scientific data

Check for updates

OPEN

DATA DESCRIPTOR

A Comprehensive Video Dataset for Surgical Laparoscopic Action Analysis

Zi Ye[1,2,6], Ru Zhou[1,6], Zili Deng[3], Dan Wang[3], Ying Zhu[3], Xiaoli Jin[1], Lijun Zhang[4], Tianxiang Chen[5], Hanwei Zhang[2] ✉ & Mingliang Wang[1]

Laparoscopic surgery has been widely used in various surgical fields due to its minimally invasive and rapid recovery benefits. However, it demands a high level of technical expertise from surgeons. While advancements in computer vision and deep learning have significantly contributed to surgical action recognition, the effectiveness of these technologies is hindered by the limitations of existing publicly available datasets, such as their small scale, high homogeneity, and inconsistent labeling quality. To address the above issues, we developed the SLAM dataset (Surgical LAparoscopic Motions), which encompasses various surgical types such as laparoscopic cholecystectomy and appendectomy. The dataset includes annotations for seven key actions: *Abdominal Entry*, *Use Clip*, *Hook Cut*, *Suturing*, *Panoramic View*, *Local Panoramic View*, and *Suction*. In total, it includes 4,097 video clips, each labeled with corresponding action categories. In addition, we comprehensively validated the dataset using the ViViT model, and the experimental results showed that the dataset exhibited superior training and testing capabilities in laparoscopic surgical action recognition, with the highest classification accuracy of 85.90%. As a publicly available benchmark resource, the SLAM dataset aims to promote the development of laparoscopic surgical action recognition and artificial intelligence-driven surgery, supporting intelligent surgical robots and surgical automation.

Background & Summary

Due to its minimally invasive nature, laparoscopic surgery is widely used in multiple surgical specialties, including gastrointestinal, gynecological, and urological fields. Using slender rod-shaped instruments and a camera inserted through small incisions, laparoscopic procedures allow precise manipulation and visualization of internal organs in real-time[1]. Compared to traditional open surgery, laparoscopic surgery provides numerous benefits, including faster recovery times, less postoperative pain, and shorter hospital stays[2]. However, the practice of laparoscopic surgery requires a high level of skill due to the limited operating space, the need for precision, and the limited visual feedback[3].

Current research in computer vision and deep learning within the medical field demonstrates significant potential for automatic recognition and classification of laparoscopic surgical actions[4]. These technologies can improve the recognition and analysis of surgical actions, such as clamping tissue, cutting, and suturing, during procedures, thus improving intraoperative navigation, surgical training, and postoperative evaluation[5,6]. Artificial intelligence (AI)-driven surgical action recognition can offer surgeons real-time assistance, optimize surgical workflows, generate automated surgical reports, and assess operator skill levels[7]. Furthermore, action classification models are pivotal for intelligent surgical robots, enhancing surgical automation, reducing operating times, and minimizing the risk of complications[8]. Research in laparoscopic surgical action recognition serves as an essential and fundamental step toward advancing AI-driven surgical robotics on a larger scale[9,10].

Today, datasets are essential for the advancement of AI research[11]. While basic action recognition in surgery has been studies, its real-world applicability remains limited due to technical challenges, particularly the lack

[1]Department of General Surgery, RuiJin Hospital LuWan Branch, Shanghai Jiaotong University School of Medicine, Shanghai, 200020, China. [2]Institute of Intelligent Software, Guangzhou, Guangdong, 511400, China. [3]Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences, Zhejiang, 310024, China. [4]Institute of Software, Chinese Academy of Sciences, Beijing, 100190, China. [5]School of Cyber Space and Technology, University of Science and Technology of China, Anhui, 230026, China. [6]These authors contributed equally: Zi Ye, Ru Zhou. ✉e-mail: zhanghanwei0912@gmail.com

Dataset	Year	Size	Surgical Procedure	Action Categories	Ref
LapGyn4	2018	30,000+ images	Laparoscopic Gynecologic Surgery	8	[13]
ESAD	2021	16 hours	Radical Prostatectomy	21	[14]
PSI-AVA	2022	20.45 hours	Robot-assisted Radical Prostatectomy	16	[15]
HeiChole	2022	22 hours	Laparoscopic Cholecystectomy	4	[16]
Cholect50	2023	50 videos	Laparoscopic Cholecystectomy	10	[17,18]
LapGyn6-Actions	2023	18 videos	Gynecology laparoscopy	6	[19]
GraSP	2024	32 hours	Robot-Assisted Radical Prostatectomy	14	[15]
SLAM(Ours)	2024	4097 videos	Multiple laparoscopic surgeries	7	

Table 1. Publicly Accessible Dataset on Laparoscopic Surgical Maneuvers.

of diverse, accurate, and comprehensive dataets. In recognition of laparoscopic surgical action, several publicly available datasets have been developed with varying aims to support this vital work[12], as shown in Table 1. LapGyn4[13] provides a rich image resource by offering over 30, 000 images of 111 gynecological laparoscopic surgeries with eight common surgical actions. ESAD[14], designed for minimally invasive endoscopic surgery, serves as a crucial benchmark for surgical action recognition, containing 16 hours of radical prostatectomy (RARP) videos annotated with 46, 325 action instances across 21 categories. PSI-AVA[15], the first comprehensive dataset for robotic-assisted RARP surgery, supports long-term activity recognition, short-term reasoning tasks, and the identification of new atomic actions, featuring 11 phases, 20 steps, seven instrument types, and 16 atomic actions. HeiChole[16] is a multicenter resource focused on surgical action recognition, containing 33 laparoscopic cholecystectomy videos with a total duration of 22 hours, annotated with seven surgical phases, four maneuvers, seven types of surgical instruments, and five skill dimensions. CholecT50[17,18] contains 50 laparoscopic cholecystectomy videos totaling approximately 100, 900 frames labeled with 100 surgical action triad categories, but the test set is not publicly available. LapGyn6-Actions[19] aims to improve surgical action recognition research with 18 randomly selected gynecologic laparoscopic video clips labeled with six categories of surgical actions. GraSP[15] focuses on human-computer collaboration in both long-term and short-term surgical tasks, comprising 32 hours of radical prostatectomy videos annotated with 14 atomic action categories, totaling over 9, 031 instances.

While these datasets offer significant value within their specific domains, they generally have several limitations: 1) limited size, as most datasets feature relatively small sample sizes, making them insufficient for training deep learning models that require large volumes of data; 2) homogeneity, as many datasets focus on a specific surgical type, lacking diversity and comprehensiveness; 3) restricted publicity, with some datasets having unpublished test sets, which limits their potential as benchmark datasets[17,18]; and 4) labeling quality, as some datasets may have incomplete or biased annotations, which can impact the generalization ability of models. To overcome these limitations, we introduce a new video dataset named SLAM[20], which stands for Surgical LAparoscopic Motion, providing a comprehensive resource to advance research in laparoscopic surgery. Our dataset encompasses various types of surgeries, such as cholecystectomy, appendectomy, radical resection of gastric stromal tumors, etc., enhancing the diversity of laparoscopic procedures and supporting more consistent surgical action recognition across different surgery types. Our current focus is on foundational actions (e.g., Use Clip, Hook Cut) within the overall workflow. This serves as a crucial initial step in breaking down procedures into detailed steps, facilitating AI-assisted surgical procedures.

Methods

This study was approved by the Ethics Committee of Ruijin Hospital, Luwan Branch, Shanghai Jiao Tong University School of Medicine (Approval Number: LWEC2024022), which explicitly waived the requirement for individual consent, permitting open sharing and publication of the surgical video dataset without restrictions. The study adheres to the principles outlined in the Declaration of Helsinki, and all applicable regulations for clinical research were followed. The strict adherence to ethical standards and privacy regulations during the collection and processing of the dataset ensured the confidentiality and integrity of patient information[21]. We thoroughly anonymize the dataset during collection to protect patient privacy and remove or mask all personally identifiable information (PII) to prevent data from linking to individual patients. In addition, we have included only the essential surgical scenario data to minimize the potential to disclose sensitive personal information.

The general demographic information of the patients, including variables such as sex, age, and BMI, is provided in Table 2. We specifically noted whether the patient underwent a single procedure or multiple combined procedures in the *Surgical Combination Classification*, along with the total count of surgery types. We carefully selected this information to balance scientific utility with privacy protection, essentially for research integrity. The dataset does not contain identifying details such as names, dates of birth, or other unique identifiers that could compromise patient confidentiality, which protects the high level of privacy of the dataset. The dataset also contains valuable information for medical research, especially for motion recognition and surgical action analysis.

Our dataset comprises videos of various common laparoscopic surgical procedures, which have been thoroughly studied and annotated to provide high-quality training and validation data for action recognition tasks. Sample frames, as shown in Fig. 1, are provided as an illustrative example. We focus on seven fundamental actions commonly performed across most surgeries according to medical doctors: *Abdominal Entry*, *Suction*, *Use Clip*, *Hook Cut*, *Suturing*, *Panoramic View*, and *Local Panoramic View*. The actions selected for annotation in this dataset were chosen based on their clinical relevance and practical significance in laparoscopic surgery. These seven actions are fundamental and play a key role in ensuring surgical safety and performance assessment[22].

Age (years)	53 ± 32 (21~78)
BMI (kg/m^2)	23.6 ± 6.0
Sex	
Female	17 (50%)
Male	17 (50%)
Surgical Combination Classification	
Single Procedure	32
Combined Procedures	2
Types of Surgery	
Cholecystectomy	22
Appendectomy	6
Resection of Gastric Stromal Tumor	2
Abdominal Wall Hernia Repair	1
Sigmoid Colectomy	1
Radical Resection of Rectal Cancer	1
Radical Resection of Sigmoid Colon Cancer	1
Splenectomy	1
VATS for Lung Surgery	1

Table 2. Patient Profile in SLAM Dataset.

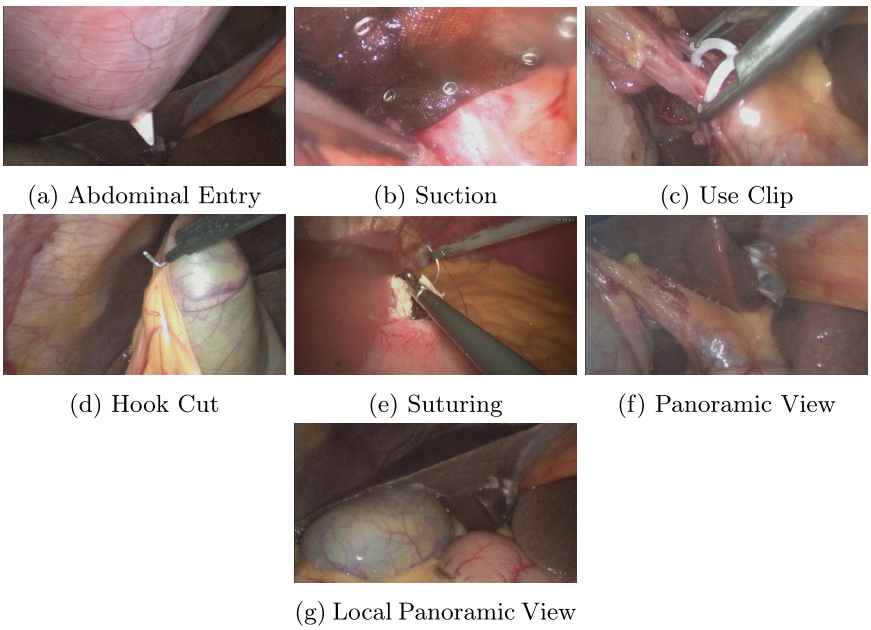

Fig. 1 Example Frames of Seven Fundamental Actions in Laparoscopic Surgery.

Among them, *Abdominal Entry* is the first step in laparoscopic surgery, where instruments initially enter the abdominal cavity. This critical procedure sets the stage for the operation and is essential for the surgery's overall success, holding significant clinical value. *Abdominal Entry* is a universal step in all laparoscopic procedures, with its technique being essential for minimizing the risk of injury to surrounding organs. Proper execution is crucial for ensuring adequate instrument access and preventing complications such as inadvertent damage to blood vessels or the bowel[23]. It is important to emphasize that since the first trocar is placed before camera insertion, its placement cannot be visually captured through video footage, which forms the basis of this study. Consequently, our analysis focuses on subsequent trocars, as their placement occurs under direct visualization once the camera is operational.

Suction removes fluid or blood build-up to keep the surgical field clear, creating optimal conditions for the next steps of the operation. *Use Clip* involves using a clamping tool to secure tissue or blood vessels, usually to control bleeding or isolate vascular structures. Actions like *Suction* and *Use Clip* are pivotal for preserving the operative field and controlling bleeding. *Suction* is often necessary to remove blood or fluids from the surgical site, ensuring the surgeon has a clear view of the surgical field, thus reducing the risk of unintended harm or incomplete treatments. Whereas *Use Clip* is commonly employed to maintain hemostasis and prevent

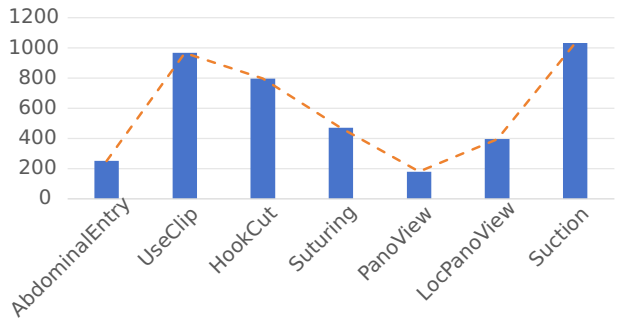

Fig. 2 Total Clip Count for Each Surgical Action Category in the Dataset (Simplified for Clarity, e.g., "Use Clip" → "UseClip").

bleeding-related complications during surgery[24]. Both actions are directly associated with patient safety, as bleeding or poor visibility might increase the risk of surgical errors.

Hook Cut describes the precise cutting and separation of specific tissues using an electric hook. *Hook Cut* minimizes damage to surrounding structures as the hook shape allows for high precision in cutting and separating tissues. This action requires high precision and skill to maintain both safety and effectiveness. *Suturing* involves inserting and withdrawing a needle for tissue suturing or similar tasks. *Suturing*, another key action, is vital for wound closure and tissue repair. Its precision facilitates proper healing and reduces post-surgical problems, such as infections or dehiscence[25]. Suturing is a technically demanding skill in laparoscopic surgery.

Unlike open surgery, where the entire operative field is visible, *Panoramic View* and *Local Panoramic View* are used to assist surgeons in navigating and orienting within the abdominal cavity. *Panoramic View* and *Local Panoramic View* provide different perspectives during surgery, with Panoramic View capturing an overall view and Local Panoramic View focusing on specific details. These views provide the surgeon with a broader and more focused view of the surgical site, aiding in identifying anatomical structures, planning the subsequent process, and preventing errors in placement. The assessment of both views was included to leverage their complementary roles: Panaromic views standardize the operation process, reducing spatial disorientation and aiding in initial target localization, while local views enable detailed examination of target tissues, such as vascular structures.

We uniformly sample and annotate surgical actions, then record the occurrence frequencies of each action. The results, displayed in Fig. 2, illustrate the distribution of each action throughout the surgeries. Abdominal Entry and Panoramic View are the two actions with the least data. Abdominal Entry occurs only once per laparoscopic surgery, resulting in limited instances, while Panoramic View is also infrequently needed. More details are provided in Table 3, where we offer the exact number of clips annotated for each surgical action according to each patient index, along with the types of surgeries each patient underwent. We also include a small dataset on lung surgery (*Video-Assisted Thoracic Surgery (VATS) for Lung Surgery*, index 34 of Table 3) to validate two key points in our experiments: (1) laparoscopic surgical action recognition forms a foundation, with knowledge transferable to other surgical types, and (2) cross-surgical data diversity improves augmentation. This also provides opportunities to explore transfer learning across surgical domains.

It is worth noting that while not all laparoscopic surgeries involve each annotated action, the chosen actions were identified as the most commonly performed in various prevalent laparoscopic procedures. Previous work, such as LapGyn6-Actions[19], have explored similar classification approaches where not all annotated actions appear in every procedure. Our dataset, however, expands upon these classifications by incorporating a broader range of surgical actions. In addition, videos collected during the study periods confirm that these classes are among the most frequently observed actions. Focusing on these essential surgical actions, our dataset provides a framework for recognizing laparoscopic action, supporting advancements in surgical intelligence and automation across diverse clinical settings.

Our dataset encompasses a diverse range of surgical procedures to address the challenge of surgical action recognition across various interventions. We aim to enhance AI models' adaptability and generalization capabilities to a broad spectrum of clinical settings, as numerous fundamental surgical acts, including Abdominal Entry and Suturing, are common to multiple procedures. This clip-based methodology, in which each surgery case contributes multiple action clips, enhances the dataset's efficiency and the AI's ability to adapt to different surgical contexts. Furthermore, we plan to expand this dataset continuously, adding more samples to improve its coverage or considering extending it to Action Triplets.

Video recording. We carefully designed and strictly adhered to a video collection protocol to create a comprehensive, high-quality video dataset of laparoscopic surgery. We collected a total of 34 complete laparoscopic surgery videos recorded during procedures performed between 2021 and 2023 at RuiJin Hospital LuWan Branch, Shanghai Jiaotong University School of Medicine, an academic and teaching hospital in Shanghai, China. As a leading medical institution, it ensures that our mono-centric dataset reflects standardized surgical practices performed by experienced surgeons.

These videos span a substantial period of time and include a wide variety of surgical types, ensuring both diversity and representativeness. Specifically, our dataset primarily consists of laparoscopic cholecystectomy

Index	Types of Surgery	Abdominal Entry	Use Clipper	Hook Cut	Suturing	Panoramic View	Local Panoramic View	Suction	Total
01	Cholecystectomy	0	48	33	0	7	0	20	108
02	Cholecystectomy	3	21	52	0	4	0	10	90
03	Appendectomy	0	0	8	0	8	0	8	24
04	Cholecystectomy	6	43	88	0	5	12	0	154
05	Resection of Gastric Stromal Tumor	0	0	0	126	0	0	0	126
06	Abdominal Wall Hernia Repair	12	4	0	108	18	42	18	202
07	Cholecystectomy	0	4	13	0	0	1	6	24
08	Cholecystectomy	25	72	21	0	0	0	0	118
09	Cholecystectomy	0	23	16	0	0	1	5	45
10	Cholecystectomy with Appendectomy	0	27	6	0	0	2	11	46
11	Cholecystectomy	6	37	92	0	0	13	39	187
12	Resection of Gastric Stromal Tumor	0	0	0	75	0	0	0	75
13	Cholecystectomy	0	31	36	0	11	0	22	100
14	Cholecystectomy	0	25	34	0	17	0	13	89
15	Sigmoid Colectomy	0	0	0	9	0	0	0	9
16	Cholecystectomy	11	25	29	0	1	0	8	74
17	Cholecystectomy	6	44	7	0	0	0	3	60
18	Cholecystectomy	8	5	0	58	0	0	3	74
19	Radical Resection of Rectal Cancer	29	323	0	0	25	159	407	943
20	Cholecystectomy	8	0	0	32	18	0	24	82
21	Cholecystectomy	0	0	0	26	0	0	72	98
22	Cholecystectomy	8	25	36	0	4	4	16	93
23	Appendectomy	18	0	24	0	17	23	32	114
24	Appendectomy	0	0	8	0	0	15	0	23
25	Radical Resection of Sigmoid Colon Cancer	0	0	0	37	0	0	0	37
26	Cholecystectomy with Appendectomy	5	14	16	0	0	0	43	78
27	Splenectomy	16	51	0	0	8	19	9	103
28	Cholecystectomy	44	31	66	0	0	8	79	228
29	Cholecystectomy	6	20	26	0	1	2	9	64
30	Cholecystectomy	8	25	39	0	8	11	27	118
31	Appendectomy	20	0	13	0	10	33	29	105
32	Cholecystectomy	9	34	112	0	1	6	20	182
33	Cholecystectomy	3	15	23	0	1	1	6	49
34	VATS for Lung Surgery	0	21	0	0	16	44	94	175
Total	—	251	968	798	471	180	396	1033	4097

Table 3. Annotated Clip Count for Each Surgical Action in SLAM Dataset.

surgeries, supplemented by other common laparoscopic procedures such as appendectomy, cholecystectomy with appendectomy, lung segmentectomy, sleeve gastrectomy, and colon resection. This diversity enriches the dataset by encompassing a wide range of laparoscopic operations and complexities, establishing a strong foundation for analyzing various actions within the surgical process. It further supports medical research and AI development, particularly in automating and optimizing laparoscopic surgical procedures.

The surgical video recording uses standard high-definition medical camera equipment, and the video collection process follows standardized procedures to ensure data quality. With a frame rate of 25 frames per second (fps), the recordings ensure smooth visualization of surgical operations. The video resolution of 1920×1080 provides high-fidelity detail, particularly highlighting precise instrument movements and anatomical operation sites. The equipment includes a 26003BA HOPKINS 30° endoscope, which features an ultra-wide field of view, a 10-mm diameter, and a length of 31 centimeters. It is designed to withstand high-temperature and high-pressure disinfection. Moreover, it integrates fiber optic transmission to maintain anti-interference capabilities and stability in complex surgical environments, ensuring continuous, high-quality video recording. Before initiating each surgery, professional technicians thoroughly calibrate and test the camera equipment. All cameras are securely installed in the standard operating room, with the camera positioned next to the operating table and connected to the laparoscopic equipment, enabling real-time transmission and high-definition recording of the procedure. This fixed installation method minimizes visual interference or obstruction from movement, ensuring a complete and uninterrupted recording of each surgery without missing any critical steps. However, variation in video quality can arise due to technical factors, such as lighting conditions, camera settings, and compression applied during the recording process. These variations were not intentionally added but are inherent to surgical video capture. Furthermore, a subset of videos in the dataset exhibit vertical line artifacts that were introduced during the original surgical recording process rather than during post-processing or encoding. Training on imperfect and variable datasets enables the models to generalise and perform well under various

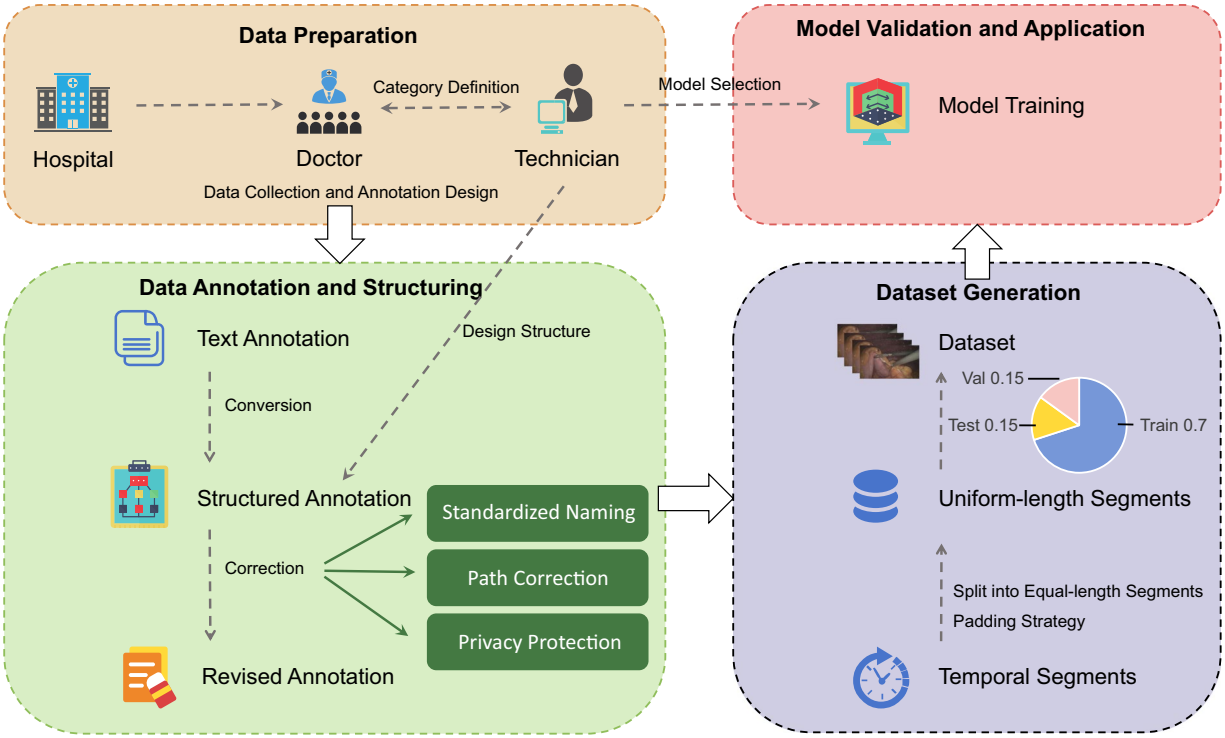

Fig. 3 Workflow of Data Collection and Annotation.

real-world conditions, where video quality can vary. Therefore, their presence can make AI models more robust, although they affect the videos' visual quality. We also provide an auxiliary metadata file that explicitly lists the video clips affected by visible artifacts. This file is available as part of the supplementary material and included in the Figshare repository.

Dataset workflow. The workflow for dataset development and model validation, illustrated in Fig. 3, consists of four primary stages. The first stage, *Data Preparation*, begins with close collaboration between medical experts and technical teams. Medical experts define the categories of surgical actions, while technical teams handle data collection and assist in designing annotation protocols. The second stage, *Data Annotation and Structuring stage*, focuses on annotating and organizing the collected data. Initial textual annotations are created and subsequently converted into structured formats. These annotations are refined for accuracy through standardized naming conventions, path corrections, and privacy protection measures, ensuring high-quality, ethically compliant data. In the third stage, *Dataset Generation*, the annotated data is segmented into uniform-length temporal chunks suitable for model training. The entire dataset is split into a training subset (70%), a validation subset (15%), and a testing subset (15%), ensuring a balanced distribution for a thorough evaluation. A looping strategy handles variability in segment lengths by filling shorter segments by cycling frames from the beginning. We also provide a supplementary CSV file that specifies the original start and end frames for each unlooped video segment. The final stage, *Model Validation and Application*, utilizes the processed dataset to train and validate a machine-learning model. Advanced algorithms evaluate and enhance the model's ability to recognize surgical actions. This comprehensive workflow ensures the creation of a high-quality, diverse dataset while enabling the development of effective machine-learning models for laparoscopic action recognition[26].

Data Records

Our laparoscopic surgery video dataset called Surgical LAparo Motion (SLAM)[20] is publicly available on *Figshare* without registration. The dataset is divided into four layers for enhanced organization and usability. At the top level, the LaparoMotion directory consists of a clip subfolder and three CSV files: train.csv, val.csv, and test.csv. Each cvs file contains three columns: **patinet**, representing the patient's ID; **file_path**, indicating the path to the video file; and **label**, denoting the category of surgical actions it belongs to. This organization allows efficient access to specific files and their labels, thus simplifying the training, validation, and testing process.

The clip subfolder comprises 34 patient-specific directories, each distinctly identified by an anonymized patient ID generated using a hashing method to ensure patient confidentiality. Each patient folder contains between 15 and 38 video clips, each uniformly segmented into 30 frames using a patching approach to ensure consistent input sizes for model analysis. This standardization not only improves the compatibility of the dataset with machine learning models but also enables a more precise and efficient analysis by maximizing the utilization of all available segments.

FrameNo.	ImageRes.	UseClipper	HookCut	PanoView	Suction	Abdominal Entry	Suturing	LocPano View	Test Accuracy
8	224	89.04%	89.26%	32.14%	76.28%	82.05%	84.72%	71.67%	80.77%
8	448	91.10%	95.04%	42.86%	80.13%	76.92%	84.72%	60.00%	82.37%
8	768	92.47%	95.04%	39.29%	83.97%	84.62%	91.67%	58.33%	84.62%
16	224	91.10%	96.69%	42.86%	72.44%	87.18%	87.50%	70.00%	82.69%
16	448	90.41%	95.04%	60.71%	78.85%	82.05%	83.33%	75.00%	84.29%
16	768	92.47%	94.21%	71.43%	79.49%	66.67%	88.89%	85.00%	85.90%

Table 4. Experimental Results (Presented as Percentages %) of Model ViViT.

Technical Validation

To ensure the robustness and applicability of the laparoscopic surgical video dataset, we conducted a comprehensive validation using the Video Vision Transformer (ViViT) model, a cutting-edge architecture for action identification in video sequences[27].

Overview of the ViViT Model. In this study, we chose the ViViT model for its exceptional ability to handle complex dynamic visual patterns in laparoscopic surgical videos. ViViT embodies a cutting-edge approach to video classification by leveraging the strengths of Transformer architectures to extract both spatial and temporal features. Unlike traditional convolutional networks, ViViT employs self-attention mechanisms to model long-range dependencies within video data, enabling it to effectively learn intricate spatio-temporal patterns. Its hierarchical, transformer-based architecture is particularly well-suited for capturing fine-grained visual changes over time, making it an ideal choice for action recognition and event detection tasks.

Experimental Setup. Our dataset was carefully structured and comprehensively annotated to encompass specific laparoscopic actions, including "UseClipper," "HookCut," "PanoView," "Suction," "AbdominalEntry," "Suturing," and "LocPanoView." To assess the quality and utility of these annotations and the overall structure of the data set, the ViViT model was trained for 80 epochs in a subset of the data representing these actions. The model's performance was validated on a separate validation set, and the model weights that achieved the highest metrics on the validation set were subsequently applied to an independent test set for further assessment. The model configuration included input video segments of 8 frames, each resized to 224 $\times$ 224 pixels, and training was carried out in batches of size 4. This setup ensured the model could adequately capture temporal dynamics and spatial details.

Evaluation on Performance Metrics. As shown in Table 4, the overall test accuracy highlights how configurations of time step length and image resolution significantly improve the performance of the ViViT model in laparoscopic surgical action classification. Regarding time-step length, longer time steps (e.g., 16 frames) markedly improve the recognition accuracy for specific surgical actions, notably in the "HookCut" and "Suturing" categories, achieving accuracies of 96.69% and 87.50%, respectively. This suggests that extending the time step enables the model to capture more comprehensive temporal features, thereby enhancing its ability to recognize these actions accurately. However, for shorter actions such as "UseClipper," an extended time step may introduce redundant information and noise, leading to decreased performance.

Furthermore, increasing image resolution positively impacts the model's performance in certain categories. The accuracy for "UseClipper" attains 92.47% at 768 pixels, approximately 3% higher than at 224 pixels, indicating that higher resolutions allow the model to capture fine-grained features more effectively. However, for action classes characterized by frequent scene transitions (e.g., LocPanoView and Suturing), the benefit of higher resolution is relatively minimal. This may be because the rapid temporal transitions in these actions pose a challenge in achieving substantial accuracy improvements at higher resolutions.

Furthermore, the performance differences across action categories highlight the difficulties the ViViT model encounters when managing actions with frequent viewpoint changes (e.g., "PanoView"). The classification accuracy for these actions remains low, especially at lower resolutions and shorter time steps, varying between 32.14% and 71.43%. These results suggest that future research could focus on improving the model's ability to recognize such actions through additional pre-processing techniques or improving the model's capacity to capture temporal context.

At its peak test accuracy (85.90%), the model achieved a test loss of 0.4554, indicating that the dataset effectively supports the training and validation of laparoscopic surgical action recognition models. These performance metrics further indicate the potential of this dataset as a benchmark, poised to play a key role in the advancement of action recognition algorithms within minimally invasive surgical environments.

Analysis on Laparoscopic and Thoracic Data. We conduct experiments to assess the transferability of surgical action recognition from laparoscopic to thoracic surgery by evaluating the model's accuracy on test datasets containing exclusively laparoscopic or thoracic data. Since knowledge from different surgical data can benefit each other, we also evaluate the impact of thoracic data as augmented data on model performance. To do this, we train two models: one using only laparoscopic data (referred to as the laparoscopic-only model) and another using a mix of laparoscopic and thoracic data (referred to as the mixed model). As shown in Table 5, the laparoscopic-only model achieves 43.75% accuracy on the thoracic test set, validating the transferability of

Model	Mixed Test Accuracy	Laparoscopic Test Accuracy	Thoracic Test Accuracy
Laparoscopic-Only Model	79.97%	81.59%	43.75%
Mixed Model	85.90%	86.66%	65.62%

Table 5. Test Accuracy (Presented as Percentages %) of Laparoscopic-Only and Mixed Models on Laparoscopic, Thoracic, and Mixed Data.

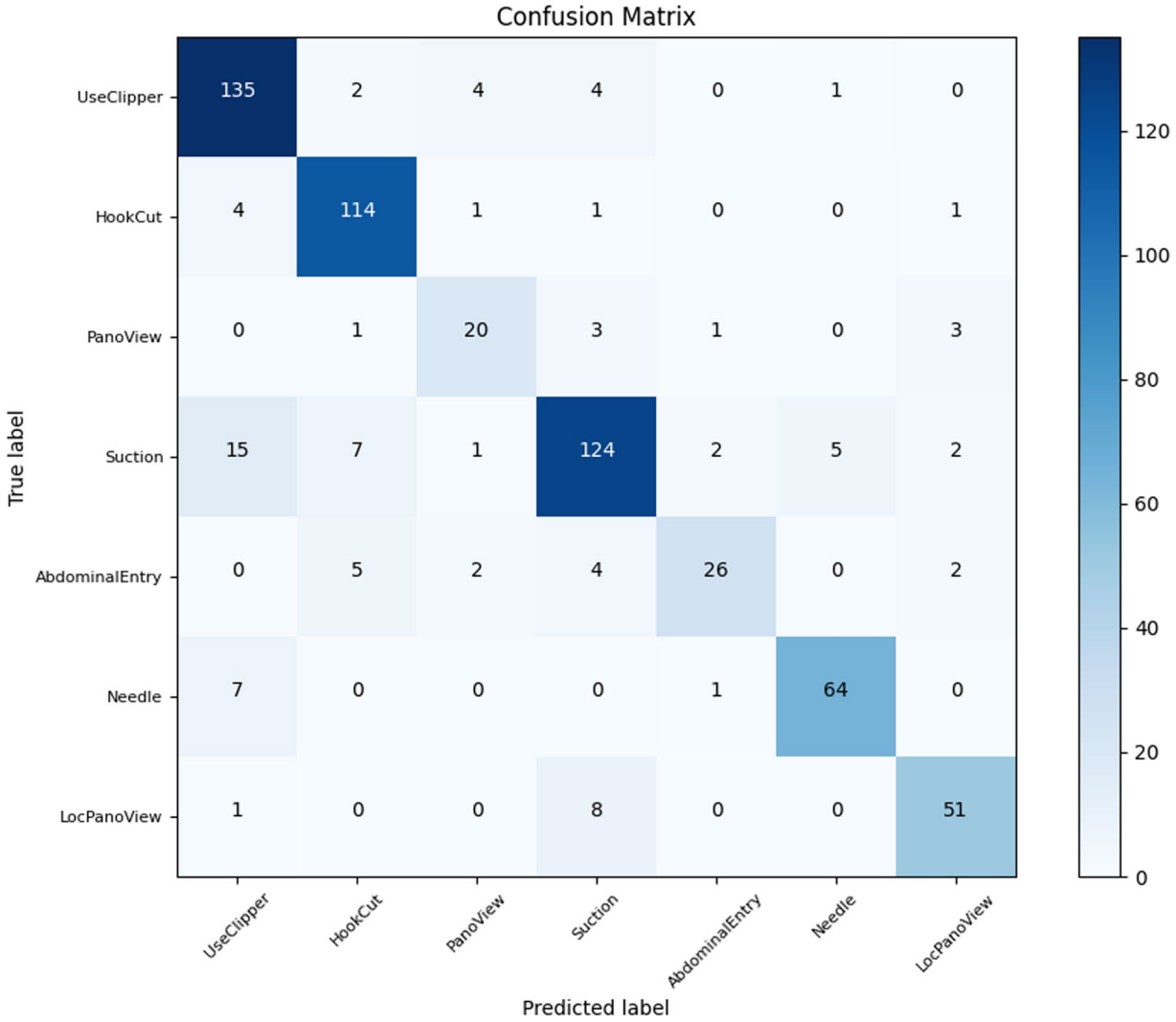

Fig. 4 Confusion matrix when the number of input frames is 16 and image resolution is 768.

laparoscopic surgical action recognition. On the other hand, the mixed model demonstrates improved performance across laparoscopic, thoracic, and mixed test data, highlighting that thoracic data augmentation enhances model performance and that diversity plays a beneficial role.

Confusion Matrix. We generated confusion matrices to evaluate the model performance across specific action categories, with Fig. 4 presenting the matrix when the number of input frames is 16 and the resolution is 768. This highlights misclassification cases, particularly in visually similar or context-dependent actions such as "PanoView" and "LocPanoView." For example, "PanoView" has twenty accurate classifications but also has three instances incorrectly classed as "LocPanoView." Additionally, overlap was observed between "Suction" and "Suturing," probably due to visual or contextual ambiguities contributing to these misclassifications.

The results emphasize the intrinsic challenges of accurately recognizing specific laparoscopic actions, particularly those requiring nuanced contextual understanding for proper differentiation. This indicates that expanding the dataset with more examples or enhancing the model's ability to capture contextual cues could help lower the misclassification rate. Such improvements could significantly boost overall recognition accuracy, especially for actions that appear visually similar but differ in context.

Dataset Reliability. To ensure data consistency, we followed strict protocols during video capture and annotation. High-definition recordings at a resolution of 1920 × 1080 and a standardized frame rate of 25 fps provided clear visual detail, crucial for identifying fine motor actions. Additionally, anonymization measures and the structured organization of video clips into training, validation, and test sets promote reproducibility and enable reliable performance comparisons across models.

The annotation process was conducted by a team of three highly qualified medical experts to ensure the dataset's precision and reliability. Two deputy chief physicians specializing in laparoscopic procedures independently performed the initial annotations, each labeling half of the dataset. To further enhance annotation quality and consistency, a chief physician with extensive expertise in surgical practice conducted a secondary review on a randomly selected 10% subset, making necessary modifications in case ambiguities were identified. Within this subset, approximately 1% of the annotations were adjusted to correct. This review process addressed potential inconsistencies and provided an additional layer of validation, reinforcing the reliability of the dataset. A limitation of inter-rater agreement analysis is acknowledged. However, as we grow the dataset, more annotators and a secondary review process will be added to cross-check annotions, reducing potential inconsistencies systematically.

In summary, the validation experiments confirm the dataset's effectiveness for laparoscopic surgical action recognition, underscoring its significance in advancing AI applications in surgical contexts. Future research could focus on leveraging advanced model architectures or implementing additional data augmentation techniques to further enhance performance, especially for classes with high variability or significant interclass similarity.

Code availability

The dataset preparation codes are accessible via the GitHub repository: https://github.com/yezizi1022/SLAM-Vivit_Cls.

Received: 9 January 2025; Accepted: 29 April 2025;
Published online: 24 May 2025

References

1. Twinanda, A. P. *et al*. Endonet: a deep architecture for recognition tasks on laparoscopic videos. *IEEE transactions on medical imaging* **36**, 86–97, https://doi.org/10.1109/TMI.2016.2593957 (2016).
2. Buia, A., Stockhausen, F. & Hanisch, E. Laparoscopic surgery: a qualified systematic review. *World journal of methodology* **5**, 238, https://doi.org/10.5662/wjm.v5.i4.238 (2015).
3. Anteby, R. *et al*. Deep learning visual analysis in laparoscopic surgery: a systematic review and diagnostic test accuracy meta-analysis. *Surgical endoscopy* **35**, 1521–1533, https://doi.org/10.1007/s00464-020-08168-1 (2021).
4. Mascagni, P. *et al*. Artificial intelligence for surgical safety: automatic assessment of the critical view of safety in laparoscopic cholecystectomy using deep learning. *Annals of surgery* **275**, 955–961, https://doi.org/10.1097/SLA.0000000000004351 (2022).
5. Khatibi, T. & Dezyani, P. Proposing novel methods for gynecologic surgical action recognition on laparoscopic videos. *Multimedia Tools and Applications* **79**, 30111–30133, https://doi.org/10.1007/s11042-020-09540-y (2020).
6. Madhok, B., Nanayakkara, K. & Mahawar, K. Safety considerations in laparoscopic surgery: A narrative review. *World journal of gastrointestinal endoscopy* **14**, 1, https://doi.org/10.4253/wjge.v14.i1.1 (2022).
7. Guo, K. *et al*. Current applications of artificial intelligence-based computer vision in laparoscopic surgery. *Laparoscopic, Endoscopic and Robotic Surgery*. https://doi.org/10.1016/j.lers.2023.07.001 (2023).
8. Nwoye, C. I. *et al*. Recognition of instrument-tissue interactions in endoscopic videos via action triplets. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)* (pp. 364–374). Springer. https://doi.org/10.1007/978-3-030-59716-0_35 (2020).
9. Golany, T. *et al*. Artificial intelligence for phase recognition in complex laparoscopic cholecystectomy. *Surgical Endoscopy* **36**, 9215–9223, https://doi.org/10.1007/s00464-022-09405-5 (2022).
10. Bian, G.-B., Peng, Y., Zhang, L., Li, J. & Li, Z. Algorithms in Surgical Action Recognition: A Survey. In *2024 IEEE International Conference on Real-time Computing and Robotics (RCAR)* (pp. 366–371). IEEE. https://doi.org/10.1109/RCAR61438.2024.10670748 (2024).
11. L'heureux, A., Grolinger, K., Elyamany, H. F. & Capretz, M. A. M. Machine learning with big data: Challenges and approaches. *IEEE Access* **5**, 7776–7797, https://doi.org/10.1109/ACCESS.2017.2696365 (2017).
12. Zhang, J. *et al*. Laparoscopic Image-Based Critical Action Recognition and Anticipation With Explainable Features. *IEEE Journal of Biomedical and Health Informatics*. https://doi.org/10.1109/JBHI.2023.3306818 (2023).
13. Leibetseder, A. *et al*. Lapgyn4: a dataset for 4 automatic content analysis problems in the domain of laparoscopic gynecology. In *Proceedings of the 9th ACM multimedia systems conference* (pp. 357–362). ACM. https://doi.org/10.1145/3204949.3208127 (2019).
14. Bawa, V. S. *et al*. The saras endoscopic surgeon action detection (esad) dataset: Challenges and methods. Preprint at https://arxiv.org/abs/2104.03178 (2021).
15. Valderrama, N. *et al*. Towards holistic surgical scene understanding. In *International conference on medical image computing and computer-assisted intervention* (pp. 442–452). Springer. https://doi.org/10.1007/978-3-031-16449-1_42 (2022).
16. Wagner, M. *et al*. Comparative validation of machine learning algorithms for surgical workflow and skill analysis with the heichole benchmark. *Medical image analysis* **86**, 102770, https://doi.org/10.1016/j.media.2023.102770 (2023).
17. Nwoye, C. I. *et al*. CholecTriplet2021: A benchmark challenge for surgical action triplet recognition. *Medical Image Analysis* **86**, 102803 (2023).
18. Nwoye, C. I. *et al*. CholecTriplet2022: Show me a tool and tell me the triplet-An endoscopic vision challenge for surgical action triplet detection. *Medical Image Analysis* **89**, 102888, https://doi.org/10.1016/j.media.2023.102888 (2023).
19. Nasirihaghighi, S., Ghamsarian, N., Stefanics, D., Schoeffmann, K. & Husslein, H. Action recognition in video recordings from gynecologic laparoscopy. In *2023 IEEE 36th International Symposium on Computer-Based Medical Systems (CBMS)* (pp. 29–34). IEEE. https://doi.org/10.1109/CBMS58004.2023.00187 (2023).
20. Ye, Z. Surgical LAparoscopic Motions. *figshare* https://doi.org/10.6084/m9.figshare.28104782.v3 (2025).
21. Kitaguchi, D. *et al*. ChAutomated laparoscopic colorectal surgery workflow recognition using artificial intelligence: experimental research. *International journal of surgery* **79**, 88–94, https://doi.org/10.1016/j.ijsu.2020.05.015 (2020).
22. Moorthy, K., Munz, Y., Sarker, S. K. & Darzi, A. Objective assessment of technical skills in surgery. *BMJ* **327**, 1032–1037, https://doi.org/10.1136/bmj.327.7422.1032 (2003).

23. Madhok, B., Nanayakkara, K. & Mahawar, K. Safety considerations in laparoscopic surgery: A narrative review. *World Journal of Gastrointestinal Endoscopy* **14**, 1–16, https://doi.org/10.4253/wjge.v14.i1.1 (2022).
24. Van Hove, P. D., Tuijthof, G. J. M., Verdaasdonk, E. G. G., Stassen, L. P. S. & Dankelman, J. Objective assessment of technical surgical skills. *Journal of British Surgery* **97**, 972–987, https://doi.org/10.1002/bjs.7115 (2010).
25. Chen, J. *et al*. Objective assessment of robotic surgical technical skill: a systematic review. *The Journal of urology* **201**, 461–469, https://doi.org/10.1016/j.juro.2018.06.078 (2019).
26. Nwoye, C. I. & Padoy, N. Data Splits and Metrics for Benchmarking Methods on Surgical Action Triplet Datasets. Preprint at https://arxiv.org/abs/2204.05235 (2022).
27. Arnab, A. *et al*. Vivit: A video vision transformer. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 6836–6846). IEEE. https://doi.org/10.1109/ICCV48922.2021.00676 (2021).

Acknowledgements

We would like to express our gratitude to the Department of General Surgery, RuiJin Hospital LuWan Branch, Shanghai Jiaotong University School of Medicine, Shanghai, China, for their invaluable support in providing the data resources that made this study possible. Their contributions were integral to the development of the SLAM dataset and the advancement of our research.

Author contributions

Z.Y. conceived the experiments and drafted the manuscript. H.Z., M.W., and L.Z. organized and supervised the study. Z.D. and D.W. were responsible for data collection and preprocessing. R.Z. and X.J. performed data annotation. Y.Z., Z.D., and T.C. conducted and validated the experiments. All authors reviewed and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to H.Z.