Proceedings of the 4th International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/2026.37195

Levels of Autonomy in Robot-Assisted Laparoscopic Surgery: Current Capabilities and Evolutionary Pathways

Qingxuan Guo

The Grainger College of Engineering, University of Illinois Urbana-Champaign, Urbana, USA
qingxuan_guo@outlook.com

Abstract. Robotic laparoscopic surgery is moving from conventional teleoperation toward greater autonomy. Previous studies usually only examined individual parts of this transition. The technologies enabling autonomous laparoscopic surgery are less often discussed as one integrated system. This review examines the system through three components: hardware infrastructure, environmental perception, and decision-making. The hardware discussion covers instrument actuation, vision and force sensing, and various robotic platforms. The perception section follows proprioceptive, visual and force data as they are converted into a unified representation of the surgical state. It focuses on 3D scene reconstruction, instrument segmentation, workflow recognition, and multimodal fusion. The decision and control section discusses classical control and data-driven methods, especially the hierarchical imitation learning and reinforcement learning for long-horizon surgical tasks. The review concludes with the main barriers to clinical translation, including the sim-to-real gap, deformable tissue modeling, and safety and regulatory requirements. These areas together show how embodied intelligence may influence future autonomous surgery in the operating room.

Keywords: Surgical autonomy, robot-assisted laparoscopic surgery, multimodal perception, imitation learning, reinforcement learning

1. Introduction

Robotic surgery has become an important research area in robotics. During the past two decades, surgical robotic systems have entered clinical practice in orthopedics, urology, and other minimally invasive procedures. Robot-assisted minimally invasive surgery (RAMIS) is now their main application area. Laparoscopic surgery is also a promising setting for surgical autonomy. Its procedures are relatively standardized, and its workflow has an inherent hierarchical structure.

According to the Levels of Autonomy in Surgical Robotics (LASR) taxonomy [1], surgical robots are classified from Level 1 to Level 5. Figure 1 summarizes the characteristics of each level. Most surgical robots cleared by the United States Food and Drug Administration (FDA) remain at Level 1 (Robot Assistance). At this level, the robot is a teleoperated platform under the surgeon's direct control. Advances in artificial intelligence are now pushing robotic laparoscopic surgery beyond simple teleoperation toward higher levels of autonomy.

DOI: 10.54254/2755-2721/2026.37195

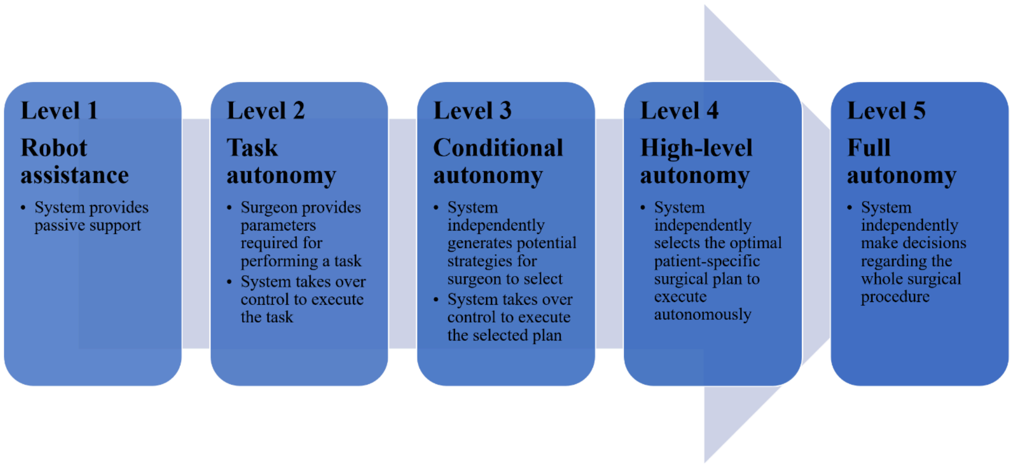

Figure 1. Levels of Autonomy in Surgical Robotics (LASR) taxonomy [1] (picture credit: original)

Most existing reviews still focus on conventional robot-assisted laparoscopic surgery. Very recent works on autonomous laparoscopic robots deserve more attention. The field also lacks a system-level synthesis of the technologies that enable autonomy.

This review organizes those technologies into three core pillars. Figure 2 shows their overall structure and their relationships. The first pillar is the hardware foundation. The second is the perception pipeline that includes information processing methods for handling raw data from upstream hardware systems. The third is decision-making and control for complex autonomous laparoscopic procedures and environments. This organization provides an integrated view of current progress and future development.

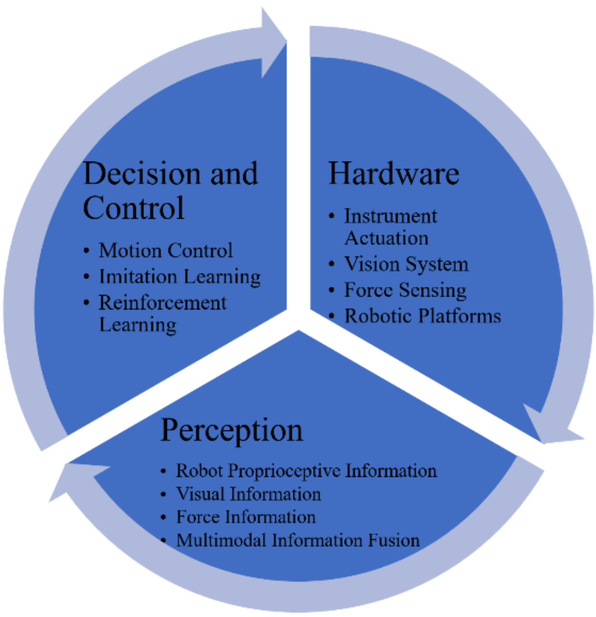

Figure 2. Core pillars enabling autonomous robotic laparoscopic surgery (picture credit: original)

DOI: 10.54254/2755-2721/2026.37195

2. Hardware foundations enabling autonomous laparoscopic surgery

Autonomous robotic laparoscopic surgery first relies on hardware systems and platforms. Instrument actuation, vision, and force sensing determine how the robot moves in confined spaces and how it perceives the surgical environment.

2.1. Hardware systems

Laparoscopic instruments enter the abdominal cavity through trocars. Their motion must satisfy the Remote Center of Motion (RCM) constraint to limit loading at the abdominal-wall entry point. The RCM may be built into the mechanism of an integrated single-cart system. In modular systems with independent arms, it can instead be enforced virtually through the controller. Once this constraint is satisfied, wristed instruments provide additional dexterity inside the abdominal cavity. These wrists are generally rigid articulated or continuum mechanisms. Rigid wrists use serial or parallel links and offer stiff, precise motion transmission. Continuum wrists bend through distributed deformation and can reach around confined anatomy. Chen et al. developed the modular SHURUI platform as one example. Its dual-continuum instruments use tight-bending wrists and support multiport, hybrid-port, and single-port configurations [2].

The vision system provides the main input for an autonomous laparoscopic robot. A stereo endoscope with illumination is the most common hardware solution for capturing the 3D intraoperative scene. Near-Infrared (NIR) fluorescence imaging can add information about vascular and biliary structures.

Force sensing provides information about the interaction between the robot and biological tissue. Four main mechanisms are used to convert force into electrical signals. Strain-gauge sensors are mature and easy to miniaturize. They can be integrated into instruments individually or arranged in parallel for multi-axis measurements. Capacitive sensors provide high resolution and low drift. They are usually made into compact multi-axis designs. Piezoelectric sensors are sensitive and have a wide bandwidth, but they are not suitable to measure static force. Optical fiber sensors are also easy to miniaturize and can be embedded directly into instruments.

Actuators and sensors alone are not enough for an autonomous laparoscopic robot. The system also needs an onboard computing device that can process several data streams in real time. Robot kinematics, stereo images, and force measurements arrive at different frequencies and must be synchronized. All the following information-processing procedures may also run at the same time. These tasks need massive computing resources and memory bandwidth. External computing resources can handle some calculations, but network latency and connection failures could violate the strict safety requirements for surgical practice. The onboard design must also account for task scheduling, deterministic data exchange, power consumption, and heat dissipation. Computing hardware and software should therefore be designed together. The goal is robustness within the space, power, and cooling limits of the surgical platform.

2.2. Robotic platforms

These hardware requirements define the physical basis of autonomous laparoscopic surgery. Robotic platforms implement them through different configurations and designs. Most platforms were initially developed for mainly surgeon-controlled teleoperation. Table 1 compares representative commercial laparoscopic robotic systems and their main technical and clinical features. Commercial

DOI: 10.54254/2755-2721/2026.37195

platforms put clinical safety and reliability as top priorities. Medical regulations therefore limit access to low-level control, which reduces their potential as native testbeds for autonomy.

Open research platforms address this limitation. RAVEN, for example, is a telesurgery research platform with a modular design and open software architecture. Multiple institutions have since adopted it for research. The da Vinci Research Kit (dVRK) takes a different approach. It converts a mature clinical system into a configurable platform and gives researchers full-stack access to the control interfaces [3].

Table 1. Overview of hardware features and autonomy levels across representative commercial laparoscopic robotic platforms

Robotic Platforms	Actuation System	Visual System	Force Sensing System	Level of Autonomy
Da Vinci 5 by Intuitive Surgical	Mechanical RCM, Rigid Articulated	Stereo + NIR	Distal Instrument	Level 1
SR-ENS-600 by SHURUI	Mechanical RCM, Continuum	Stereo	None	Level 1
Senhance by Asensus Surgical	Mechanical RCM, Rigid Articulated	Stereo	None	Level 2
Versius by CMR Surgical	Virtual RCM, Rigid Articulated	Stereo	None	Level 1

3. Perception and information processing

Autonomous laparoscopic robots receive signals from their hardware to support decision-making and control. Raw sensor measurements are heterogeneous and have limitations. Perception and information processing must convert these low-level signals into a unified representation of the surgical state.

Robot proprioception includes actuator feedback and robot state information. It supports motion control and provides the robot's internal status. For example, in cable-driven mechanisms widely adopted by robotic laparoscopic surgery, nonlinear behavior reduces the accuracy of mathematical geometry models. Data-driven methods can compensate for these effects and improve motion accuracy. Bai et al. developed an LSTM kinematic model trained on task-specific data. This model improved path-following accuracy [4].

Visual perception has to do more than just identify objects. It must understand the parts of the surgical scene that matter for the current task. The main areas are 3D scene reconstruction, instrument segmentation, and workflow recognition. For 3D scene reconstruction, Zha et al. introduced EndoSurf. It represents deformable tissue geometry and appearance with signed-distance and radiance fields [5]. Free-SurGS uses 3D Gaussian Splatting to improve reconstruction efficiency [6]. SurgicalGS supports real-time rendering of dynamic soft tissue [7]. EndoSD-SLAM achieves online monocular tracking and dense reconstruction in dynamic laparoscopic scenes [8]. Figure 3 compares the pipeline of these frameworks. Instrument segmentation turns endoscopic images into structured descriptions of surgical tools. Relevant information includes tool presence, type, tip pose, and interaction state. Recent methods are moving beyond binary masks toward fine-grained and standardized representations. These representations remain reliable even when instruments are partly obstructed or only limited labeled data are available. Workflow recognition deals with the temporal dimension of perception. It identifies surgical phases, tasks, subtasks, and actions from video.

DOI: 10.54254/2755-2721/2026.37195

Workflow recognition models do not treat every frame in isolation. They use procedure transitions and longer context to estimate stages and provide task-level information for downstream decisions. The three branches therefore describe different parts of the laparoscopic scene: environment geometry, operative entities, and temporal state.

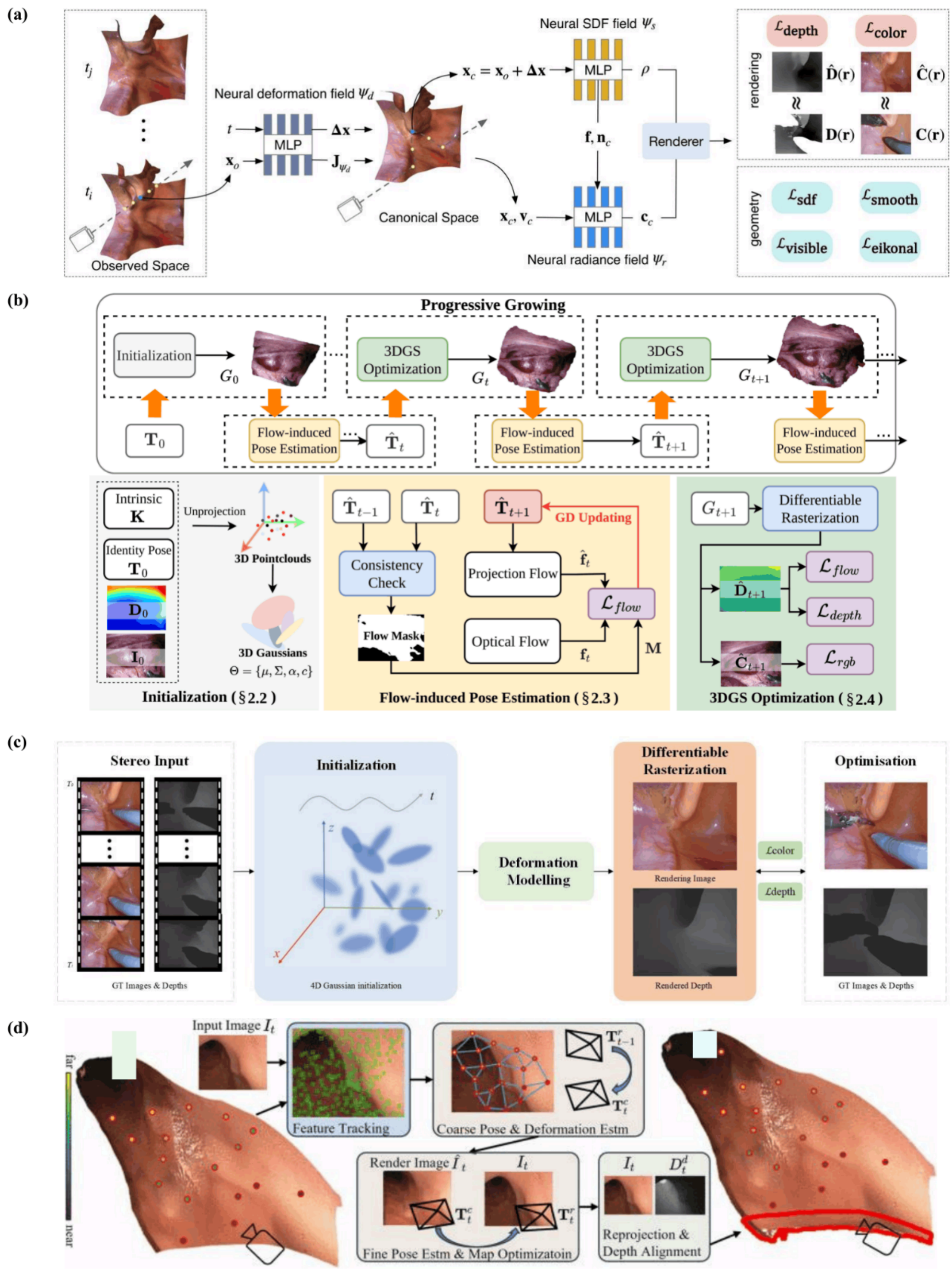

Figure 3. Overview of pipelines and results of 3D scene reconstruction frameworks: (a) EndoSurf [5]; (b) Free-SurGS [6]; (c) SurgicalGS [7]; (d) EndoSD-SLAM [8]

Force information can be recovered through model-based estimation from robot states or vision-based inference from tissue deformation. These complementary approaches support safer tissue contact when direct force sensing is unavailable.

DOI: 10.54254/2755-2721/2026.37195

Robot states, visual information, and force feedback change simultaneously during a laparoscopic procedure. No single modality can fully describe the entire surgical environment. Each modality captures a different part of the surgical state. Proprioception describes the robot's internal condition. Vision provides external geometry, while force signals describe physical interaction. A reliable state estimate must align these measurements despite differences in sampling rate, uncertainty, and form. Researches like DaFoEs combine mixed vision-haptic datasets with a vision-state architecture. It was used to study whether force-estimation models generalize across domains [9]. The fused result then becomes the input to decision-making and control.

4. Decision and control

Laparoscopic procedures generally follow a structured and standardized workflow with phases, tasks, subtasks, and individual actions. A laparoscopic cholecystectomy, for example, may proceed through tissue exposure, dissection, clipping, and specimen retrieval. Each phase can be divided further into actions such as grasping, retracting, positioning, and cutting. This organization resembles the hierarchical decision frameworks used in autonomous robotics. Many laparoscopic tasks also require two instruments to work together. One instrument may grasp, retract, or stabilize tissue while the other dissects, cuts, or sutures. An autonomous robot therefore has to do more than reproduce precise motion. It must account for tissue response and coordinate both instruments as the scene changes.

Once a unified surgical state is available, the robot must turn that information into motion and task decisions. Current work mainly covers motion control, imitation learning, and reinforcement learning. Research is also beginning to move toward more general forms of autonomy.

Motion-control studies cover autonomous endoscope manipulation, visual servoing, and control under the remote center-of-motion (RCM) constraint. Autonomous camera control can combine deep-learning-based visual perception with robot kinematics to adjust the laparoscopic field of view. Li et al. proposed a constraint-consistent dynamic controller for the manipulator. The controller preserves the RCM constraint and improves the torque profiles' smoothness [10]. Saeidi et al. demonstrated a system-level example in STAR. Their system combines tissue tracking, autonomous suture planning, and closed-loop control for supervised autonomous laparoscopic intestinal anastomosis in porcine models [11].

Imitation learning allows a surgical robot to learn manipulation skills from expert demonstrations collected by teleoperation. It is therefore widely used in data-driven surgical autonomy. On the dVRK platform, Kim et al. introduced the Surgical Robot Transformer (SRT), which uses transformer-based imitation learning and a relative action formulation for tissue manipulation, needle pickup and handover, and knot tying [12]. In subsequent work, the same group introduced the Hierarchical Surgical Robot Transformer (SRT-H), which incorporates language-guided task planning to support long-horizon manipulation in autonomous laparoscopic surgery [13]. In Figure 4(a), the high-level (HL) policy generates a language instruction from visual observations. The low-level (LL) policy combines that instruction with images to generate Cartesian motion. Figure 4(b) shows the implementation. Swin-T and a transformer decoder form the HL pathway. In the LL pathway, frozen DistilBERT embeddings condition EfficientNet features through FiLM. A second transformer decoder then predicts an action sequence as incremental changes in position and orientation.

A single end-to-end policy can struggle with the duration and complexity of a surgical procedure. Errors made early in a long sequence may accumulate and cause the task to fail. Hierarchical imitation learning separates decision-making and control procedures into high-level planning and

DOI: 10.54254/2755-2721/2026.37195

low-level skill execution. This separation can improve scalability and make the decision process easier to interpret. It may also connect foundation models and robotic execution together. Language or semantic instructions focus on task-level reasoning, while learned policies handle fine-grained physical manipulation.

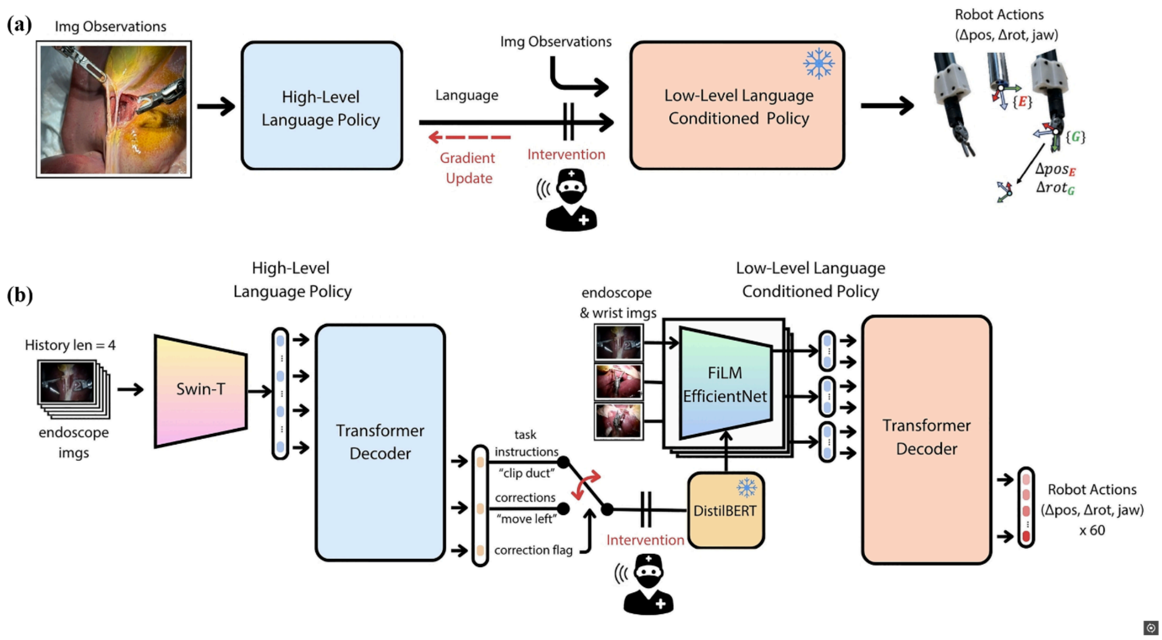

Figure 4. Model overview and architecture of SRT-H [13]

Reinforcement learning develops control policies through repeated interaction with laparoscopic surgical environment. This offers a route to autonomous skill acquisition, but direct exploratory attempts based on a physical surgical robot are expensive. Xu et al. developed SurRoL to support this research [14]. It is an open reinforcement-learning platform for laparoscopic robotic surgery and provides standardized simulation environments. Algorithms can be trained and evaluated in simulation before transfer to a physical system. Long et al. later proposed a framework for surgical embodied intelligence [15]. Figure 5(a) connects SurRoL with a dVRK master tool manipulator. Figure 5(b) shows the shared-control training scheme. An RL reference path is converted into haptic guidance. In the peg-transfer example, Figure 5(c) compares instrument motion, path deviation, and feedback force over time. Assistance is activated when the deviation becomes large. Figure 5(d) reports the user study and uses post-training completion time to evaluate learning efficiency. Simulation is especially important in surgery because unrestricted trial and error is not acceptable in the operating room. Conventional RL also requires many interactions before convergence. Direct optimization on delicate biological tissue is therefore impractical. A standardized virtual environment moves most training away from physical hardware. The agent can encounter and correct failures without placing patients or tissue at risk. Surgical embodied intelligence may extend this idea beyond isolated motor skills. It combines multisensory perception with dynamic reasoning, which may help the robot adapt its control strategy to anatomical differences among patients.

DOI: 10.54254/2755-2721/2026.37195

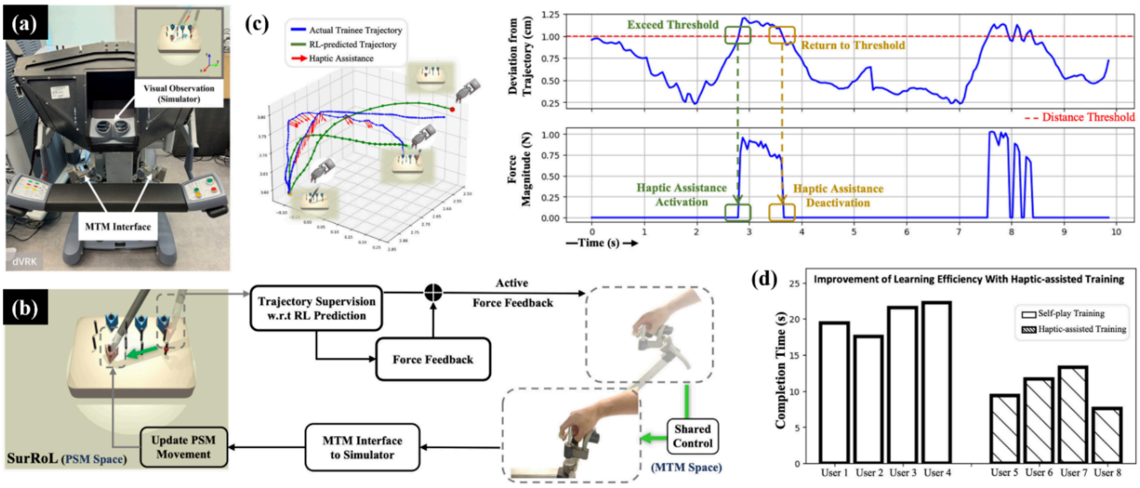

Figure 5. Schematic overview and evaluation outcomes of the SurRoL human-machine interface [15]

Clinical deployment raises challenges in robust perception, data availability, the measurement of safety and reliability, regulation, liability, and workflow integration. Schmidgall et al. summarized these issues [16]. Recent work also proposes foundation models as a basis for future autonomous surgical robots. Other studies describe reasoning-enabled AI copilots for task-level or supervised autonomy rather than high-level autonomous surgery. The field is therefore moving gradually from isolated skill learning toward broader embodied intelligence based on foundation models, world models, and integrated decision-making.

5. Conclusion

This review examined autonomous laparoscopic surgery through three connected components: specialized hardware, multimodal perception, and decision-making and control. The hardware base includes RCM-constrained manipulators, stereo vision, force sensors, and several types of robotic platforms. Proprioceptive, visual, and force data are then converted into a unified state representation. Visual perception includes 3D reconstruction, instrument segmentation, and workflow recognition. Force may be estimated across modalities when direct sensing is limited. The resulting state representation supports visual servoing, hierarchical imitation learning, and simulation-based reinforcement learning. These methods allow increasingly complex bimanual laparoscopic tasks. Clinical translation still faces three main obstacles: the sim-to-real gap, deformable-tissue modeling, and strict safety and regulatory requirements. Progress will also require smaller force sensors, real-time perception of dynamic tissue, and more general surgical embodied intelligence. Foundation and world models may improve generalization and task-level reasoning. Their outputs, however, must remain within verifiable safety constraints. Standardized safety evaluations will be needed to connect algorithm development with clinical validation and regulation.

References

[1] Lee, A., Baker, T. S., Bederson, J. B., & Rapoport, B. I. (2024). Levels of autonomy in FDA-cleared surgical robots: A systematic review. *npj Digital Medicine, 7*(1), Article 103.

DOI: 10.54254/2755-2721/2026.37195

[2] Chen, Y., Zhang, C., Wu, Z., Zhao, J., Yang, B., Huang, J., Luo, Q., Wang, L., & Xu, K. (2022). The SHURUI system: A modular continuum surgical robotic platform for multiport, hybrid-port, and single-port procedures. *IEEE/ASME Transactions on Mechatronics, 27*(5), 3186–3197.

[3] D'Ettorre, C., Mariani, A., Stilli, A., Rodriguez y Baena, F., Valdastri, P., Deguet, A., Kazanzides, P., Taylor, R. H., Fischer, G. S., DiMaio, S. P., Menciassi, A., & Stoyanov, D. (2021). Accelerating surgical robotics research: A review of 10 years with the da Vinci Research Kit. *IEEE Robotics & Automation Magazine, 28*(4), 56–78.

[4] Bai, W., Cursi, F., Guo, X., Huang, B., Lo, B., Yang, G.-Z., & Yeatman, E. M. (2022). Task-based LSTM kinematic modeling for a tendon-driven flexible surgical robot. *IEEE Transactions on Medical Robotics and Bionics, 4*(2), 339–342.

[5] Zha, R., Cheng, X., Li, H., Harandi, M., & Ge, Z. (2023). EndoSurf: Neural surface reconstruction of deformable tissues with stereo endoscope videos. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2023* (pp. 13–23). Springer.

[6] Guo, J., Wang, J., Kang, D., Dong, W., Wang, W., & Liu, Y.-H. (2024). Free-SurGS: SfM-free 3D Gaussian splatting for surgical scene reconstruction. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2024* (pp. 350–360). Springer.

[7] Chen, J., Zhang, X., Hoque, M. I., Vasconcelos, F., Stoyanov, D., Elson, D. S., & Huang, B. (2025). SurgicalGS: Dynamic 3D Gaussian splatting for accurate robotic-assisted surgical scene reconstruction. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2025* (pp. 572–582). Springer.

[8] Wang, K., Yang, C., Lin, R., Yang, X., & Shen, W. (2025). EndoSD-SLAM: Real-time deformable endoscopic SLAM via sparse-dense hybrid representation. In *2025 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (pp. 4214–4219). IEEE.

[9] De Iturrate Reyzabal, M., Chen, M., Huang, W., Ourselin, S., & Liu, H. (2024). DaFoEs: Mixing datasets towards the generalization of vision-state deep-learning force estimation in minimally invasive robotic surgery. *IEEE Robotics and Automation Letters, 9*(3), 2527–2534.

[10] Li, Y., Sadeghian, H., Yang, Z., Le Mesle, V., & Haddadin, S. (2026). RCM constraint-consistent dynamic control in surgical robots. In *2026 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE.

[11] Saeidi, H., Opfermann, J. D., Kam, M., Wei, S., Leonard, S., Hsieh, M. H., Kang, J. U., & Krieger, A. (2022). Autonomous robotic laparoscopic surgery for intestinal anastomosis. *Science Robotics, 7*(62), Article eabj2908.

[12] Kim, J. W., Zhao, T. Z., Schmidgall, S., Deguet, A., Kobilarov, M., Finn, C., & Krieger, A. (2025). Surgical Robot Transformer (SRT): Imitation learning for surgical tasks. In *Proceedings of the 8th Conference on Robot Learning (Vol. 270)* (pp. 130–144). Proceedings of Machine Learning Research.

[13] Kim, J. W., Chen, J.-T., Hansen, P., Shi, L. X., Goldenberg, A., Schmidgall, S., Scheikl, P. M., Deguet, A., White, B. M., Tsai, D. R., Cha, R. J., Jopling, J., Finn, C., & Krieger, A. (2025). SRT-H: A hierarchical framework for autonomous surgery via language-conditioned imitation learning. *Science Robotics, 10*(104), Article eadt5254.

[14] Xu, J., Li, B., Lu, B., Liu, Y.-H., Dou, Q., & Heng, P.-A. (2021). SurRoL: An open-source reinforcement learning centered and dVRK compatible platform for surgical robot learning. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (pp. 1821–1828). IEEE.

[15] Long, Y., Lin, A., Kwok, D. H. C., Zhang, L., Yang, Z., Shi, K., Song, L., Fu, J., Lin, H., Wei, W., Chen, K., Chu, X., Hu, Y., Yip, H. C., Chiu, P. W. Y., Kazanzides, P., Taylor, R. H., Liu, Y., Chen, Z., Wang, Z., Au, S. K. W., & Dou, Q. (2025). Surgical embodied intelligence for generalized task autonomy in laparoscopic robot-assisted surgery. *Science Robotics, 10*(104), Article eadt3093.

[16] Schmidgall, S., Opfermann, J. D., Kim, J. W., & Krieger, A. (2025). Will your next surgeon be a robot? Autonomy and AI in robotic surgery. *Science Robotics, 10*(104), Article eadt0187.