MELBA Journal Machine Learning for Biomedical Imaging

SurgiSR4K: A High-Resolution Endoscopic Video Dataset for Robotic-Assisted Minimally Invasive Procedures

Fengyi **Jiang** [1], Xiaorui **Zhang** [1], Lingbo **Jin** [1], Ruixing **Liang** [1,2,3], Yuxin **Chen** [1,4], Adi Chola **Venkatesh** [1], Jason **Culman** [1], Tiantian **Wu** [5], Lirong **Shao** [1], Wenqing **Sun** [1], Cong **Gao** [1], Hallie **McNamara** [1], Jingpei **Lu** [1], Omid **Mohareri** [1]

1 Intuitive Surgical, Inc., Sunnyvale, CA, USA
2 Johns Hopkins Medicine Neurosurgery, Baltimore, MD, USA
3 Johns Hopkins University Electrical and Computer Engineering, Baltimore, MD, USA
4 University of British Columbia Electrical and Computer Engineering, Vancouver, BC, Canada
5 Wilford & Kate Bailey Small Animal Teaching Hospital, Auburn, AL, USA

Abstract

High-resolution imaging is crucial for enhancing visual clarity and enabling precise computer-assisted guidance in minimally invasive surgery (MIS). Despite the increasing adoption of 4K endoscopic systems, there remains a significant gap in publicly available native 4K datasets tailored specifically for robotic-assisted MIS. We introduce SurgiSR4K, the first publicly accessible surgical imaging and video dataset captured at a native 4K resolution, representing realistic conditions of robotic-assisted procedures. SurgiSR4K comprises diverse visual scenarios including specular reflections, tool occlusions, bleeding, and soft tissue deformations, meticulously designed to reflect common challenges faced during laparoscopic and robotic surgeries. This dataset opens up possibilities for a broad range of computer vision tasks that might benefit from high resolution data, such as super resolution (SR), smoke removal, surgical instrument detection, 3D tissue reconstruction, monocular depth estimation, instance segmentation, novel view synthesis, and vision-language model (VLM) development. SurgiSR4K provides a robust foundation for advancing research in high-resolution surgical imaging and fosters the development of intelligent imaging technologies aimed at enhancing performance, safety, and usability in image-guided robotic surgeries.

Keywords

Endoscopy, Surgical Robotics, Super-Resolution, Segmentation, Depth Estimation, Tool Tracking, 3D Tissue Reconstruction, Monocular Depth Estimation

Article informations

https://doi.org/10.59275/j.melba.2025-f593©2025 Jiang, Zhang, Jin, Liang, Chen, Venkatesh, Culman, Wu, Shao, Sun, Gao, McNamara, Lu, and Mohareri. License: CC-BY 4.0
Published 12/2025
Corresponding author: Fengyi Jiang art.jiang@intusurg.com
Special issue: MICCAI Open Data special issue
Guest editors: Martijn Starmans, Apostolia Tsirikoglou, Lidia Garrucho Moras, Kaouther Mouheb

1. Background

Minimally invasive surgery (MIS), including laparoscopic and robot-assisted procedures, has transformed surgical practice by reducing patient recovery time, minimizing tissue trauma, and improving cosmetic outcomes (Mack, 2001). As these techniques become the standard of care in a wide range of clinical specialties, from general surgery to gynecology and urology, the demand for enhanced surgical vision and intra-operative guidance has increased in parallel (Conrad et al., 2015; Oberlin et al., 2016). In particular, the quality of intra-operative imaging has become increasingly vital for ensuring surgical safety, precise dissection, and the detection of critical anatomical features such as vessel bifurcations, lesion margins, or hidden pathology (Marshall et al., 2010; Goh et al., 2010). For example, accurate margin assessment during colorectal or liver surgery can directly impact long-term patient outcomes (Montalti et al., 2015).

In recent years, advances in medical imaging have driven a shift toward higher-resolution acquisition systems. In the domain of MIS, modern laparoscopy platforms such as Olympus VISERA or Stryker 1688 have begun offering ultra-high-definition (UHD) 4K images, promising substantial benefits in surgical navigation and decision making (Abdelrahman et al., 2018). High resolution visualization enables surgeons to better differentiate tissue types, observe microvasculature, and assess resection margins, capabilities that are crucial for both oncologic and reconstructive interventions, as shown in Figure 1. Recent clinical and simulator studies demonstrate that 4k laparoscopy shortens the operative time and reduces blood loss compared to HD systems (Puccetti et al., 2023). Randomized skill acquisition trials also show faster learning curves and fewer errors for 4K vision (Abdelrahman et al., 2018), and systematic evidence confirms that many performance metrics are close to those of dedicated 3D systems when 4K is used (Harada et al., 2018). In addition, detailed imaging facilitates the development of downstream computer-assisted intervention (CAI) tools, such as automated segmentation, tissue classification, and image-based depth estimation. 4K resolution is widely available in non-surgical commercial space, yet most public surgical vision instruments remain trapped in the HD/sub-HD era, preventing algorithms from unleashing their full potential on utilizing high-resolution images.

1.1 Related Works

Despite the growing availability of 4K endoscopic systems, the integration of 4K imaging into routine surgical practice and robotic-assisted practice remains limited due to high hardware costs, bandwidth constraints, and legacy infrastructure. Compounding this challenge is the surprising lack of publicly available 4K surgical datasets particularly those designed to support super-resolution (SR) and AI-driven computer-assisted intervention research. Early deep SR frameworks such as EndoL2H demonstrated clinically perceptible gains in capsule footage (Almalioglu et al., 2020). More recent transformer architectures further boost video PSNR in operative scenes (Song et al., 2022; Zhang and Yang, 2023), but are still constrained by sub-4K sources. SurgiSR4K provides the **first** native-resolution benchmark to close this training–deployment mismatch.

The benefits of high-resolution imaging extend beyond SR, offering significant advantages for tasks such as monoscopic depth estimation and novel view synthesis (Xu et al., 2024). Unlike stereo endoscopes, which require dual-camera calibration and increase hardware complexity, monoscopic systems can potentially infer depth from a single high-resolution input. This is particularly valuable in robotic platforms where camera movement is restricted or in laparoscopic systems that use fixed-angle scopes. When trained on 4K video, models can better extract fine spatial details and textural cues necessary for robust depth estimation. These depth maps, in turn, support consistent novel view synthesis, enabling virtual camera repositioning, depth-aware overlays, or immersive visualization in education and planning.(Masuda et al., 2024)

As shown in Table 1, currently available surgical vision datasets are limited in resolution, with most capped at 1280x1024 pixels (Allan et al., 2020), 1350×1080 pixels for odometry-focused datasets such as EndoSLAM (Fredrick, 2022) or 1920×1080 pixels (Allan et al., 2019). Although these datasets have enabled progress in many vision tasks, their resolution is insufficient for training high-performance models that rely on capturing fine anatomical and contextual details. To address this critical gap, we introduce SurgiSR4K, the first high-quality, 4K-resolution surgical image and video dataset dedicated to enabling super-resolution and image understanding research in the context of minimally invasive procedures.

To our knowledge, SurgiSR4K represents the first dataset of its kind designed explicitly for applications that require high-resolution data in surgical settings. We anticipate that it will play a pivotal role in bridging the gap between state-of-the-art vision models and the operational needs of computer-assisted surgical platforms. As surgical AI continues to move from proof-of-concept to operating room integration, data sets such as SurgiSR4K will be indispensable in training, validating, and deploying reliable visual intelligence systems.

2. Summary

SurgiSR4K represents a significant advancement in surgical imaging datasets, providing the first publicly available collection of native 4K resolution endoscopic images and videos specifically designed for robotic-assisted minimally invasive surgery research. The dataset comprises 800 high-quality 4K PNG images and 50 video clips (each 5 seconds at 30 fps) (Table 2) captured from porcine animal models under realistic surgical conditions using a Da Vinci Xi surgical robot system.

Unlike existing surgical datasets that are limited to HD or sub-HD resolutions, SurgiSR4K offers unprecedented visual detail at 3840×2160 pixels, enabling researchers to explore the full potential of high-resolution imaging in surgical AI applications. The dataset includes diverse surgical scenarios featuring various robotic instru-

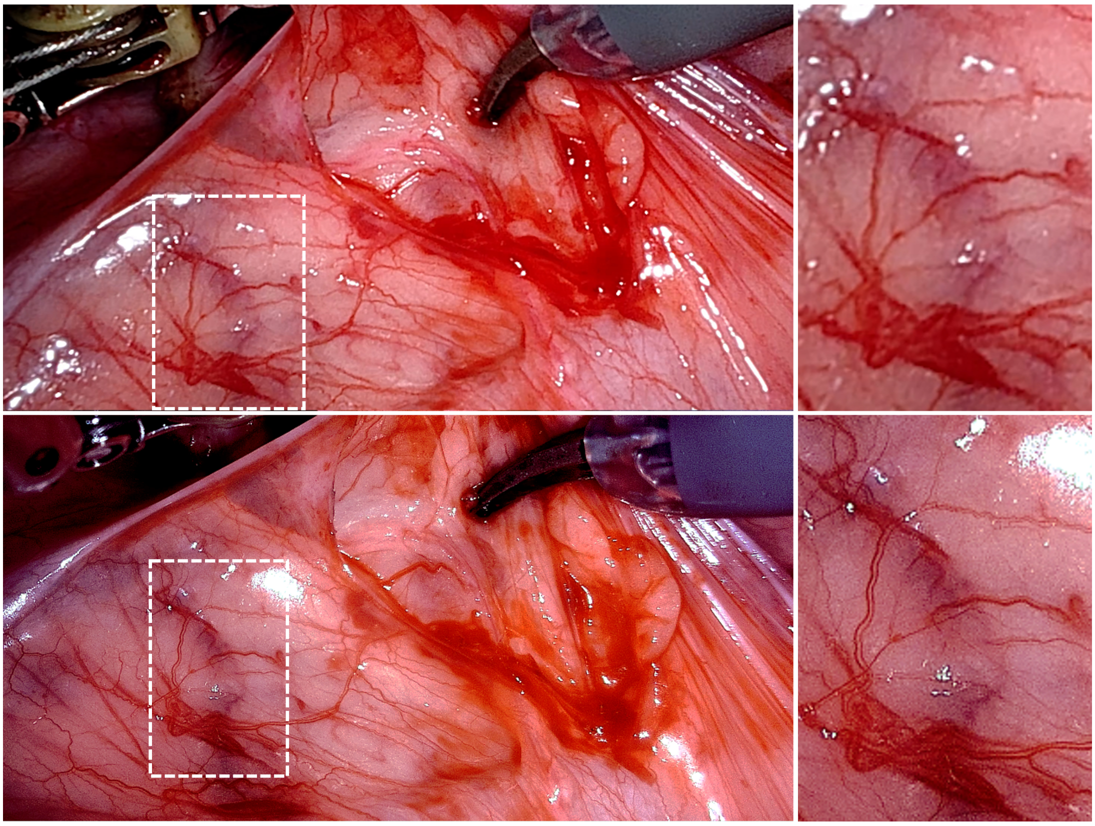

Figure 1: Side-by-side comparison of 1080p (top) and 4K (bottom) endoscopic images captured simultaneously with separate 1080p and 4K cameras.

ments (monopolar curved scissors, force bipolar, bipolar forceps, grasping retractor, and Cadiere forceps) and challenging visual conditions such as specular reflections, tool occlusions, bleeding, smoke dispersion, and soft tissue deformations.

To support multiscale research, each 4K image and video frame is accompanied by downsampled versions at 960x540p and 480x270p resolutions using Lanczos resampling (Fig. 3), creating perfectly aligned multi-resolution triplets ideal for super-resolution training and evaluation. The dataset is structured to facilitate a broad range of computer vision tasks, including superresolution, monoscopic depth estimation, surgical instrument detection and tracking, instance segmentation, 3D tissue reconstruction, and novel view synthesis.

All content has been clinically validated by experienced medical professionals and is released under the CC-BY-NC-SA 4.0 license to promote responsible use of research. SurgiSR4K addresses a critical gap in surgical AI research by providing the highest resolution surgical imaging data set available to date, enabling the development of next-generation computer-assisted surgical systems that can leverage the full benefits of 4K endoscopic imaging technology.

3. Discussion

Traditional fully laparoscopic methods (closed-cavity) were not feasible due to the size of the external camera and the limited space within the pig's abdomen. Instead, a large incision (open abdominal cavity) was used to insert the camera.

This open approach introduced certain constraints: the camera had to remain in the chest / abdominal opening, with restricted maneuverability compared to an endoscope inside a trocar. Consequently, the freedom to adjust the camera field of view is limited, and the surgical scene is sometimes occluded due to the line-of-sight issues caused by instruments or tissue folds.

The use of an off-the-shelf 4k camera provides optics larger than those in conventional endoscopes. Although the larger aperture size increases the diffraction limit and enables high optical resolution, it has constrained the depth of field (DOF) compared to endoscopes. Frames with large out-of-focus areas have been removed during quality control. For future work, we plan to include scenes captured using large-DoF cameras.

Dataset (Year)	Modality	Size	Native Res.	Primary Tasks	Licence / Access
ETIS-LaribPolyp (Silva et al., 2014)	Colonoscopy stills	196 imgs	1225×966p	polyp seg.	free-research
CVC-ClinicDB (Bernal et al., 2015)	Colonoscopy stills	612 imgs	384×288p	polyp seg.	academic-only
Cholec80 (Twinanda et al., 2016)	Laproscopic video	80 vids	854×480p	cls., det, seg	CC-BY 4.0
Kvasir (Pogorelov et al., 2017)	Gastroscopy stills	8 000 imgs	720p	cls., det.	CC-BY 4.0
Kvasir-SEG (Jha et al., 2019)	Gastroscopy stills	1 000 imgs	1000×1000p	polyp seg.	CC-BY 4.0
HyperKvasir (Borgli et al., 2020)	Gastroscopy video/stills	110 k imgs, 374 vids	720–1080p	cls., seg., QA	CC-BY 4.0
GIANA (Guo and Matuszewski, 2019)	Colonoscopy imgs/videos	300 imgs, 30 vids	384–1080p	det., seg.	challenge-EULA
EndoVis 2017 (Allan et al., 2019)	Robotic surgery video	15 seqs.	1920×1080p	tool seg., track	request-form
EndoVis 2018 (Allan et al., 2020)	Robotic surgery video	15 seqs.	1280×1024p	tool seg., track	request-form
HeiSURF (Wagner et al., 2023)	Cholecystectomy video	33 vids	720–1080p	cls., det	CC-BY 4.0
SurgVU (Zia et al., 2025)	Robotic surgery video	280 vids	1280×720p	seg, det, cls	CC-BY 4.0
SurgiSR4K (2025)*	Robotic surgery imgs/videos	**800 imgs, 50 vids**	**3840×2160p**	**SR, seg, det**	**CC-BY-NC-SA 4.0/request-form**

Table 1: Comparison of public endoscopic and surgical datasets. **SurgiSR4K** is our proposed dataset introduced in this work.

4. Potential Use Cases

SurgiSR4K enables a broad range of machine learning and computer vision applications in surgical imaging. Below we outline the primary use cases and their associated clinical domains.

4.1 Super-Resolution

Super-resolution is a critical enabler for enhancing legacy or bandwidth-limited surgical video streams, particularly in low-cost or resource-constrained settings. SurgiSR4K provides a unique benchmark for SR algorithms by offering perfectly aligned multi-resolution image triplets (270p, 540p, and native 4K, as shown in Figure 3). These triplets allow for supervised training of state-of-the-art models (e.g., SwinIR (Liang et al., 2021), Real-ESRGAN (Wang et al., 2021)) as well as quantitative evaluation using perceptual and fidelity metrics.

Clinical Domains: Telesurgery and remote surgical consultation, resource-limited healthcare settings, legacy endoscopic system enhancement, surgical video archival and retrospective analysis.

4.2 Monoscopic Depth Estimation

Native 4K frames retain subtle photometric cues such as specular highlights, sub-pixel texture flow, and shading that improve scale-ambiguous monocular depth networks such as MiDaS. The accompanying 5s clips (30 fps) enable temporal photometric self-supervision without requiring stereo rigs or structured light, providing a valuable benchmark for dense 3D reconstruction of flexible organs. In addition, the high-resolution depth information supports tissue tracking, enabling more accurate modeling of tissue deformation and interaction in real-time. Tissue tracking challenge (Schmidt et al., 2025) and method (Chen et al., 2025) emphasize the critical role of high-resolution depth estimation in precise tissue tracking. This is crucial to improve surgical precision and support the development of autonomous systems (Figure 5 c).

Clinical Domains: Laparoscopic navigation, autonomous surgical subtask execution.

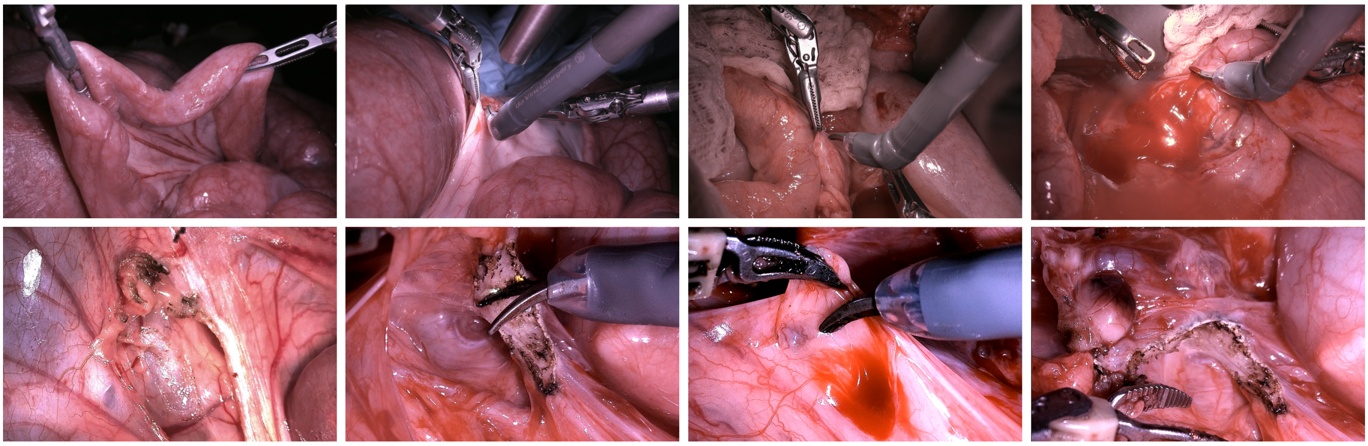

Figure 2: Example frames from the training dataset, showcasing various tools used in different scenarios. These frames highlight the diversity of situations included in the dataset.

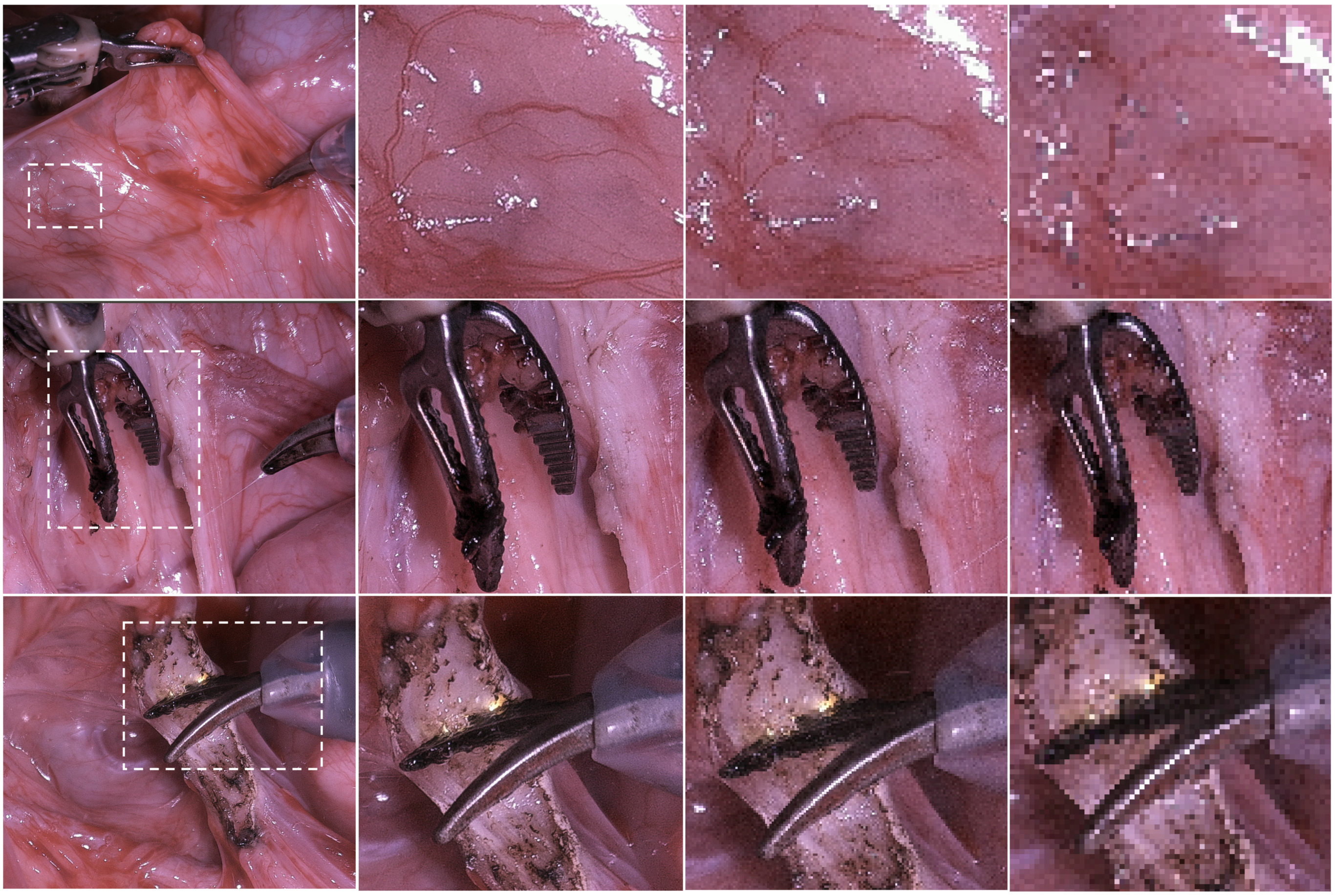

Figure 3: Comparison of image quality at different resolutions. From left to right: native 4K image (3840×2160), zoomed-in region of the 4K image, zoomed-in region of the downsampled 540p image (960×540), and zoomed-in region of the downsampled 270p image (480×270).

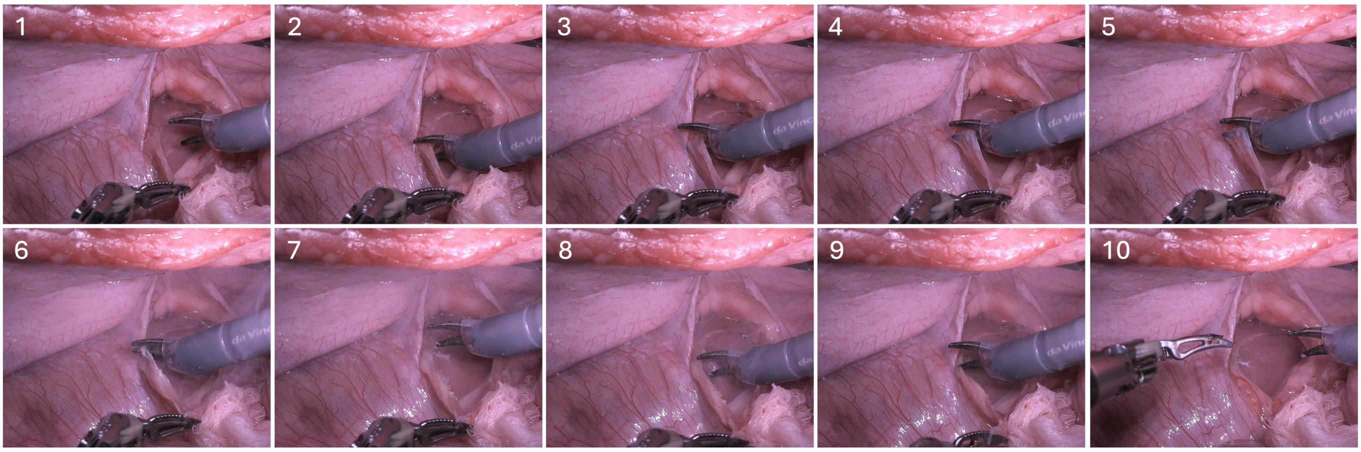

Figure 4: Frame-by-frame example of a 5-second video clip sampled at 2 fps, showing training data with a bipolar forceps instrument in a cauterizing motion.

4.3 Surgical Instrument Detection and Tracking

Fine instrument shafts (with 0 to 2mm) and grasper tips often collapse into a single pixel in 1080p footage. Thanks to its 4k resolution, our dataset exposes clear tool edges and metallic reflections, enabling detection of thin surgical instruments and pixel-level mask curation for semantic segmentation frameworks. Precise tool detection and localization are critical for downstream manipulation tasks such as suturing and tissue manipulations. The video nature of *SurgiSR4K* also allows researchers to go beyond frame-wise detection to *multi-instance tracking (Li et al. (2023))*. Accurate

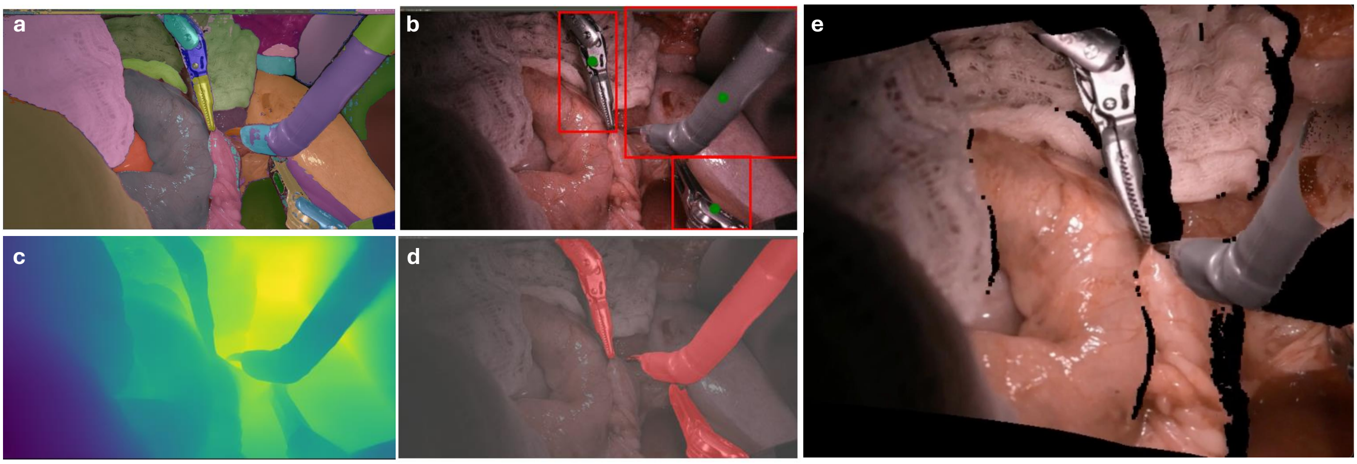

Figure 5: Examples of downstream applications: (a) instance segmentation (Ravi et al. (2024)), (b) surgical tool detection with bounding boxes (Liu et al. (2025)), (c) depth estimation (Bochkovskii et al. (2024)), (d) tool segmentation Ravi et al. (2024), and (e) 3D reconstruction. (Hu et al. (2025))

3D trajectories allow quantitative workflow analysis, objective skill assessment, context-aware autonomy, and safety monitoring in minimally invasive surgery (Figure 5 b).

Clinical Domains: Surgical workflow analysis and phase recognition, real-time safety monitoring and collision avoidance, surgeon training and simulation.

4.4 Instance Segmentation

Fine-grained instance segmentation in surgical scenes is particularly challenging due to occlusions, specular highlights, and visual similarity between instruments and surrounding tissue. With its native 4K resolution, *SurgiSR4K* allows accurate delineation of individual tool instances, even in overlapping or cluttered scenarios, by preserving fine spatial boundaries and textural cues. This enables the development of robust instance-aware models that can distinguish between multiple instruments of the same type. High-resolution instance masks also support pixel-level supervision for training advanced segmentation architectures, including transformer-based or graph-based models. Instance-level understanding is essential for precise robotic control, tool usage analysis, and temporal reasoning across frames, facilitating intelligent surgical assistance and real-time decision support (Figure 5 a).

Clinical Domains: Multi-arm robotic surgery coordination, instrument usage pattern analysis, surgical error detection.

4.5 3D Scene Reconstruction and Novel View Synthesis

High-resolution 4K video data is fundamental for accurate 3D scene reconstruction in surgical environments, where precise spatial understanding is critical for safe navigation and intervention planning. The native 4K resolution in SurgiSR4K preserves fine geometric details such as tissue surface topology, instrument-tissue contact points, and subtle anatomical landmarks that are essential for situational understanding. Unlike lower-resolution datasets, where fine details are lost to aliasing and compression artifacts, 4K footage retains submillimeter features necessary for dense point cloud generation and mesh reconstruction. The temporal consistency provided by 30 fps video sequences enables sophisticated SLAM (Simultaneous Localization and Mapping) approaches that can track camera motion while simultaneously building detailed 3D models of the surgical field. Furthermore, high-resolution depth cues—including specular highlights, texture gradients, and shading variations—support more accurate photometric stereo techniques and 3D Gaussian Splatting training for novel view synthesis. This enhanced 3D understanding directly benefits downstream applications such as augmented reality surgical guidance (Shu et al. (2023), Zhang et al. (2023)), preoperative planning overlay registration, and autonomous instrument navigation in complex anatomical environments (Figure 5 e).

Clinical Domains: Augmented reality surgical navigation, preoperative planning and simulation, digital twins.

4.6 Vision-Language Model

The diverse surgical scenes captured in SurgiSR4K, combined with its high visual fidelity, make it suitable for training and evaluating vision-language models (VLMs) in the surgical domain. Recent advances in surgical VLMs, such as SurgicalGPT Seenivasan et al. (2023) and GP-VLS Schmidgall et al. (2024), have demonstrated the potential of multimodal models for visual question answering and scene understanding in robotic surgery. The dataset's variety of instruments, anatomical contexts, and surgical events provides rich material for developing models capable of surgical scene understanding and interactive surgical assistance.

Clinical Domains: Automated surgical documentation, intraoperative decision support, surgical education and training systems, procedure summarization and reporting.

5. Resource Availability

5.1 Data/Code Location

The SurgiSR4K dataset is publicly available through Synapse. Complete access details, including the repository URL and DOI, are provided in the Data Availability section at the end of this manuscript.

5.2 Licensing

The SurgiSR4K dataset is released under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0) license. This permits users to share and adapt the dataset for non-commercial purposes, provided appropriate credit is given, any changes are indicated, and derivative works are distributed under the same license.

5.3 Ethical Considerations

This animal model dataset was collected in vivo from healthy Yorkshire swine under general anesthesia. No human subjects were involved in this study; therefore, Institutional Review Board (IRB) approval and informed consent processes were not applicable.

All procedures were approved and conducted in accordance with Intuitive Surgical's Institutional Animal Care and Use Committee (IACUC) protocol No. IACUC -2022-1200 (Title: [004] Use of Porcine Models for Robotic Surgical and Procedure Training). The study was conducted in alignment with the 3Rs principles: Refinement, Reduction, and Replacement—to ensure the ethical and humane treatment of animals. All procedures were supervised by qualified veterinary staff, and personnel involved were trained and approved under the relevant IACUC protocols.

This dataset represents original data collection and is not derived from any existing datasets; therefore, no compliance with prior licensing terms is required.

6. Methods

Dataset Component	Specification
Native Resolution	3840×2160 (4K)
Downsampled Resolutions	960×540 (540p), 480×270 (270p)
Resampling Method	Lanczos
Training Set	
Images	800 PNG files
Video clips	50 AVI files
Video duration	5 seconds each
Frame rate	30 fps
Test Set (Hidden)	
Images	300 PNG files
Video clips	50 AVI files
Surgical Instruments	
Monopolar curved scissors	✓
Force bipolar	✓
Bipolar forceps	✓
Grasping retractor	✓
Cadiere forceps	✓

Table 2: SurgiSR4K dataset composition and specifications.

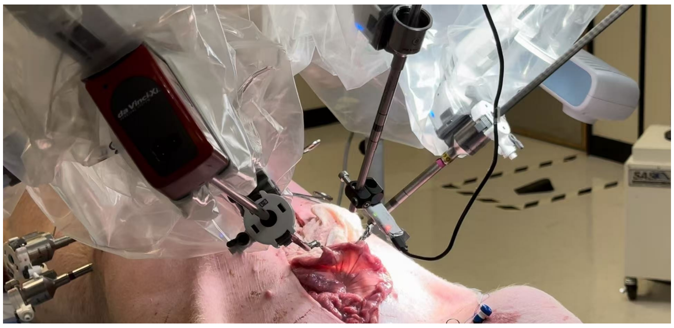

Figure 6: Experimental setup for SurgiSR4K data collection. Live porcine surgeries were performed using a Da Vinci Xi robot.

6.1 Experimental Setup

To construct the SurgiSR4K dataset, we performed a series of open-scene porcine lab surgeries using endoscopic imaging, designed to emulate realistic minimally invasive surgical conditions. These procedures were performed by experienced veterinarians on live anesthetized swines, using an open abdominal and thoracic approach to allow direct access for native 4K external camera capture. A Da Vinci Xi surgical robot performed interventions on internal organs during each session, ensuring authentic instrument motion and tissue interaction, as shown in Figure 6. Each recorded sequence spans several minutes and features diverse surgical events, including tool usage, organ manipulation, cautery smoke, bleeding, and tissue ablation, providing a rich source of visual complexity for downstream computer vision tasks (see Figure 4).

Equipment The imaging setup used a Sony IMX412 external camera module, manufactured by Sony Corporation, featuring a 12.3 MP CMOS sensor (Sony Starvis series), with an optical stack of 74.4 ° diagonal (60.2 ° horizontal).

Acquisition Settings Images and videos were recorded at a native 4K Ultra HD resolution of 3840×2160 pixels at 60 frames per second (fps). The camera was interfaced through CSI-2 to an NVIDIA Jetson TX2 Development Kit for real-time data handling and storage.

Subject Information

- Demographics: The dataset comprises adult pigs used under controlled experimental conditions. The animals were healthy and standardized by weight, anesthetized following institutional animal care guidelines.
- Cohort Description: The selection criteria ensured that the subjects were free of systemic diseases or anatomical anomalies that could affect surgical simulations. Multiple wet-lab procedures were performed to maximize the diversity of captured surgical scenarios.

6.2 Dataset Composition

Images and videos were captured at native 4K resolution (3840×2160), then downsampled to 540p (960×540) and 270p (480×270) using Lanczos resampling to support a range of super resolution tasks and multiscale training. The curated dataset includes 800 high-quality PNG images and 50 short video clips (AVI format), each 5 seconds in duration at 30 fps, selected for their diversity and visual richness (see Figure 2, 4). The data set features a variety of minimally invasive robotic surgical instruments, including monopolar curved scissors, bipolar forceps, grasping retractor, and cadiere forceps, improving its utility for tool recognition and interaction modeling tasks, as shown in Figure 7. An additional 300 images and 50 video clips have been reserved as a hidden test set for benchmarking and evaluation purposes.

7. Validation

All images and videos incorporated in the SurgiSR4K dataset were individually reviewed by an experienced Clinical Development Engineer (CDE) and an experienced veterinarian (VMD) to ensure clinical relevance and visual clarity for downstream applications.

Acknowledgments

This work was supported by Intuitive Surgical, Inc. We appreciate the valuable discussions with Max Allan and Xingtong Liu, and the support and encouragement from Kelvin (Fei) Chu, Qunming Peng, Ran Sendrovitz and Jonathan Halderman.

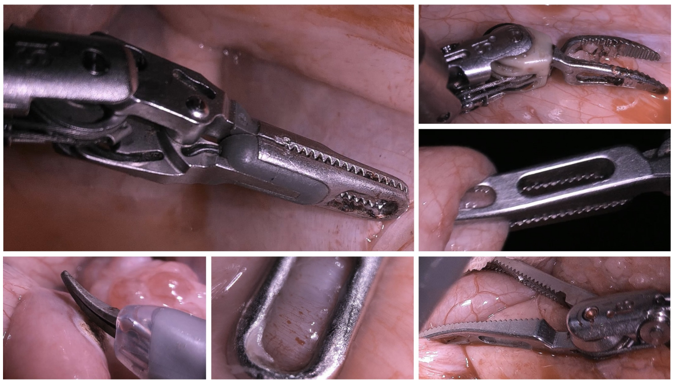

Figure 7: Example of various instruments included in the dataset, showcasing a range of tools used in different surgical or procedural contexts.

Ethical Standards

This study used porcine animal models only; no human subjects were involved. All procedures were approved by Intuitive Surgical's IACUC (Protocol No. IACUC-2022-1200) and conducted in accordance with the 3Rs principles.

Conflicts of Interest

We declare we don't have conflicts of interest.

Data availability

The SurgiSR4K dataset—including video clips and corresponding high-resolution images—is publicly available through our dedicated repository on Synapse:

- **Repository:** `https://www.synapse.org/Synapse:syn68756003`
- **DOI:** 10.7303/SYN68756003

The portal provides open access to the complete dataset, along with detailed metadata, project wiki, and supporting documentation.

References

M Abdelrahman, A Belramman, R Salem, and B Patel. Acquiring basic and advanced laparoscopic skills in novices using two-dimensional (2d), three-dimensional (3d) and ultra-high definition (4k) vision systems: a randomized control study. *International Journal of Surgery*, 53:333–338, 2018.

Max Allan, Alex Shvets, Thomas Kurmann, Zichen Zhang, Rahul Duggal, Yun-Hsuan Su, Nicola Rieke, Iro Laina, Niveditha Kalavakonda, Sebastian Bodenstedt, et al. 2017 robotic instrument segmentation challenge. *arXiv preprint arXiv:1902.06426*, 2019.

Max Allan, Satoshi Kondo, Sebastian Bodenstedt, Stefan Leger, Rahim Kadkhodamohammadi, Imanol Luengo, Felix Fuentes, Evangello Flouty, Ahmed Mohammed, Marius Pedersen, et al. 2018 robotic scene segmentation challenge. *arXiv preprint arXiv:2001.11190*, 2020.

Yasin Almalioglu, Mehmet Turan, Helder Araujo, Ender Konukoglu, Metin Sitti, and Mert R. Turan. Endol2h: Deep super-resolution for capsule endoscopy. *IEEE Transactions on Medical Imaging*, 39(12):4297–4309, 2020. .

Jorge Bernal, F Javier Sánchez, Gloria Fernández-Esparrach, Debora Gil, Cristina Rodríguez, and Fernando Vilariño. Wm-dova maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. *Computerized medical imaging and graphics*, 43:99–111, 2015.

Aleksei Bochkovskii, AmaÃĢl Delaunoy, Hugo Germain, Marcel Santos, Yichao Zhou, Stephan R Richter, and Vladlen Koltun. Depth pro: Sharp monocular metric depth in less than a second. *arXiv preprint arXiv:2410.02073*, 2024.

Hanna Borgli, Vajira Thambawita, Pia H Smedsrud, Steven Hicks, Debesh Jha, Sigrun L Eskeland, Kristin Ranheim Randel, Konstantin Pogorelov, Mathias Lux, Duc Tien Dang Nguyen, et al. Hyperkvasir, a comprehensive multi-class image and video dataset for gastrointestinal endoscopy. *Scientific data*, 7(1): 283, 2020.

Yuxin Chen, Zijian Wu, Adam Schmidt, and Septimiu E Salcudean. A-mfst: adaptive multi-flow sparse tracker for real-time tissue tracking under occlusion. *International Journal of Computer Assisted Radiology and Surgery*, pages 1–9, 2025.

Lesley B Conrad, Pedro T Ramirez, William Burke, R Wendel Naumann, Kari L Ring, Mark F Munsell, and Michael Frumovitz. Role of minimally invasive surgery in gynecologic oncology: an updated survey of members of the society of gynecologic oncology. *International Journal of Gynecological Cancer*, 25(6): 1121–1127, 2015.

Dikpal Fredrick. *EndoSLAM: Large-Scale Endoscopic Visual SLAM Dataset*. PhD thesis, Norwegian University of Science and Technology, 2022.

Hood Keng Christopher Goh, Yuk Hui Ng, and Dawn Tju Wei Teo. Minimally invasive surgery for head and neck cancer. *The lancet oncology*, 11(3):281–286, 2010.

Yun Bo Guo and Bogdan Matuszewski. Giana polyp segmentation with fully convolutional dilation neural networks. In *Proceedings of the 14th international joint conference on computer vision, imaging and computer graphics theory and applications*, pages 632–641, 2019.

Hiroyuki Harada, Yusuke Hirasawa, Takeshi Hachisuka, Masami Watanabe, Daisuke Kawano, Mikio Sakashita, and Yuichiro Tanaka. The effect on surgical skills of expert surgeons using 3d/hd and 2d/4k resolution monitors in laparoscopic phantom tasks. *Surgical Endoscopy*, 32(10):4228–4234, 2018. .

Wenbo Hu, Xiangjun Gao, Xiaoyu Li, Sijie Zhao, Xiaodong Cun, Yong Zhang, Long Quan, and Ying Shan. Depthcrafter: Generating consistent long depth sequences for open-world videos. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 2005–2015, 2025.

Debesh Jha, Pia H Smedsrud, Michael A Riegler, Pål Halvorsen, Thomas De Lange, Dag Johansen, and Håvard D Johansen. Kvasir-seg: A segmented polyp dataset. In *International conference on multimedia modeling*, pages 451–462. Springer, 2019.

Zhaoshuo Li, Hongchao Shu, Ruixing Liang, Anna Goodridge, Manish Sahu, Francis X Creighton, Russell H Taylor, and Mathias Unberath. Tatoo: vision-based joint tracking of anatomy and tool for skull-base surgery. *International journal of computer assisted radiology and surgery*, 18(7):1303–1310, 2023.

Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1833–1844, 2021.

Ziyu Liu, Zeyi Sun, Yuhang Zang, Xiaoyi Dong, Yuhang Cao, Haodong Duan, Dahua Lin, and Jiaqi Wang. Visual-rft: Visual reinforcement fine-tuning. *arXiv preprint arXiv:2503.01785*, 2025.

Michael J Mack. Minimally invasive and robotic surgery. *Jama*, 285(5):568–572, 2001.

Milton V Marshall, John C Rasmussen, I-Chih Tan, Melissa B Aldrich, Kristen E Adams, Xuejuan Wang, Caroline E Fife, Erik A Maus, Latisha A Smith, and Eva M Sevick-Muraca. Near-infrared fluorescence imaging in humans with indocyanine green: a review and update. *Open surgical oncology journal (Online)*, 2(2):12, 2010.

Takeshi Masuda, Ryusuke Sagawa, Ryo Furukawa, and Hiroshi Kawasaki. View synthesis of endoscope images by monocular depth prediction and gaussian splatting. In *2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 1–6, 2024. .

Roberto Montalti, Federico Tomassini, Stéphanie Laurent, Peter Smeets, Marc De Man, Karen Geboes, Louis J Libbrecht, and Roberto I Troisi. Impact of surgical margins on overall and recurrence-free survival in parenchymal-sparing laparoscopic liver resections of colorectal metastases. *Surgical endoscopy*, 29:2736–2747, 2015.

Daniel T Oberlin, Andrew S Flum, Jeremy D Lai, and Joshua J Meeks. The effect of minimally invasive prostatectomy on practice patterns of american urologists. In *Urologic Oncology: Seminars and Original Investigations*, volume 34, pages 255–e1. Elsevier, 2016.

Konstantin Pogorelov, Kristin Ranheim Randel, Carsten Griwodz, Sigrun Losada Eskeland, Thomas de Lange, Dag Johansen, Concetto Spampinato, Duc-Tien Dang-Nguyen, Mathias Lux, Peter Thelin Schmidt, et al. Kvasir: A multi-class image dataset for computer aided gastrointestinal disease detection. In *Proceedings of the 8th ACM on Multimedia Systems Conference*, pages 164–169, 2017.

Federico Puccetti, Laura Licari, Elisa Bertocchi, Roberto Berchiolli, Alberto Falsetti, and Saverio Pucciarelli. Impact of imaging magnification on colorectal surgery: A matched analysis of a single tertiary centre. *Techniques in Coloproctology*, 27(11):1057–1063, 2023. .

Nikhila Ravi, Valentin Gabeur, Yuan-Ting Hu, Ronghang Hu, Chaitanya Ryali, Tengyu Ma, Haitham Khedr, Roman Rädle, Chloe Rolland, Laura Gustafson, et al. Sam 2: Segment anything in images and videos. *arXiv preprint arXiv:2408.00714*, 2024.

Samuel Schmidgall, Joseph Cho, Cyril Zakka, and William Hiesinger. Gp-vls: A general-purpose vision language model for surgery. *arXiv preprint arXiv:2407.19305*, 2024.

Adam Schmidt, Mert Asim Karaoglu, Soham Sinha, Mingang Jang, Ho-Gun Ha, Kyungmin Jung, Kyeongmo Gu, Ihsan Ullah, Hyunki Lee, Jonáš Šerỳch, et al. Point tracking in surgery–the 2024 surgical tattoos in infrared (stir) challenge. *arXiv preprint arXiv:2503.24306*, 2025.

Lalithkumar Seenivasan, Mobarakol Islam, Gokul Kannan, and Hongliang Ren. Surgicalgpt: End-to-end language-vision gpt for visual question answering in surgery. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 281–290. Springer, 2023.

Hongchao Shu, Ruixing Liang, Zhaoshuo Li, Anna Goodridge, Xiangyu Zhang, Hao Ding, Nimesh Nagururu, Manish Sahu, Francis X Creighton, Russell H Taylor, et al. Twin-s: a digital twin for skull base surgery. *International journal of computer assisted radiology and surgery*, 18(6):1077–1084, 2023.

Juan Silva, Aymeric Histace, Olivier Romain, Xavier Dray, and Bertrand Granado. Toward embedded detection of polyps in wce images for early diagnosis of colorectal cancer. *International journal of computer assisted radiology and surgery*, 9:283–293, 2014.

Xiang Song, Mingming Shi, Yifan Li, Yi Xu, and Zhiwei Liu. Deformable transformer for endoscopic video super-resolution. *Biomedical Signal Processing and Control*, 77:103827, 2022. .

Andru P Twinanda, Sherif Shehata, Didier Mutter, Jacques Marescaux, Michel De Mathelin, and Nicolas Padoy. Endonet: a deep architecture for recognition tasks on laparoscopic videos. *IEEE transactions on medical imaging*, 36(1):86–97, 2016.

Martin Wagner, Beat-Peter Müller-Stich, Anna Kisilenko, Duc Tran, Patrick Heger, Lars Mündermann, David M Lubotsky, Benjamin Müller, Tornike Davitashvili, Manuela Capek, et al. Comparative validation of machine learning algorithms for surgical workflow and skill analysis with the heichole benchmark. *Medical Image Analysis*, 86:102770, 2023.

Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1905–1914, 2021.

Zhen Xu, Sida Peng, Haotong Lin, Guangzhao He, Jiaming Sun, Yujun Shen, Hujun Bao, and Xiaowei Zhou. 4k4d: Real-time 4d view synthesis at 4k resolution. In *Proceedings of the IEEE/CVF conference* *on computer vision and pattern recognition*, pages 20029–20040, 2024.

Tian Zhang and Jingru Yang. Transformer with hybrid attention mechanism for stereo endoscopic video super-resolution. *Symmetry*, 15(10):1947, 2023. .

Xiaorui Zhang, Andreas Keller, Mehran Armand, and Alejandro Martin Gomez. Feasibility study of using augmented mirrors for alignment task during orthopaedic procedures in mixed reality. In *2023 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pages 650–651. IEEE, 2023.

Aneeq Zia, Max Berniker, Rogerio Nespolo, Conor Perreault, Ziheng Wang, Benjamin Mueller, Ryan Schmidt, Kiran Bhattacharyya, Xi Liu, and Anthony Jarc. Surgical visual understanding (surgvu) dataset. *arXiv preprint arXiv:2501.09209*, 2025.