Proceedings of ICBioMed 2026 Symposium: Biomedical Engineering and Public Health: Real-Time Wearables for Disease Detection
DOI: 10.54254/2753-8818/2026.ES36570

Research Progress on Visual Navigation and Control Methods for Surgical Robots

Xinmin Jiang

School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai, China
jjjjjjjxm@sjtu.edu.cn

Abstract. In the last two decades, surgical robots have moved from specialized platforms to widely used tools in minimally invasive surgery. Their performance increasingly depends on visual navigation, because intraoperative images connect tissue perception, spatial localization, and motion control. Recent work has focused on how cameras, ultrasound, CT/MRI-derived models, optical tracking, and learning-based image analysis can be combined under the time pressure and uncertainty of the operating room. The discussion covers visual perception, multimodal imaging, segmentation, target recognition, three-dimensional reconstruction, image registration, real-time tracking, and vision-guided control in surgical robotics. Particular attention is given to the technical bottlenecks created by occlusion, specular reflection, tissue deformation, limited field of view, calibration drift, and data scarcity. The reviewed evidence shows that visual navigation is most useful when perception outputs are tightly linked to registration accuracy, safety constraints, and closed-loop control rather than treated as isolated image-processing results. Future progress will depend on more robust multimodal fusion, clinically reliable validation, safer intelligent control, and better integration of visual and haptic feedback in dynamic surgical scenes.

Keywords: surgical robotics, visual navigation, intraoperative imaging, multimodal fusion

1. Introduction

During the past two to three decades, minimally invasive surgery and advances in information technology have jointly reshaped the development of surgical robotic systems. Thousands of robotic platforms are now used in clinical environments worldwide. The da Vinci system, for example, has been applied in more than ten million procedures, showing that robot-assisted surgery has moved beyond proof-of-concept use and into large-scale clinical practice.

Visual navigation is one of the technical foundations that allows these systems to sense the surgical field and support intraoperative decisions. Cameras and medical images provide information about tissues, instruments, and target anatomy; image-processing algorithms then translate these signals into environment models, localization results, registration maps, and planned paths. In this way, visual feedback becomes a practical bridge between the surgical scene and robotic motion.

The operating room, however, is a difficult visual environment. Illumination changes, blood, smoke, fluid, and specular reflection can blur or distort images. Soft tissues move with respiration,

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0
(https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-8818/2026.ES36570

heartbeat, and tool contact, so anatomical targets rarely remain fixed. At the same time, navigation algorithms must satisfy strict requirements for speed, accuracy, and safety. A method that performs well on static images may fail once deformation, occlusion, and narrow fields of view appear together.

A further issue is that visual navigation has to operate inside a clinical workflow, not only inside a technical pipeline. Camera cleaning, trocar placement, surgeon preference, patient positioning, and the availability of preoperative imaging all influence whether a visual algorithm can be used at the right moment. Reports on autonomous and AI-enabled robotic systems show that autonomy remains limited in cleared devices, and this limitation is partly related to the difficulty of proving that perception, planning, and control remain safe under clinical variation [1, 2].

Against this background, the following review discusses recent progress in visual navigation and control for surgical robots, with visual assistance and autonomy treated as coupled problems rather than separate engineering tracks [1, 3].

2. Visual perception methods for surgical robots

2.1. Intraoperative imaging and multimodal perception

Surgical robots acquire information from several imaging sources, and each source captures only part of the operative reality. Direct imaging methods can reveal skeletal structures, lesions, or three-dimensional anatomy, but many of them are intermittent rather than continuous. Because scanning time and radiation exposure are limiting factors, CT or related modalities are usually introduced at selected procedural stages rather than throughout the entire operation.

The role of visual aids also differs by robotic configuration. A camera-guided positioning arm, a teleoperated laparoscopic platform, and a semi-autonomous orthopedic milling system all use visual information, but the information enters different parts of the control chain. Some systems use images mainly for display and orientation; others use them to update robot pose, constrain motion, or trigger automated subtasks. The reason why this distinction matters is that, in a display-only mode, the same segmentation or registration error may be innocuous, but become consequential when it directly modifies robot motion [3].

Indirect imaging methods involve preoperative 3D models along with implanted, attached or visible markers. CT-based preoperative models can be aligned to intraoperative camera or optical-tracking information in orthopedics and neurosurgery where the surfaces of bones or skin are relatively stiff as compared to abdominal soft tissue. These marker-assisted workflows have been reported to aid in accurate orthopedic navigation but require high quality calibration and stability of the markers.

In terms of the continuity of acquisition, intraoperative images can be continuous, intermittently continuous, or discontinuous. Laparoscopic video and real-time ultrasound provide continuous streams, whereas intraoperative CT offers staged updates at key moments. Preoperative CT and MRI supply detailed anatomical priors but cannot by themselves describe surgical changes that occur after incision, insufflation, or tissue manipulation.Practical surgical navigation often combines modalities rather than relying on one source. Laparoscopy supplies a real-time surface view, ultrasound adds subsurface information, and CT or MRI contributes preoperative anatomical context. The choice of modality depends on the surgical site, required accuracy, acceptable invasiveness, and available time for image acquisition.

One common limitation is that no one modality can offer high temporal resolution, high spatial detail, large field of view, deep anatomical visibility and low invasiveness at the same time.

DOI: 10.54254/2753-8818/2026.ES36570

Multimodal perception can be thought of as a compromise management strategy: it spreads the shortcomings of a single modality over complementary sources and adds new requirements to registration, synchronization, and uncertainty estimation [3].

2.2. Image segmentation and target recognition

Intraoperative segmentation and target recognition transform complicated surgical images into structures that can be utilized by a robot. The targets can be organ boundaries, tumor margins, blood vessels, nerves, sutures and surgical tools. Deep learning has become prominent in this area because convolutional networks and Transformer-based architectures can learn visual patterns that are difficult to define manually.

For instrument segmentation, semantic segmentation networks classify pixels and identify tool regions inside the endoscopic field. Some methods extend this task to detection and pose estimation, which gives motion planners more specific information about instrument orientation and workspace occupancy. Mask R-CNN and YOLO-style detectors have been adapted for tool localization, while U-Net variants and attention-based networks remain widely used for tissue and organ segmentation [4, 5]. As shown in Fig. 1, Cerón et al. [5] developed a lightweight YOLACT++-based framework that combines multi-scale feature fusion and attention mechanisms for real-time surgical instrument segmentation.

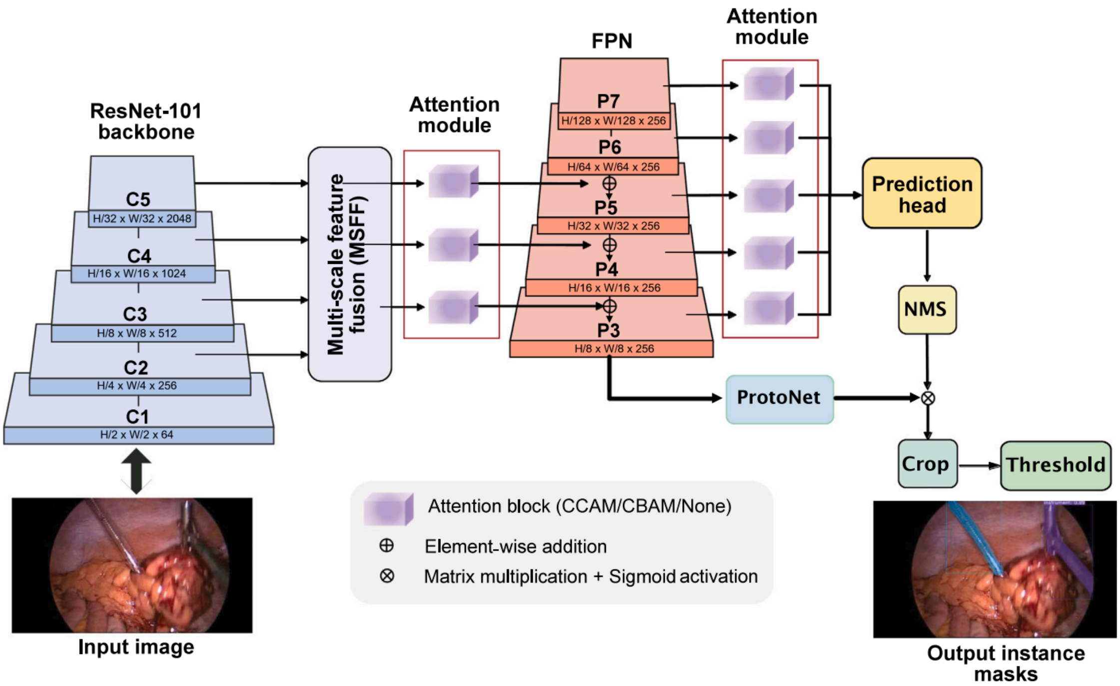

Figure 1. Real-time surgical instrument instance segmentation framework [5]

Clinical deployment still depends on robustness rather than benchmark accuracy alone. Blood, smoke, glare, image blur, and anatomical variability can shift image distributions away from curated training datasets. A segmentation method that distinguishes a clean vessel boundary in a public dataset may become less reliable when the camera angle changes or the tissue surface is partially covered. Domain adaptation, self-supervised learning, and uncertainty-aware prediction are increasingly important in this setting [6, 7].

DOI: 10.54254/2753-8818/2026.ES36570

Segmentation and recognition methods provide the visual primitives for localization, registration, and path planning. Their value in surgery is highest when the output is calibrated to the control task: a boundary mask, for example, is more useful when its uncertainty is known and when downstream control can respond safely to that uncertainty.

Recent reviews of instrument segmentation datasets make the same point from a data perspective. Endoscopic datasets are valuable because they create common benchmarks, but many still represent a narrow range of devices, procedures, lighting conditions, and annotation conventions. Methods trained on one dataset may show weaker performance when the camera, instrument type, tissue background, or surgical phase changes. A practical navigation system needs both high accuracy and evidence that the segmentation remains stable outside the dataset that produced the reported score [4, 6, 8].

2.3. Three-dimensional reconstruction and tissue motion estimation

Recovering three-dimensional structure from two-dimensional surgical images is a central problem in robot vision. Binocular endoscopy estimates depth from disparities between paired images, structured light uses patterns projected onto the surface to infer shape, and time-of-flight sensors use reflected light to estimate distance. These can be used to support spatial modelling, augmented reality visualisation and safe navigation of instruments around critical structures [9].

The reconstruction problem is more difficult in endoscopic surgery due to the possibility of weak texture, nonuniform lighting and the proximity of the camera to deformable tissue. Although a fixed surface can be reconstructed, the geometry can change very quickly as tissue is raised, cut, pressed or moved by physiological movement. Reconstruction algorithms should be able to manage spatial uncertainty as well as temporal instability.

The soft-tissue motion estimation is concerned with this dynamic aspect. Displacement has been estimated using optical flow, physical deformation models, and recurrent or temporal deep-learning models. These techniques may be used to refresh virtual anatomical models, correct respiratory and cardiac motion, and align planned routes with the real tissue surface.

A reconstruction pipeline that is clinically useful must not be limited to generating visually plausible three-dimensional models. It also needs to report errors, update quickly enough to support the control loop and be stable when a portion of the scene is blocked by tools or fluid. These demands are the reason why hybrid methods, which integrate geometric constraints with learning-based prediction, are gaining an increasing amount of attention.

The three-dimensional perception also interacts with tissue tracking. Small errors in depth can alter the location of a hidden structure that a surgeon thinks is there when an augmented reality overlay or motion constraint is applied to a reconstructed surface. Other researchers use geometric reconstruction with temporal priors, motion-flow cues, deformable models and point-cloud representations. These techniques do not eliminate uncertainty, but they can bring it into the open and make it easier to manage by a control system [10, 11].

3. Intraoperative localization and spatial registration

3.1. Image registration and multimodal fusion

Image registration aligns information collected from different viewpoints, modalities, or time points. In surgical navigation, this usually means matching a preoperative CT or MRI model with intraoperative images so that the target anatomy can be displayed and updated during the procedure.

DOI: 10.54254/2753-8818/2026.ES36570

Rigid registration is suitable when the relevant anatomy changes little, while non-rigid registration is needed when organs shift or deform [9].

A typical registration workflow includes feature extraction, feature matching, similarity measurement, and optimization of alignment parameters. Features may be artificial markers, anatomical landmarks, surface point clouds, or dense pixel and voxel information. Marker-based approaches are often easier to implement but add procedural burden; markerless approaches reduce this burden but require stronger image interpretation and more careful error control. As shown in Fig. 2, the CAI4CAI framework integrates multimodal data and contextual knowledge to support adaptive decision-making in computer-assisted interventions [9].

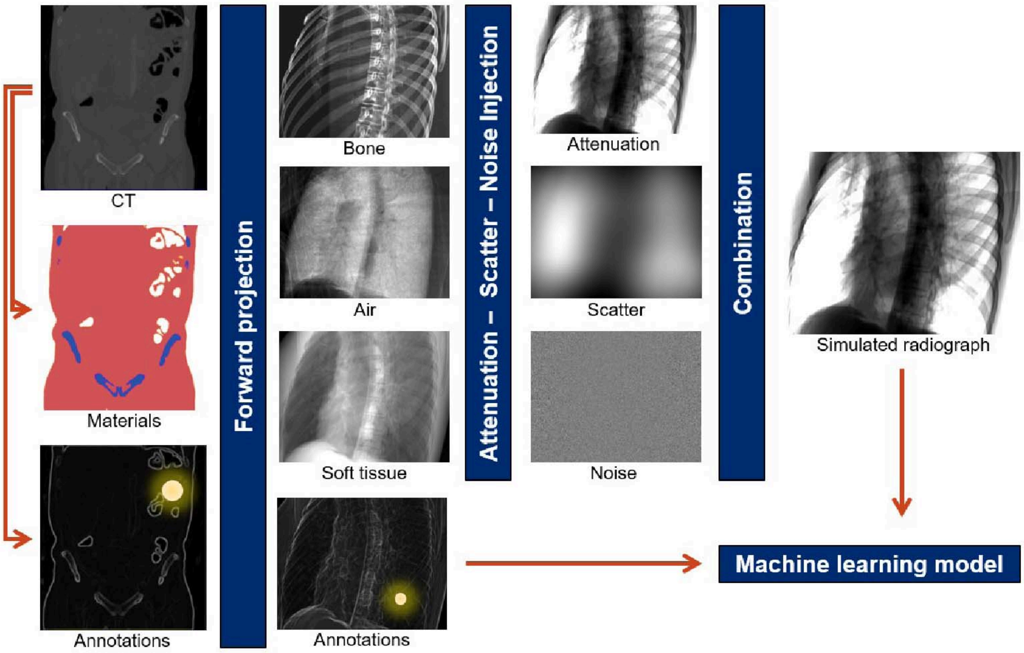

Figure 2. Conceptual framework of contextual artificial intelligence for computer-assisted interventions (CAI4CAI) [9]

Non-rigid registration remains one of the difficult links in surgical navigation. Traditional algorithms may be computationally expensive, whereas purely data-driven models can struggle when the target domain differs from the training data. Recent studies increasingly combine anatomical constraints, biomechanical priors, and neural networks to improve both speed and plausibility. In practice, registration quality must be judged by the error that remains at the surgical target, not only by global image similarity.

In soft-tissue surgery, a registration result should be treated as time-sensitive. A preoperative model may be accurate at the beginning of a procedure and less reliable after insufflation, retraction, bleeding, or partial resection. Continuous or periodically updated registration can reduce this drift, but it also increases computational load. The clinical question is not whether a model can be aligned once, but whether the alignment remains useful when the anatomy and camera viewpoint are changing together [9].

3.2. Real-time localization and tracking

During surgery, the robot must know its relationship to the patient, the surgical instruments, and the target anatomy. Optical and electromagnetic tracking systems have long been used for this purpose.

DOI: 10.54254/2753-8818/2026.ES36570

Reflective spheres, electromagnetic sensors, and calibrated coordinate frames can provide three-dimensional position information with high precision, but they also introduce line-of-sight limitations, susceptibility to interference, and extra setup steps.

Vision-based tracking can be marker-based or markerless. Marker-based methods detect visual tags on instruments or patient surfaces and estimate their position from camera images. Markerless methods instead use natural image features, tool contours, tissue textures, or deep-learning detections to follow motion without additional markers. Markerless tracking is more flexible, but it has to cope with occlusion, deformation, and changing illumination.

Camera self-localization is part of the same problem. An endoscopic camera mounted on a robotic arm must be calibrated with respect to the robot base, and its pose must be updated so that image coordinates correspond to robot motion. Accuracy is commonly evaluated by comparing estimated target positions with ground-truth measurements, but surgical relevance also depends on whether the error occurs near a critical structure or in a low-risk region.

3.3. Accuracy evaluation and error analysis

Reliable navigation requires explicit error analysis. Deterministic errors may arise from robot kinematic parameters, calibration deviations, and manufacturing tolerances; random errors may come from sensor noise, image artifacts, and environmental disturbance. Among these sources, geometric parameter error is often a major contributor to robotic positioning error.

Compensation strategies usually combine model calibration with feedback correction. In kinematic calibration, experimentally measured end-effector errors are used to update the robot model. In online visual correction, visually estimated positions are compared with expected positions during operation, and the control output is adjusted accordingly. Force or tactile feedback may further reduce error when visual information is incomplete.

Error evaluation should be connected to surgical risk. A millimeter-level deviation has different implications in bone drilling, neurosurgery, soft-tissue manipulation, and endoscope positioning. Validation studies need task-specific metrics, clinically meaningful tolerance thresholds, and transparent reporting of how errors propagate from image acquisition to registration and finally to robot motion.

Error analysis is also a communication problem. Navigation systems often report numerical target-registration error, but surgeons need to understand whether the uncertainty is near a nerve, vessel, duct, tumor margin, or planned cutting plane. A richer interface could show confidence, recent drift, and the source of the current estimate, allowing the operator to distinguish a stable visual cue from a stale or uncertain one. Such reporting is especially important as systems move from passive display toward active guidance [1, 12].

4. Vision-guided robot control methods

Vision-guided control uses image-derived information to guide robotic motion. The main strategies include visual servoing, motion planning under surgical constraints, visuo-haptic fusion for safe interaction, and data-driven control methods that learn policies from demonstrations, simulation, or prior procedures. These strategies differ in how directly visual information enters the control loop.

DOI: 10.54254/2753-8818/2026.ES36570

4.1. Vision feedback-based control

Visual servoing is used to produce motion commands based on visual feedback. Image-based visual servoing directly maps image-plane error in features to robot motion without explicitly reconstructing the three-dimensional space and can be used to support real-time operation. Position-based visual servoing estimates the three dimensional pose of the target first and then computes the control command in Cartesian space. Hybrid forms are a combination of image and pose information to strike a balance between stability and geometric interpretability [2, 3].

Visual servoing is applied to endoscope viewpoint adjustment, instrument tracking and fine alignment in surgical robots. An automated endoscope holder, such as, can maintain the tool tip close to the center of the field of view by reacting to its image offset. The same feedback can aid in suturing alignment or controlled movement along a planned visual trajectory.

This is also true for camera control. While automatic endoscope guidance can alleviate the surgeon's burden, a camera controller that responds to brief tool movements can disturb the visual scene. Smoothness, latency, field-of-view preservation, and handover between human and automated control belong to the control problem itself. A visually centered instrument is not always a clinically optimal view if the tissue target, bleeding source, or dissection plane lies elsewhere [2, 3].

The central difficulty is that surgical images are not stable measurements. Feature points can disappear behind tissue or instruments, depth estimation may drift, and the camera itself may move. Robust visual servoing depends on feature selection, outlier rejection, camera calibration, and safety constraints that prevent unstable feedback from generating unsafe motion.

4.2. Motion planning and path optimization

In a vision-guided framework, motion planning determines how the robot reaches a target while respecting surgical constraints. These constraints include collision avoidance, instrument entry-point restrictions, workspace boundaries, and safe distances from vulnerable tissues. Sampling-based methods such as rapidly exploring random trees and probabilistic roadmaps, as well as optimization-based planners, have been applied to this problem.

Surgical planning cannot rely on a fixed map for long. Tissue morphology changes, the field of view is narrow, and the target may move as instruments interact with the anatomy. Online planning methods address this by updating the environment model through visual feedback and adjusting the trajectory during execution. Path optimization must account not only for reachability and collision avoidance but also for tissue injury, motion smoothness, and operative efficiency.

Planning methods are most convincing when they incorporate local surgical context. A shortest collision-free path may still be poor if it crosses fragile tissue, blocks the surgeon's view, or requires an awkward instrument angle at the entry point. Vision-guided planning has to combine geometric feasibility with procedural priorities. In semi-autonomous systems, this often means that the robot proposes or updates a path, while the surgeon retains authority over whether that path matches the operative intention [1, 13].

A useful planner treats perception uncertainty as part of the planning problem. If a vessel boundary is uncertain or a target is partially occluded, the robot should slow down, request a better view, or choose a safer trajectory. The link between visual confidence and motion policy is likely to become more important as surgical robots move toward higher levels of autonomy.

DOI: 10.54254/2753-8818/2026.ES36570

4.3. Visuo-haptic fusion control

Vision provides spatial information, but contact tasks often require force feedback. During grasping, cutting, suturing, or palpation, force signals help determine tissue stiffness, contact state, and the risk of excessive pressure. Visuo-haptic fusion combines image-based localization with force-sensor measurements so that robotic motion can respond to both geometry and contact [14].

Common control strategies include impedance control, hybrid force/position control, and adaptive control. Impedance control creates a virtual spring-damper relationship between force and motion, allowing compliant responses to contact. Hybrid force/position control regulates position in some directions while controlling force in others, which is useful when a tool must follow a surface while limiting tissue load.

The practical barriers are hardware and synchronization. Miniaturized force sensors must remain reliable after sterilization and repeated use; visual estimates and force measurements must be temporally aligned; and heterogeneous signals must be fused without adding unsafe delay. Progress in high-bandwidth tactile sensing and probabilistic fusion may make visuo-haptic control more dependable in teleoperation and semi-autonomous procedures. Figure 3 illustrates an AI-assisted interface that integrates visual analysis and force information for safer intraoperative guidance [12].

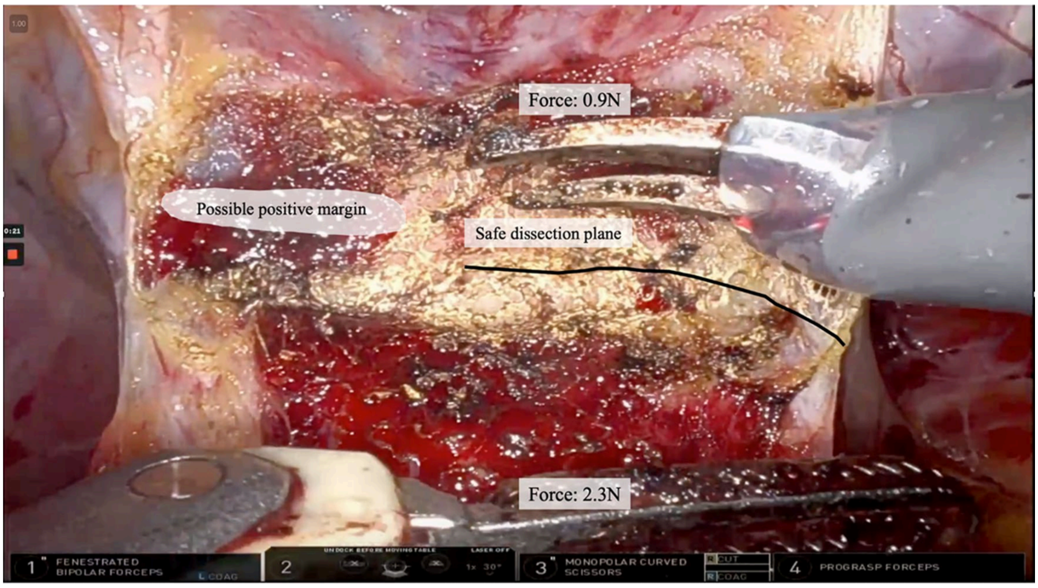

Figure 3. AI-assisted intraoperative visualization integrating tissue assessment, surgical plane identification, and force measurement for enhanced visuo-haptic robotic guidance [12]

Haptic vision studies illustrate how contact information can compensate for visual ambiguity. In intracardiac catheter navigation, contact cues and anatomical sensing have been used to support autonomous movement in a beating-heart environment. Surgical robots in the abdomen or pelvis face different mechanics, but the lesson is similar: visual navigation becomes safer when the robot can detect whether its geometric plan is consistent with the forces occurring at the tissue interface [14].

4.4. Data-driven and intelligent control

Artificial intelligence has introduced new control strategies for surgical robotics. Deep reinforcement learning can learn action sequences through interaction with a simulated environment,

DOI: 10.54254/2753-8818/2026.ES36570

while imitation learning can derive policies from expert demonstrations. These approaches are attractive for complex tasks in which hand-designed control rules are difficult to specify.

Real surgery imposes constraints that simulation alone cannot solve. Training data are limited, unsafe exploration is unacceptable, and policies must generalize across anatomy, instruments, and imaging conditions. Researchers have explored simulation-to-real transfer, combinations of imitation learning and reinforcement learning, safety-constrained learning, and model-based learning to improve reliability.

The autonomous intestinal anastomosis literature shows both the promise and the boundary of data-driven control. Autonomous execution can improve consistency in selected tasks, yet the system still depends on defined task structure, imaging assumptions, and operator selection of a plan. This supports a more cautious interpretation of surgical intelligence: the near-term goal is not a fully independent surgeon replacement, but a set of supervised capabilities that can reduce variability in narrow, well-validated task segments [13, 15].

Data-driven control is most credible when it is integrated with conventional safety mechanisms. A learned model may propose motion commands based on visual segmentation, but a planner or supervisory controller can enforce workspace limits, speed limits, and collision constraints. Such combinations allow partial autonomy while preserving an explicit safety layer [1, 13].

The next stage of intelligent control will likely depend on transparent validation rather than algorithmic novelty alone. Surgical systems must demonstrate stable performance under distribution shift, provide interpretable failure signals, and maintain a clear boundary between autonomous adjustment and clinician authority.

5. Challenges and future outlook

Visual navigation and control have substantially improved the capabilities of surgical robots. Multimodal imaging and learning-based segmentation have strengthened scene understanding; registration, tracking, and error compensation have improved localization; visual servoing and multimodal feedback control have made robotic motion more responsive to the operative field [1, 12, 16].

Several unresolved problems still limit clinical translation. Image quality can degrade quickly in dynamic surgical environments. Multimodal fusion must be performed in real time while preserving spatial accuracy. Learning-based algorithms need better generalization under limited and heterogeneous data. Safety and reliability must be verified not only in laboratory demonstrations but also in task-specific clinical settings.

The reference base also shows an uneven maturity across subsystems. Instrument segmentation has become data-rich compared with deformable registration and force-vision fusion, where clinical datasets and shared benchmarks are less established. This imbalance can encourage visually impressive demonstrations that are not yet complete navigation solutions. A mature surgical robot must connect perception, localization, planning, and actuation under a validation framework that reflects the procedure being performed [4, 17, 18].

Further development is likely to follow several connected paths. Self-supervised and federated learning may help improve perception models without relying entirely on large centralized datasets. More accurate real-time registration could reduce the gap between preoperative planning and intraoperative anatomy. Closed-loop systems that combine vision, force, and intelligent control may support safer semi-autonomous operation. The decisive step will be to connect technical performance with clinically meaningful evidence, so that visual navigation becomes not only more accurate but also more dependable during real surgery.

DOI: 10.54254/2753-8818/2026.ES36570

References

[1] Lee, A., Baker, T., Bederson, J., & Rapoport, B. (2024). Levels of autonomy in FDA-cleared surgical robots: A systematic review. *npj Digit Med, 7*(1), 8.

[2] Attanasio, A., Scaglioni, B., De Momi, E., Fiorini, P., & Valdastri, P. (2021). Autonomy in surgical robotics. *Annual Review of Control, Robotics, and Autonomous Systems, 4*, 651-679.

[3] Mendoza-Bautista, K., Torres-Mendez, L., & Chairez, I. (2025). Systematic review on visual aid technologies for surgical assistant robotic devices. *Prog Biomed Eng, 7*(2), 27.

[4] Rueckert, T., Rueckert, D., & Palm, C. (2024). Methods and datasets for segmentation of minimally invasive surgical instruments in endoscopic images and videos: A review of the state of the art. *Comput Biol Med, 169*, 24.

[5] Cerón, J., Ruiz, G., Chang, L., & Ali, S. (2022). Real-time instance segmentation of surgical instruments using attention and multi-scale feature fusion. *Medical Image Analysis, 81*.

[6] Allan, M., Shvets, A., Kurmann, T., Zhang, Z., & Azizian, M. (2019). 2017 robotic instrument segmentation challenge.

[7] Madani, A., Namazi, B., Altieri, M. S., Hashimoto, D. A., & Alseidi, A. (2020). Artificial intelligence for intraoperative guidance: Using semantic segmentation to identify surgical anatomy during laparoscopic cholecystectomy. *Annals of Surgery*.

[8] Liu, X., Guo, X., Liu, Y., & Yuan, Y. (2021). Consolidated domain adaptive detection and localization framework for cross-device colonoscopic images. *Medical Image Analysis, 71*, 13.

[9] Vercauteren, T., Unberath, M., Padoy, N., & Navab, N. (2020). CAI4CAI: The rise of contextual artificial intelligence in computer-assisted interventions. *Proceedings of the IEEE, 108*(1), 198–214.

[10] Jin, Y., Cheng, K., Dou, Q., & Heng, P. (2019). Incorporating temporal prior from motion flow for instrument segmentation in minimally invasive surgery video. In *Proceedings of the 10th International Workshop on Machine Learning in Medical Imaging (MLMI) / 22nd International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)* (pp. 440–448). Springer International Publishing. Shenzhen, PEOPLES R CHINA.

[11] Wang, Y., Sun, Y., Liu, Z., Sarma, S., Bronstein, M., & Solomon, J. (2019). Dynamic graph CNN for learning on point clouds. *ACM Transactions on Graphics, 38*(5), 12.

[12] Knudsen, J., Ghaffar, U., Ma, R., & Hung, A. (2024). Clinical applications of artificial intelligence in robotic surgery. *Journal of Robotic Surgery, 18*(1), 10.

[13] Saeidi, H., Opfermann, J., Kam, M., et al. (2022). Autonomous robotic laparoscopic surgery for intestinal anastomosis. *Science Robotics, 7*(62), 13.

[14] Fagogenis, G., Mencattelli, M., Machaidze, Z., et al. (2019). Autonomous robotic intracardiac catheter navigation using haptic vision. *Science Robotics, 4*(29), 12.

[15] Ray, K. (2022). Autonomous robotic laparoscopic gastrointestinal surgery. *Nature Reviews Gastroenterology & Hepatology*(3), 19.

[16] Haidegger, T., Speidel, S., Stoyanov, D., & Satava, R. (2022). Robot-assisted minimally invasive surgery-surgical robotics in the data age. *Proceedings of the IEEE, 110*(7), 835-46.

[17] Rieke, N., Hancox, J., Li, W., et al. (2020). The future of digital health with federated learning. *NPJ Digital Medicine, 3*(1), 7.

[18] Bodenstedt, S., Wagner, M., Kati, D., et al. (2017). Unsupervised temporal context learning using convolutional neural networks for laparoscopic workflow analysis.