the original PNG frames, binary lesion masks, and the label summary CSV file. The dataset is released under CC-BY 4.0, and no additional patient-identifying information is included.

Code Availability

The code for all models in this study is available on GitHub. The links are as follows: PSPNet and DeepLabV3+ are hosted at https://github.com/Tramac/awesome-semantic-segmentation-pytorch; SegNet is hosted at https://github.com/Deeachain/Segmentation-Pytorch; nnU-Netv2 is hosted at https://github.com/MIC-DKFZ/nnUNet; SegFormer is hosted at https://github.com/NVlabs/SegFormer; SETR is hosted at https://github.com/fudan-zvg/SETR; Swin Transformer is hosted at https://github.com/SwinTransformer/Swin-Transformer-Semantic-Segmentation; SAM is hosted at https://github.com/facebookresearch/segment-anything.

Funding

This work was financially supported by the Science and Technology Program of Guangzhou (grant no.2025A04J4017) and Special Funds for the Cultivation of Guangdong College Students' Scientific and Technological Innovation (grant no.pdjh2024a086).

Author Contributions

L.G., H.C., H.W., Y.H., and J.Y. conceived the study, led the design and construction of the GCPM-SL-20 dataset, developed the annotation framework, coordinated the annotation workflow, supervised dataset curation, and oversaw manuscript preparation. H.C., Y.H. and J.Y. contributed to clinical data collection, surgical video review, frame extraction, expert lesion identification, annotation verification, and dataset quality control. L.G., S.P., Y.L., Y.Z., J.C. and Z.Y. organized the extracted frames, performed pixel-level lesion annotation, curated dataset files, prepared visualizations, and drafted the manuscript. Y.L. and H.Y.C. implemented the segmentation models, conducted the five-fold cross-validation experiments, and contributed to experimental analysis and figure/table preparation. H.W. supervised the algorithmic design, model benchmarking, technical validation, and statistical analysis. Y.W., M.H., C.L. and C.P. assisted with data organization, annotation documentation, reference checking, figure/table preparation, and manuscript editing. H.C., H.W., Y.H., J.Y. and L.G. provided senior supervision, interpreted the results, critically reviewed and revised the manuscript, and approved the final presentation of the dataset and manuscript. All authors approved the submitted version of the manuscript and agreed to be accountable for their own contributions and for the integrity of the work. S.P., Y.L., Y.Z., J.C. and Z.Y. contributed equally to this work. L.G., H.C., H.W., Y.H. and J.Y. jointly supervised this work and served as corresponding authors.

Competing Interests

The authors declare no competing interests.

References

1. The global, regional, and national burden of cancer, 1990-2023, with forecasts to 2050: a systematic analysis for the Global Burden of Disease Study 2023. *Lancet* **406**, 1565-1586 (2025).

2. Cortés-Guiral D.*, et al.* Primary and metastatic peritoneal surface malignancies. *Nat Rev Dis Primers* **7**, 91 (2021).

3. Thomassen I.*, et al.* Peritoneal carcinomatosis of gastric origin: a population-based study on incidence, survival and risk factors. *Int J Cancer* **134**, 622-628 (2013).

4. Butensky S.D.*, et al.* Consensus guideline for the management of gastric cancer with synchronous peritoneal metastases. *Cancer* **131**, e35870 (2025).

5. Bonnot P.E.*, et al.* Prognosis of poorly cohesive gastric cancer after complete cytoreductive surgery with or without hyperthermic intraperitoneal chemotherapy (CYTO-CHIP study). *Br J Surg* **108**, 1225-1235 (2021).

6. Hu Y.F.*, et al.* Staging laparoscopy improves treatment decision-making for advanced gastric cancer. *World J Gastroenterol* **22**, 1859-1868 (2016).

7. Guchelaar N.A.D.*, et al.* The prognostic value of peritoneal metastases in patients with gastric cancer: a nationwide population-based study. *EClinicalMedicine* **81**, 103109 (2025).

8. Ajani J.A.*, et al.* Gastric Cancer, Version 2.2025, NCCN Clinical Practice Guidelines In Oncology. *Journal of the National Comprehensive Cancer Network : JNCCN* **23**, 169-191 (2025).

9. Fukagawa T. Role of staging laparoscopy for gastric cancer patients. *Ann Gastroenterol Surg* **3**, 496-505 (2019).

10. van 't Sant I.*, et al.* Diagnostic performance of imaging for the detection of peritoneal metastases: a meta-analysis. *Eur Radiol* **30**, 3101-3112 (2020).

11. Japanese gastric cancer treatment guidelines 2025 (7th edition). *Gastric Cancer* **29**, 271-299 (2026).

12. Lordick F.*, et al.* Gastric cancer: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up. *Ann Oncol* **33**, 1005-1020 (2022).

13. Shitara K.*, et al.* Pan-Asian adapted ESMO Clinical Practice Guidelines for the diagnosis, treatment and follow-up of patients with gastric cancer. *ESMO Open* **9**, 102226 (2024).

14. Schena C.A.*, et al.* The Role of Staging Laparoscopy for Gastric Cancer Patients: Current Evidence and Future Perspectives. *Cancers (Basel)* **15**, (2023).

15. Kim C.Y.*, et al.* Learning curve for gastric cancer surgery based on actual survival. *Gastric Cancer* **19**, 631-638 (2016).

16. Ramos R.F., Scalon F.M., Scalon M.M., Dias D.I. Staging laparoscopy in gastric cancer to detect peritoneal metastases: A systematic review and meta-analysis. *Eur J Surg Oncol* **42**, 1315-1321 (2016).

17. Rawicz-Pruszyński K.*, et al.* Techniques of staging laparoscopy and peritoneal fluid assessment in gastric cancer: a systematic review. *Int J Surg* **109**, 3578-3589 (2023).

18. Schnelldorfer T., Ware M.P., Liu L.P., Sarr M.G., Birkett D.H., Ruthazer R. Can We Accurately Identify Peritoneal Metastases Based on Their Appearance? An Assessment of the Current Practice of Intraoperative Gastrointestinal Cancer Staging. *Ann Surg Oncol* **26**, 1795-1804 (2019).

19. Suliburk J.W.*, et al.* Analysis of Human Performance Deficiencies Associated With Surgical Adverse Events. *JAMA Netw Open* **2**, e198067 (2019).

20. Maier-Hein L.*, et al.* Surgical data science - from concepts toward clinical translation. *Med Image Anal* **76**,

102306 (2022).

21. Maier-Hein L.*, et al.* Surgical data science for next-generation interventions. *Nat Biomed Eng* **1**, 691-696 (2017).

22. Tozzi F.*, et al.* Multimodal machine learning for staging laparoscopy: a combined image analysis and morphologic tool for the discrimination of peritoneal metastasis. *International Journal of Surgery (London, England)* **112**, 373-383 (2026).

23. Ortenzi M.*, et al.* A novel high accuracy model for automatic surgical workflow recognition using artificial intelligence in laparoscopic totally extraperitoneal inguinal hernia repair (TEP). *Surg Endosc* **37**, 8818-8828 (2023).

24. Yang J.H.*, et al.* Using AI and computer vision to analyze technical proficiency in robotic surgery. *Surg Endosc* **37**, 3010-3017 (2023).

25. Madad Zadeh S.*, et al.* SurgAI: deep learning for computerized laparoscopic image understanding in gynaecology. *Surg Endosc* **34**, 5377-5383 (2020).

26. Maier-Hein L.*, et al.* Heidelberg colorectal data set for surgical data science in the sensor operating room. *Sci Data* **8**, 101 (2021).

27. Ghamsarian N.*, et al.* Cataract-1K Dataset for Deep-Learning-Assisted Analysis of Cataract Surgery Videos. *Sci Data* **11**, 373 (2024).

28. Ye Z.*, et al.* A Comprehensive Video Dataset for Surgical Laparoscopic Action Analysis. *Sci Data* **12**, 862 (2025).

29. Chen H.*, et al.* Artificial intelligence assisted real-time recognition of intra-abdominal metastasis during laparoscopic gastric cancer surgery. *NPJ Digit Med* **8**, 9 (2025).

30. Schnelldorfer T., Castro J., Goldar-Najafi A., Liu L. Development of a Deep Learning System for Intraoperative Identification of Cancer Metastases. *Ann Surg* **280**, 1006-1013 (2024).

31. Ma X.*, et al.* A pioneering artificial intelligence tool to predict treatment outcomes in ovarian cancer via diagnostic laparoscopy. *Sci Rep* **15**, 14437 (2025).

32. Mousavi S.A.*, et al.* A Reference-Based Approach for Tumor Size Estimation in Monocular Laparoscopic Videos. In: *Computational Mathematics Modeling in Cancer Analysis: Third International Workshop, CMMCA 2024, Marrakesh, Morocco, October 6, 2024, Proceedings*). Springer-Verlag (2024).

33. Gou L.*, et al.* Dynamic key vascular anatomy dataset for D2 lymph node dissection during laparoscopic gastric cancer surgery. *Sci Data* **12**, 903 (2025).

34. Oliva R.*, et al.* Artificial Intelligence for the Assessment of Peritoneal Carcinosis during Diagnostic Laparoscopy for Advanced Ovarian Cancer. *arXiv*, (2025).

35. Russell B.C., Torralba A., Murphy K.P., Freeman W.T. LabelMe: A Database and Web-Based Tool for Image Annotation. *International Journal of Computer Vision* **77**, 157-173 (2008).

36. Peng S.*, et al.* GCPM-SL-20: a pixel-level segmentation dataset with location and extent labels for peritoneal metastasis during staging laparoscopy for gastric cancer. Science Data Bank (2026).

37. Isensee F., Jaeger P.F., Kohl S.A.A., Petersen J., Maier-Hein K.H. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods* **18**, 203-211 (2021).

38. Zhao H., Shi J., Qi X., Wang X., Jia J. Pyramid Scene Parsing Network. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*). IEEE (2017).

39. Chen L., Zhu Y., Papandreou G., Schroff F., Adam H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In: *15th European Conference on Computer Vision (ECCV)*) (2018).

40. Badrinarayanan V., Kendall A., Cipolla R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Trans Pattern Anal Mach Intell* **39**, 2481-2495 (2017).

41. Zheng S.*, et al.* Rethinking Semantic Segmentation From a Sequence-to-Sequence Perspective With Transformers. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE (2021).

42. Xie E., Wang W., Yu Z., Anandkumar A., Alvarez J.M., Luo P. SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers. In: *Advances in Neural Information Processing Systems*. Curran Associates, Inc. (2021).

43. Liu Z.*, et al.* Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE (2021).

44. Kirillov A.*, et al.* Segment Anything. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE (2023).

45. Wang Y., Huang G., Song S., Pan X., Xia Y., Wu C. Regularizing Deep Networks With Semantic Data Augmentation. *IEEE Trans Pattern Anal Mach Intell* **44**, 3733-3748 (2022).

46. Fernando K.R.M., Tsokos C.P. Dynamically Weighted Balanced Loss: Class Imbalanced Learning and Confidence Calibration of Deep Neural Networks. *IEEE Trans Neural Netw Learn Syst* **33**, 2940-2951 (2022).

Figure Legends

Figure 1. Overview of the data acquisition and annotation process. (a) Data collection and frame extraction: raw data were collected from 20 surgical videos of staging laparoscopy procedures for patients with gastric cancer, from which positive and negative frames were extracted by expert surgeons. **(b) Data annotation:** positive frames were assigned both peritoneal location and metastatic extent labels by expert surgeons, whereas negative frames were assigned peritoneal location labels only; the 1013 positive frames underwent pixel-level annotation by six medical annotators under the supervision of three expert surgeons. **(c) Data records:** the complete dataset is organized into twenty folders and one CSV file. Each folder corresponds to a single surgical video and contains a “Positive frames” subfolder, including positive frames and their corresponding PM lesion masks. Folders numbered 2, 4, 5, 6, 9, 10, 11, 13, 14, and 15 also contain a “Negative frames” subfolder, including negative frames extracted from the corresponding surgical video. A “label_summary.csv” file stores the filename and peritoneal location label for all 1920 frames, together with metastatic extent labels for the 1013 positive frames. PM: peritoneal metastasis. GCPM-SL-20: Gastric Cancer Peritoneal Metastasis Staging Laparoscopy dataset; SMU-NF-GC-02 Dataset: Southern Medical University - Nanfang Hospital - Gastric Cancer - 02 Dataset.

Figure 2. Video lengths and frame distribution in the dataset. (a) Lengths of the original and edited videos. Original videos include all staging laparoscopy steps; the edited videos exclude steps including peritoneal cytology, lesion excision biopsy, and closure of the abdominal wall incision, with only intra-abdominal exploration of different peritoneal locations retained. **(b)** Proportional composition of positive and negative frames within the entire dataset. **(c)** Histogram showing the distribution of positive and negative frames across 20 videos.

Figure 3. Sample annotation frames of PM lesions with different metastatic extents and peritoneal locations. This figure presents raw frames of PM lesions at varying metastatic extents and peritoneal locations during staging laparoscopy, with corresponding pixel-level segmentation masks and mask-overlay visualizations. PM: peritoneal metastasis.

Figure 4. Distribution of frames by peritoneal location and extent in the dataset and individual cases. **(a)** Distribution of “diaphragmatic peritoneum”, “lateral peritoneum”, and “pelvic peritoneum” labels in the positive

frames. **(b)** Distribution of “single”, “multiple”, and “extensive” labels in the positive frames. **(c)** Distribution of metastatic extent labels stratified by peritoneal location. **(d)** Distribution of “diaphragmatic peritoneum”, “lateral peritoneum”, and “pelvic peritoneum” labels across the positive frames for each video. **(e)** Distribution of “single”, “multiple”, and “extensive” labels across the positive frames for each video.

Figure 5. Evaluation metrics. **(a)** The graphic illustrations and formulas of Dice coefficient, IoU (Intersection over Union), and Recall. **(b)** The graphic illustrations and formulas of 95% HD_{norm}. “G” represents the ground-truth segmentation mask of a frame while “P” denotes the prediction result of this frame. “TP”, “FP”, “FN” refer to pixels of “true positive”, “false positive” and “false negative”, respectively. The symbol “$\cap$” and “$\cup$” denote the intersection and union, respectively, of the ground truth (G) and the prediction result (P). “d_{GP}” refers to the minimum Euclidean distance from a point $g \in G$ to the set P. “d_{PG}” refers to the minimum Euclidean distance from a point $p \in P$ to the set G.

Figure 6. Five-fold cross-validation setup and label distribution for PM lesion segmentation model benchmarking. **(a)** Five-fold cross-validation strategy for PM lesion segmentation. Gray and orange boxes denote the training and validation sets, respectively. **(b)** Distribution of metastatic extent and location labels in the training and validation sets across folds. PM: peritoneal metastasis.

Figure 7. Overview of the architectures of the segmentation models evaluated in this study. PSPNet: Pyramid Scene Parsing Network; SETR: SEgmentation Transformer; SAM: Segment Anything Model.

Figure 8. Comparative visualization of semantic segmentation models for PM lesion identification across different metastatic extents and peritoneal locations. PM: Peritoneal Metastasis; GT: Ground truth; PSPNet: Pyramid Scene Parsing Network; SETR: SEgmentation Transformer; SAM: Segment Anything Model.

Tables

Table 1. Comparison of current research on PM in surgical frames and the proposed GCPM-SL-20 Dataset.

Study title	Study objective	Year	Annotation approach[a]	Data labeling[b]	Frame count	Publication	Data availability
Development of a deep learning system for intraoperative identification of cancer metastases[30]	Gastrointestinal cancer PM lesion identification	2024	Bounding box	Biopsy results	3650	Annals of Surgery	Unavailable
A reference-based approach for tumor size estimation in monocular laparoscopic videos[32]	Abdominal malignancy PM lesion size estimation[c]	2024	Polygon (segmentation)	PM lesion size	2309	CMMCA 2024 Proceedings	Unavailable
A pioneering artificial intelligence tool to predict treatment outcomes in ovarian cancer via diagnostic laparoscopy[31]	Ovarian cancer survival prediction	2025	None	Survial outcomes of patients	3169	Scientific reports	Unavailable
Artificial intelligence assisted real-time recognition of intra-abdominal metastasis during laparoscopic gastric cancer surgery[29] (Ours)	Gastric cancer intra-abdominal metastasis lesion identification	2025	Polygon (segmentation)	Metastatic extents and locations	5111	npj Digital Medicine	Unavailable

Artificial intelligence for the assessment of peritoneal carcinosis during diagnostic laparoscopy for advanced ovarian cancer[34]	Ovarian Cancer PM lesion identification and scoring	2025	Polygon (segmentation)	Anatomical location and Fagotti score	7311	arXiv (unpublished)	Unavailable
Multimodal machine learning for staging laparoscopy: a combined image analysis and morphologic tool for the discrimination of peritoneal metastasis[22]	Abdominal malignancy peritoneal lesion Benign–malignant classification[d]	2026	None	Biopsy results and morphologic features	5214	International Journal of Surgery	Unavailable
GCPM-SL-20: a pixel-level segmentation dataset with location and extent labels for peritoneal metastasis during staging laparoscopy for gastric cancer (Ours)	**Gastric cancer PM lesion identification**	**2026**	**Polygon (segmentation)**	**Metastatic extents and peritoneal locations**	**1920**	**Scientific Data**	**Available**

[a]Annotation approach refers to the annotation of specific visual features or regions of interest within each frame;

[b]Data labeling refers to assigning a single overall label to the entire frame;

[c]Abdominal malignancy: not specified;

[d]Abdominal malignancy: gastric cancer, colorectal cancer, ovarian cancer, pancreatic cancer, appendiceal cancer, cholangiocarcinoma, hepatocellular carcinoma, esophageal cancer, and other malignancies;

GCPM-SL-20 dataset: Gastric Cancer Peritoneal Metastasis Staging Laparoscopy dataset; PM: peritoneal metastasis; CMMCA: Computational Mathematics Modeling in Cancer Analysis.

Table 2. Intra- and inter-annotator consistency metrics for PM lesion segmentation.

Category	**Intra-annotator consistency (annotator)**		**Inter-annotator consistency (annotator vs. annotator)**		**Inter-annotator consistency (annotators vs. expert surgeon)**	
	Dice coefficient	**95% HD_{norm}**	**Dice coefficient**	**95% HD_{norm}**	**Dice coefficient**	**95% HD_{norm}**
Overall	0.91 (0.08)	0.03 (0.07)	0.85 (0.09)	0.05 (0.11)	0.78 (0.20)	0.08 (0.12)
Metastatic extent						
Single	0.93 (0.06)	0.03 (0.09)	0.86 (0.07)	0.06 (0.14)	0.85 (0.12)	0.04 (0.08)
Multiple	0.90 (0.09)	0.04 (0.11)	0.82 (0.09)	0.05 (0.10)	0.79 (0.10)	0.09 (0.12)
Extensive	0.94 (0.05)	0.06 (0.05)	0.92 (0.04)	0.04 (0.03)	0.63 (0.37)	0.14 (0.16)
Peritoneal location						
Diaphragmatic peritoneum	0.91 (0.08)	0.03 (0.07)	0.84 (0.11)	0.05 (0.11)	0.83 (0.13)	0.09 (0.14)
Lateral peritoneum	0.92 (0.07)	0.04 (0.10)	0.84 (0.07)	0.04 (0.08)	0.76 (0.24)	0.08 (0.11)
Pelvic peritoneum	0.90 (0.09)	0.06 (0.12)	0.86 (0.07)	0.06 (0.14)	0.76 (0.23)	0.06 (0.12)

PM: peritoneal metastasis. 95% HD_{norm}: The normalized 95% Hausdorff Distance. Values are reported as mean (standard deviation).

CONFIDENTIAL CONTENT REDACTED

稿件尚未接收。为保护在审项目，仅对这一处核心内容作集中遮挡，敬请谅解。

Protected scope: Detailed model benchmarking tables (Tables 3-5)