Application of artificial intelligence in surgery

Xiao-Yun Zhou (✉)[1], Yao Guo[1], Mali Shen[1], Guang-Zhong Yang[2]

[1]The Hamlyn Centre for Robotic Surgery, Imperial College London, London SW7 2AZ, UK; [2]Institute of Medical Robotics, Shanghai Jiao Tong University, Shanghai 200240, China

Abstract Artificial intelligence (AI) is gradually changing the practice of surgery with technological advancements in imaging, navigation, and robotic intervention. In this article, we review the recent successful and influential applications of AI in surgery from preoperative planning and intraoperative guidance to its integration into surgical robots. We conclude this review by summarizing the current state, emerging trends, and major challenges in the future development of AI in surgery.

Keywords artificial intelligence; surgical autonomy; medical robotics; deep learning

Introduction

Advances in surgery have revolutionized the management of both acute and chronic diseases, prolonging life and extending the boundary of patient survival. These advances are underpinned by continuing technological developments in diagnosis, imaging, and surgical instrumentation. Complex surgical navigation and planning are made possible through the use of both pre- and intraoperative imaging techniques, such as ultrasound, computed tomography (CT), and magnetic resonance imaging (MRI) [1]. Surgical trauma is reduced through minimally invasive surgery (MIS), which is now progressively combined with robotic assistance [2]. Postoperative care is also improved by sophisticated wearable and implantable sensors for supporting early discharge after surgery, thereby enhancing patient recovery and early detection of postsurgical complications [3,4]. Numerous terminal illnesses have been transformed into clinically manageable chronic lifelong conditions, and surgery is increasingly focused on the systematic effects of this procedure on patients, avoiding isolated surgical treatment or anatomical alteration, with careful consideration of metabolic, hemodynamic, and neurohormonal consequences that can influence the quality of life.

Owing to recent advances in medicine, artificial intelligence (AI) has played an important role in supporting clinical decision-making since the early years of the development of the MYCIN system [5]. AI is now increasingly used for risk stratification, genomics, imaging and diagnosis, precision medicine, and drug discovery. AI was introduced into surgery more recently, with a strong root in imaging and navigation and early techniques focusing on feature detection and computer-assisted intervention for both preoperative planning and intraoperative guidance. Over the years, supervised algorithms, such as active-shape models, atlas-based methods, and statistical classifiers, have been developed [1]. The recent successes of deep convolutional neural network (DCNN), such as AlexNet [6], have enabled automatically learned data-driven descriptors to be used for image understanding, which have shown improved robustness and generalizability compared with ad hoc hand-crafted features.

As robotics is increasingly applied in surgery, AI is set to transform the field through the development of sophisticated functions connecting real-time sensing to robotic control. Varying levels of autonomy can allow the surgeon and robotic system to navigate together the constantly changing and patient-specific environments, a situation that may reduce the ability of either to complete a surgical task effectively. Additionally, by leveraging the parallel medical advances in early detection and targeted therapy, AI can ensure that the proper intervention is executed. Future surgical robots would expectedly be able to perceive and understand complicated surroundings, conduct real-time decision-making, and perform desired tasks with increased precision, safety, and efficiency. But what are the roles of AI in these systems and the future of surgery in general? How can these systems deal with

dynamic environments and learn from human operators? How can reliable control policies be derived to achieve human–machine symbiosis?

In this article, we review the applications of AI in preoperative planning, intraoperative guidance, and its integrated use in surgical robotics. Popular AI techniques, including an overview of their requirements, challenges, and sub-areas in surgery are summarized in Fig. 1, which outlines the main flow of the contents of the paper. We first introduce the application of AI in preoperative planning. We then discuss several AI techniques for intraoperative guidance and review the applications of AI in surgical robotics. Finally, we provide our conclusions and future outlook. Technically, we put a strong emphasis on deep learning-based approaches in this review.

AI for preoperative planning

Preoperative planning where surgeons plan the surgical procedure on the basis of existing medical records and imaging is essential for the success of a surgery. X-ray, CT, ultrasound, and MRI are the most common imaging modalities used in clinical practice. Routine tasks based on medical imaging include anatomical classification, detection, segmentation, and registration.

Classification

Classification outputs the diagnostic value of the input, which is a single or a set of medical images or volumes of organs or lesions. Aside from traditional machine-learning and image-analysis techniques, deep learning-based meth-

ods are growing in popularity [7]. The network architecture of deep learning-based methods for classification is composed of convolutional layers for extracting information from the input and fully connected layers for regressing the diagnostic value.

For example, a classification pipeline using Google's Inception and ResNet architecture was proposed to segment lung, bladder, and breast cancers [8]. Chilamkurthy *et al.* demonstrated that deep learning can recognize intracranial hemorrhage, calvarial fracture, midline shift, and mass effect from head CT scans [9]. The mortality, renal failure, and postoperative bleeding in patients after cardiosurgical care can be predicted by recurrent neural network (RNN) in real time with improved accuracy compared with standard-of-care clinical tools [10]. ResNet-50 and Darknet-19 are used to classify benign or malignant lesions in ultrasound images, showing similar sensitivity and improved specificity [11].

Detection

Detection provides the spatial localization of regions of interest, often in the form of bounding boxes or landmarks, and may also include image- or region-level classification. Similarly, deep learning-based approaches have shown promise in detecting various anomalies or medical conditions. DCNNs for detection usually consist of convolutional layers for feature extraction and regression layers to determine the bounding box properties.

A deeply stacked convolutional autoencoder was trained to extract statistical and kinetic biological features for detecting prostate cancer from 4D positron-emission tomography images [12]. A 3D CNN with roto-translation
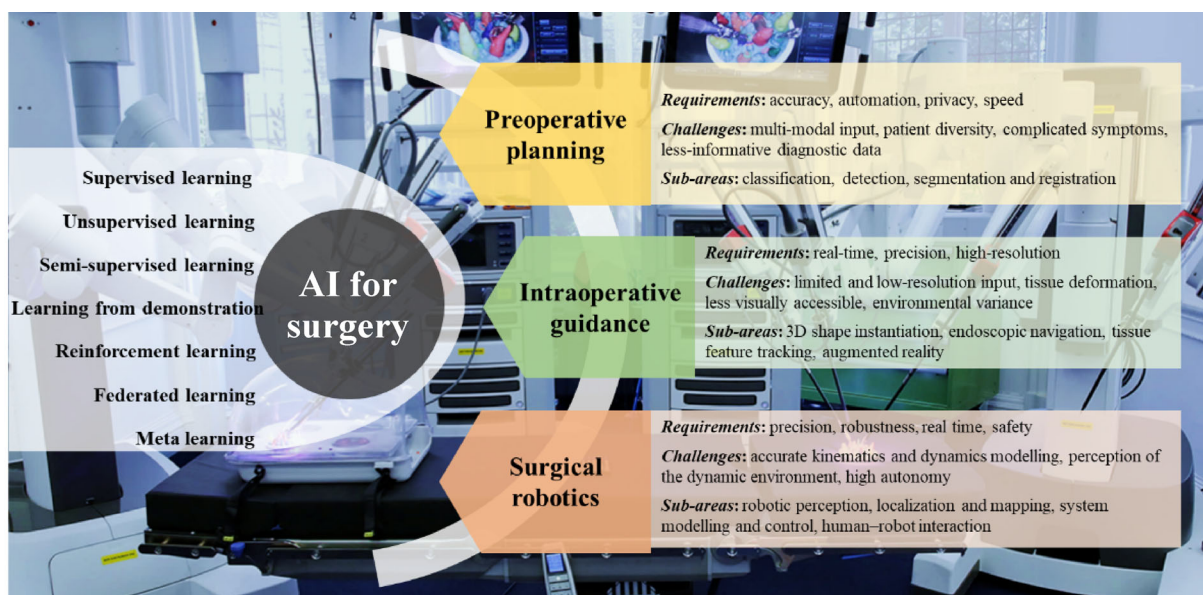
Fig. 1 An overview of popular AI techniques, as well as the key requirements, challenges, and sub-areas of AI used in preoperative planning, intraoperative guidance, and surgical robotics.

group convolutions was proposed for pulmonary nodule detection with good accuracy, sensitivity, and convergence speed [13].

Deep reinforcement learning (DRL) based on an extension of the deep Q-network was used to learn a search policy from dynamic contrast-enhanced MRI for detecting breast lesions [14]. To detect acute intracranial hemorrhage from CT scans and improve network interpretability, Lee *et al.* [15] used an attention map and an iterative process to mimic the workflow of radiologists.

Segmentation

Segmentation can be treated as a pixel- or voxel-level image classification problem. Owing to the limitation of computational resources in early works, each image or volume was previously divided into small windows, and CNNs were trained to predict the target label at the central location of the window. Image- or voxel-wise segmentation was achieved by running the CNN classifier over densely sampled image windows. For example, Deepmedic exhibited good performance for multimodal brain tumor segmentation from MRI [16]. However, the sliding window-based approach is inefficient as the network's function is repeatedly computed in regions where many windows are overlapping. For this reason, the sliding window-based approach was recently replaced by fully convolutional networks (FCNs) [17]. The key idea of FCNs is to replace fully connected layers in a classification network with convolutional layers and upsampling layers, a process that substantially improves segmentation efficiency. Encoder–decoder networks, such as U-Net [18,19], have shown promising performance in medical-image segmentation. The encoder has multiple convolutional and downsampling layers that extract image features at different scales. The decoder has convolutional and upsampling layers that recover the spatial resolution of feature maps and finally achieve pixel- or voxel-wise dense segmentation. A review of different normalization methods for training U-Net for medical-image segmentation is provided by Zhou and Yang [20].

For navigation during endoscopic pancreatic and biliary procedures, Gibson *et al.* [21] used dilated convolutions and fused-image features in multiple scales to segment abdominal organs from CT scans. For interactive segmentation of placenta and fetal brains from MRI, FCN and user-defined bounding boxes and scribbles were combined, where the last few layers of FCN were fine-tuned according to user input [22]. The segmentation and localization of surgical instrument landmarks were modeled as heatmap regression, and an FCN was used to track the instruments in near real time [23]. For pulmonary nodule segmentation, Feng *et al.* addressed the issue of requiring accurate manual annotations when training FCNs by learning discriminative regions from weakly labeled lung CT with a candidate screening method [24]. Bai *et al.* proposed a self-supervised learning strategy to improve the cardiac segmentation accuracy of U-Net with limited labeled training data [25].

Registration

Registration is the spatial alignment between two medical images, volumes, or modalities. It is particularly important for both pre- and intraoperative planning. Traditional algorithms usually iteratively calculate a parametric transformation, i.e., an elastic, fluid, or B-spline model, to minimize a given metric, i.e., mean square error, normalized cross correlation, or mutual information, between two medical inputs. Recently, deep regression models were used to replace the traditional time-consuming and optimization-based registration algorithms.

An example of deep learning-based approaches to registration is VoxelMorph, which maximizes standard image-matching objective functions by leveraging a CNN-based structure and auxiliary segmentation to map an input image pair to a deformation field [26]. An end-to-end deep learning framework was proposed for 3D medical-image registration that consisted of three stages, namely, affine transform prediction, momentum calculation, and non-parametric refinement, to combine affine registration and vector momentum-parameterized stationary velocity field [27]. A weakly supervised framework was proposed for multimodal image registration, with training on images with a higher level correspondence, i.e., anatomical labels, rather than voxel-level transformation for predicting the displacement field [28]. A Markov decision process with each agent trained with dilated FCN was applied to align a 3D volume to 2D X-ray images [29]. RegNet was proposed by considering multiscale contexts and trained on an artificially generated displacement vector field to achieve a nonrigid registration [30]. A 3D image registration can also be formulated as a strategy-learning process with 3D raw images as the input, the next optimal action (i.e., up or down) as the output, and the CNN as the agent [31].

AI for intraoperative guidance

Computer-aided intraoperative guidance has always been a cornerstone of MIS. Learning strategies have been extensively integrated into the development of intraoperative guidance to provide enhanced visualization and localization in surgery. Recent works can be divided into four main aspects: shape instantiation, endoscopic navigation, tissue tracking, and augmented reality (AR) (Fig. 2).
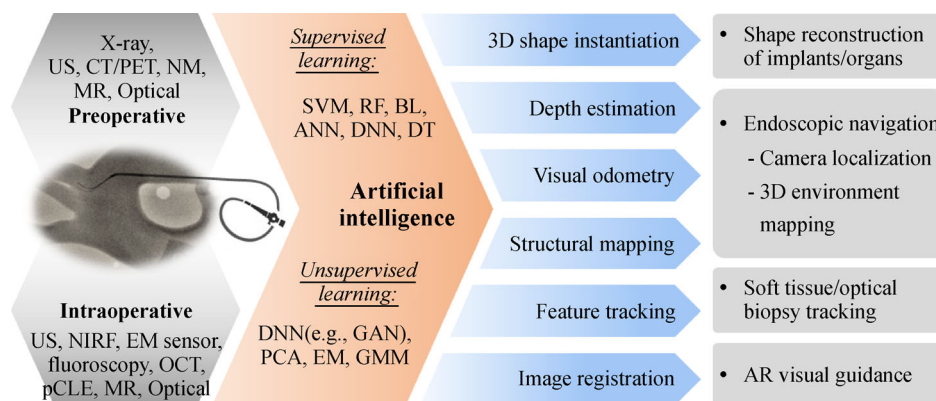
Fig. 2 AI techniques for computer-aided intraoperative guidance in MIS. Multimodal data acquired preoperatively and intraoperatively are used in either supervised or unsupervised learning-based techniques for various surgical applications. US, ultrasound; NIRF, near-infrared fluorescence; OCT, optical coherence tomography; pCLE, probe-based confocal laser endomicroscopy; EM sensor, electromagnetic sensor; RF, random forests; BL, Bayesian learning; DT, decision tree; EM, expectation maximization; GMM, Gaussian mixture models.

3D shape instantiation

For intraoperative 3D reconstruction, 3D volumes can be scanned with MRI, CT, or ultrasound. In practice, this 3D/4D process can be time consuming or produce scans with low resolution. Limiting the number of images needed for 3D shape reconstruction can enable a program to reconstruct a 3D surgical scene in real time, and superior protocols can additionally improve the resolution of the reconstruction. Intraoperative real-time 3D shape instantiation from a single or limited amount of 2D images is an emerging area of research.

For example, a 3D prostate shape was instantiated from multiple nonparallel 2D ultrasound images with a radial basis function [32]. The 3D shapes of fully compressed, fully deployed, and also partially deployed stent grafts were instantiated from a single projection of 2D fluoroscopy with mathematical modeling combined with the robust perspective-n-point method, graft gap interpolation, and graph neural networks [33–35]. Furthermore, an equally weighted focal U-Net was proposed to segment automatically the markers on stent grafts and improve the efficiency of intraoperative stent graft shape-instantiation framework [36]. Moreover, a 3D AAA skeleton was instantiated from a single projection of 2D fluoroscopy with skeleton deformation and graph matching [37]. A 3D liver shape was instantiated from a single 2D projection via principal component analysis (PCA), statistical shape model (SSM), and partial least square regression (PLSR) [38]; this work was further generalized to a registration-free shape instantiation framework for any dynamic organ with sparse PCA, SSM, and kernel PLSR [39]. Recently, an advanced deep and one-stage learning strategy that estimates 3D point cloud from a single 2D image was proposed for 3D shape instantiation [40].

Endoscopic navigation

The trend now in surgery is increasingly leaning toward intraluminal procedures and endoscopic surgery driven by early detection and intervention. Navigation techniques have been evaluated to guide the maneuvering of endoscopes toward target locations. To this end, learning-based depth estimation, visual odometry, and simultaneous localization and mapping (SLAM) have been tailored for camera localization and environment mapping with the use of endoscopic images.

Depth estimation

Depth estimation from endoscopic images plays an essential role in 6 DoF camera motion estimation and 3D structural environment mapping, which had been tackled either by supervised [41,42] or by self-supervised [43,44] deep learning methods. This process is hindered by two main challenges. First, obtaining a large amount of high-quality training data containing paired video images and depth maps is practically difficult because of both hardware constraints and labor-intensive labeling. Second, surgical scenes are often textureless that makes applying depth recovery methods that rely on feature matching and reconstruction difficult [45,46].

To address the challenge of limited training data, Ye *et al.* [47] proposed a self-supervised depth estimation approach for stereoimages using siamese networks. For monocular depth recovery, Mahmood *et al.* [41,42] learned the mapping from rendered RGB images to the corresponding depth maps with synthetic data and adopted domain transfer learning to convert real RGB images into rendered images. Additionally, self-supervised unpaired image-to-image translation [44] using a modified cycle

generative adversarial network (CycleGAN) [48] was proposed to recover the depth from bronchoscopic images. Moreover, a self-supervised CNN based on the principle of shape from motion was applied to recover the depth and achieve visual odometry for an endoscopic capsule robot [43].

Visual odometry

Visual odometry uses consecutive video frames to estimate the pose of a moving camera. CNN-based approaches [49] were adopted for camera pose estimation on the basis of temporal information. Turan *et al.* [49] estimated the camera pose for an endoscopic capsule robot using a CNN for feature extraction and long short-term memory (LSTM) for dynamics estimation. Sganga *et al.* [50] combined ResNet and FCN to calculate the pose change between consecutive video frames. However, the feasibility of localization approaches according to visual odometry was only validated in lung phantom data [50] and gastro-intestinal (GI) tract data [49].

3D reconstruction and localization

Owing to the dynamic nature of tissues, real-time 3D reconstruction of tissue environment and localization are vital prerequisites for navigation. SLAM is a widely studied technique in robotics. In SLAM, the robot can simultaneously build a 3D map of surrounding environments and localize the camera pose in the built map. Traditional SLAM algorithms are based on the assumption of a rigid environment, which is in contrast to that found in a typical surgical scene where the deformation of soft tissues and organs is involved. Hence, this flawed assumption limits its direct adoption for surgical tasks. To address this limitation, Mountney *et al.* [51] first applied the extended Kalman filter SLAM (EKF-SLAM) framework [52] with a stereoendoscope, where the SLAM estimation was compensated with periodic motion of soft tissues caused by respiration [53]. Grasa *et al.* [54] evaluated the effectiveness of monocular EKF-SLAM in hernia repair surgery for measuring hernia defects. Turan *et al.*[55] estimated the depth images from the RGB data through shape from shading. They then adopted the RGB-D SLAM framework by using paired RGB and depth images. Song *et al.* [56] implemented a dense deformable SLAM on a graphics processing unit (GPU) and an ORB-SLAM on a central processing unit (CPU) to boost the localization and mapping performance of a stereoendo-scope.

Endovascular interventions have been increasingly utilized to treat cardiovascular diseases. However, visual cameras are not applicable inside vessels. For example, catheter mapping is commonly used in radiofrequency catheter ablation for navigation [57]. To this end, recent advances in intravascular ultrasound (IVUS) have offered another avenue for endovascular intraoperative guidance. Shi *et al.* [58] first proposed the simultaneous catheter and environment (SCEM) framework for 3D vasculature reconstruction by fusing electromagnetic sensing data and IVUS images. To deal with the errors and uncertainty measured from both EM sensors and IVUS images, they improved SCEM and reconstructed the 3D environment by solving a nonlinear optimization problem [59]. To alleviate the burden of preregistration between preoperative CT data and EM sensing data, a registration-free SCEM approach was proposed for more efficient data fusion [60].

Tissue feature tracking

Learning strategies have also been applied to soft tissue tracking in MIS. Mountney and Yang [61] introduced an online learning framework that updates the feature tracker over time by selecting correct features using decision tree classification. Ye *et al.* [62] proposed a detection approach that incorporates a structured support vector machine (SVM) and online random forest for re-targeting a preselected optical biopsy region on soft tissue surfaces of the GI tract. Wang *et al.* [63] adopted a statistical appearance model to differentiate the organ from the background in their region-based 3D tracking algorithm. Their validation results demonstrated that incorporating learning strategies can improve the robustness of tissue tracking with respect to the deformations and variations in illumination.

Augmented reality

AR improves surgeons' intraoperative vision by providing a semitransparent overlay of preoperative image on the area of interest [64]. Wang *et al.* [65] used a projector to project the AR overlay for oral and maxillofacial surgery. A 3D contour matching was used to calculate the transformation between the virtual image and real teeth. Instead of using projectors, Pratt *et al.* exploited Hololens, a head-mounted AR device, to project a 3D vascular model on the lower limb of patients [66]. Given that one of the most challenging tasks is to project the overlay on markerless deformable organs, Zhang *et al.* [67] intro-duced an automatic registration framework for AR navigation, in which the iterative closet point and RANSAC algorithms were applied for 3D deformable tissue reconstruction.

AI for surgical robotics

Owing to the development of AI techniques, surgical robots can achieve superhuman performance during MIS

[68,69]. The objective of AI is to boost the capability of surgical robotic systems in perceiving complex *in vivo* environments, conducting decision-making, and performing the desired tasks with increased precision, safety, and efficiency. As illustrated in Fig. 3, the common AI techniques used for robotic and autonomous systems (RAS) can be summarized in four aspects: (1) perception, (2) localization and mapping, (3) system modeling and control, and (4) human–robot interaction.

As overlap exists between intraoperative guidance and robot localization and mapping, this section mainly covers the methods for increasing the level of surgical autonomy.

Perception

### Instrument segmentation and tracking

Instrument segmentation tasks can be divided into three groups: segmentation to distinguish the instrument from the background, multiclass segmentation of instrument parts (i.e., shaft, wrist, and gripper), and multiclass segmentation for different instruments. The advancement of deep learning in segmentation has remarkably improved the instrument segmentation accuracy from the exploitation of SVM for pixel-level binary classification [70] to recent DCNN architectures, such as U-Net, TernausNet-VGG11, TernausNetVGG16, and LinkNet, for both binary segmentation and multiclass segmentation [71]. To further improve instrument segmentation performance, Islam *et al.* developed a cascaded CNN with a multiresolution feature fusion framework [72].

Algorithms for solving tracking problems can be separated into two categories: tracking by detection and tracking via local optimization [73]. Previous works in this field mainly relied on hand-crafted features, such as Haar wavelets [73], color or texture features [74], and gradient-based features [75]. In the context of deep learning-based methods, the proposed methods were built on the basis of the concept of tracking by detection [76,77]. Various CNN architectures, such as AlexNet [76] and ResNet [23,77], were used to detect the surgical tools from RGB images. Sarikaya *et al.* [78] additionally fed the optical flow estimated from color images into the network. LSTM was integrated to smoothen the detection results to leverage spatiotemporal information [77]. In addition to position tracking, the pose of the articulated end-effector was simultaneously estimated by the methods proposed by Ye *et al.* [75] and Kurmann*et al.* [79].

Interaction between surgical tools and environment

A representative example of tool–tissue interaction during surgery is suturing. In this task, the robot needs to recover the 2D or 3D shape of thread from 2D images in real time. Other challenges in this task include the deformation of thread and variations in the environment. Padoy and Hager [80] introduced a Markov random field-based optimization method to track the 3D thread modeled by a nonuniform rational B-spline.

Recently, a supervised two-branch CNN, called deep multistage detection (DMSD), was proposed for surgical thread detection [81]. In addition, the DMSD framework
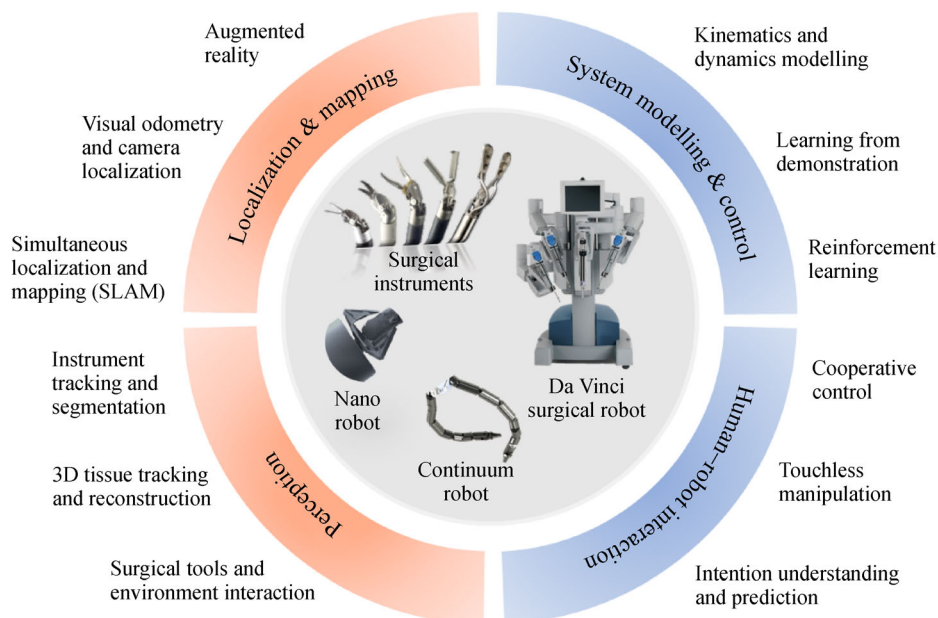
Fig. 3 AI techniques for surgical robotics, including perception, localization and mapping, system modeling and control, and human–robot interaction.

was improved with a CycleGAN [48] structure to perform domain adaption for the foreground and background [82]. On the basis of adversarial learning, additional synthetic data for thread detection were generated while preserving the semantic information that enabled the transfer of learned knowledge to the target domain.

Estimation of the interaction forces between surgical instruments and tissues can provide meaningful feedback to ensure safe robotic manipulation.

Recent works have incorporated AI techniques in the field of vision-based force sensing, which can accurately estimate the force values from visual inputs. The LSTM-RNN architecture can automatically learn accurate mapping between visual–geometric information and applied force in a supervised manner [83]. In addition to supervised learning, a semisupervised DCNN was proposed by Marban *et al*. [84], where the convolutional auto-encoder learns the representation from RGB images, followed by minimizing the error between the estimated force and ground truth using LSTM.

System modeling and control

Learning from human demonstrations

Learning from demonstration (LfD), also known as programming by demonstration, imitation learning, and apprenticeship learning, is a popular paradigm for enabling robots to perform autonomously new tasks with learned policies. This paradigm is beneficial for complicated automation tasks, such as surgical procedures, for which surgical robots can autonomously execute specific motions or tasks through learning from surgeons' demonstrations without tedious programming procedures. The robots can reduce surgeons' tedium, as well as provide superhuman performance in terms of execution speed and smoothness. The common framework of LfD is to first segment a complicated surgical task into several motion primitives or subtasks, followed by recognition, modeling, and execution of these subtasks sequentially.

Surgical task segmentation and recognition

The JHU-ISI Gesture and Skill Assessment Working Set data set [85] is the first publicly available benchmark data set for surgical activity segmentation and recognition. This data set contains synchronized video and kinematic data of three subtasks captured from the Da Vinci robot: suturing, needle passing, and knot tying. Unsupervised clustering algorithms are the most popular for surgical task segmentation. Fard *et al*. [86] proposed a soft boundary-modified Gath–Geva clustering algorithm for segmenting kinematic data. A transition state clustering (TSC) method

[87] was presented to exploit both the video and kinematic data to detect and cluster transitions between linear dynamic regimes on the basis of kinematic, sensory, and temporal similarities. The TSC method was further improved by applying DCNNs to extract features from video data [88]. For surgical subtask recognition, most previous methods [85,89,90] were developed according to variations in hidden Markov model (HMM), conditional random field (CRF), and linear dynamic systems (LDS). Particularly, joint segmentation and recognition frameworks were proposed by Despinoy *et al*. [91] and DiPietro *et al* [92]. DiPietro *et al*. [92] specifically modeled complex and nonlinear dynamics of kinematic data with RNN to recognize both surgical gestures and activities, where the simple RNN, forward LSTM, bidirectional LSTM, gated recurrent unit, and mixed history RNN were compared with traditional methods. Liu and Jiang [93] introduced a novel method by modeling the recognition task as a sequential decision-making process and trained an agent by RL with hierarchical features from a DCNN model.

Surgical task modeling, generation, and execution

After acquiring the segmented motion trajectories representing surgical subtasks, the dynamic time warping algorithm can be applied to align temporally different demonstrations before modeling. To generate the motion in a new task autonomously, previous works extensively studied Gaussian mixture model (GMM) [94,95], Gaussian process regression (GPR) [96], dynamics model [97], finite state machine [98], and RNN [99] for modeling the demonstrated trajectories. Experts' demonstrations are encoded by the GMM algorithm, and the parameters of mixture model can be iteratively estimated by the expectation maximization algorithm. With a given GMM, the Gaussian mixture regression is then used to generate the target trajectory of the desired surgical task [94,95]. GPR is a nonlinear Bayesian function learning technique that models a sequence of observations generated by a Gaussian process. Osa *et al.* [96] chose GPR for online path planning in a dynamic environment. Given the predicted motion trajectory, different control strategies, e.g., linear–quadratic regulator controller [97], sliding mode control [96], and neural network [100], can be applied to improve robustness in surgical task execution.

Reinforcement learning

In many surgical tasks, reinforcement learning (RL) is another popular machine-learning paradigm to solve problems, such as control of continuum robots, soft tissue

manipulation, and tube insertion, that are difficult to model analytically and explicitly observe [101]. In the learning process, the controller of the autonomous surgical robot, known as an agent, attempts to find the optimized policies that yield highly accumulated reward through iterative interaction with the surrounding environment. The RL environment is modeled as a Markov decision process. The RL algorithm can be initialized with the learned policies from expert demonstrations instead of learning from scratch to reduce efficiently the learning time [95,102,103]. Tan *et al.* [103] trained a generative adversarial imitation learning [104] agent to imitate latent patterns existing in human demonstrations. This agent can deal with mismatched distributions caused by multimodal behaviors. DRL with advanced policy search methods allowed robots to execute autonomously a wide range of tasks [105]. However, repeating these experiments on a surgical robotic platform over a million times is unrealistic. To this end, the agent can be first trained in a simulation environment and then transferred to a real robot [106]. The agent can first learn tensioning policies from a finite-element simulator via DRL, and then it can be transferred to a real physical system. However, the discrepancy between simulation data and real-world environment needs to be reconciled.

Human–robot interaction

Human–robot interaction (HRI) is a field that integrates knowledge and techniques from multiple disciplines to build effective communication between humans and robots. With the help of AI, surgical task-oriented HRI allows surgeons to control cooperatively surgical robotic systems with touchless manipulation. Interaction media between surgeons and intelligent robots are usually through surgeons' gaze, head movement, speech/voice, and hand gestures. By understanding the intention of humans, robots can then perform the most appropriate actions that satisfy the surgeons' needs.

Tracking 2D/3D eye-gaze points of surgeons can assist surgical instrumental control and navigation [107]. For surgical robots, the eye-gaze contingent paradigm can facilitate the transmission of images and enhance procedure performance, thereby enabling more accurate navigation of the instruments [107]. Yang *et al.* [108] first introduced the concept of gaze-contingent perceptual docking for robot-assisted MIS, in which the robot can learn the operators' specific motor and perceptual behavior through their saccadic eye movements and ocular vergence. Inspired by this idea, Visentini-Scarzanella *et al.* [109] used the gaze-contingent docking to reconstruct the surgeon's area of interest with a Bayesian chains method in real time. Fujii *et al.* [110] performed gaze gesture recognition with the HMM so as to pan, zoom, and tilt the laparoscope during surgery. In addition to human

gaze, surgeons' head movements can also be used to control remotely a laparoscope or endoscope [111,112].

Robots have the potential to interpret human intentions or commands through voice commands. However, robot assistance during surgery remains challenging because of the noisy environment in the operation room. With the development of deep learning in speech recognition, the precision and accuracy of speech recognition have considerably improved [113], thus leading to more reliable control of the surgical robot [114].

Hand gesture is another popular medium in different HRI scenarios. Learning-based real-time hand gesture detection and recognition methods have been developed by harnessing different sensors. Jacob *et al.* [115] designed a robotic scrub nurse, called Gestonurse, to understand nonverbal hand gestures. They used the Kinect sensor to localize and recognize different gestures generated by surgeons to help in delivering surgical instruments to surgeons. Wen *et al.* introduced an HMM-based hand gesture recognition method for AR control [116]. More recently, vision-based hand gesture recognition with high precision [117] can be achieved with the help of deep learning. This development can therefore substantially improve the HRI safety in surgery.

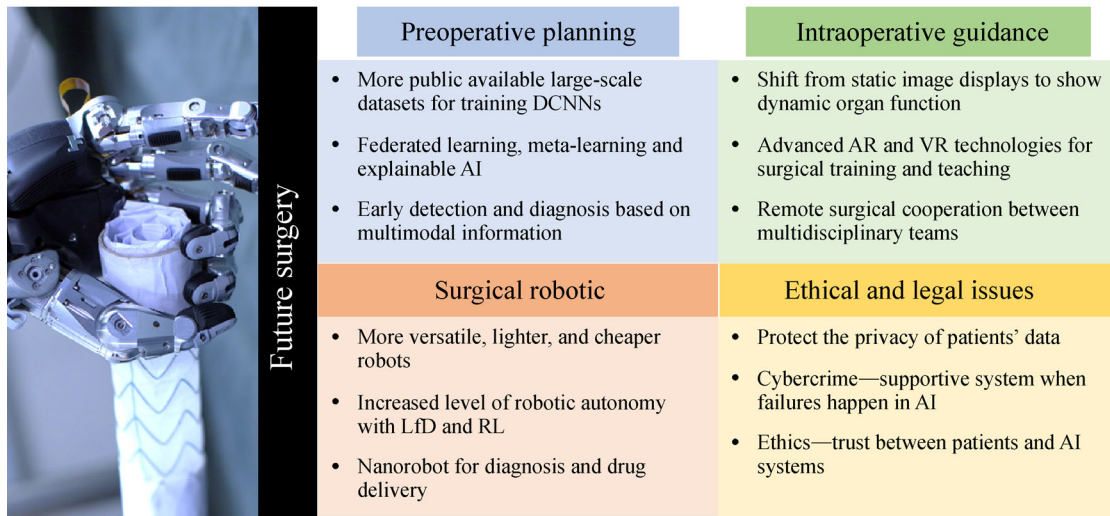## Conclusion and future outlook

The advancement in AI has been transforming modern surgery toward more precise and autonomous intervention for treating both acute and chronic symptoms. By leveraging such techniques, notable progress has been achieved in preoperative planning, intraoperative guidance, and surgical robotics. Herein, we summarize the major challenges for these three aspects (Fig. 4). We then discuss achievable visions of future research directions. Finally, we further examine other key issues, such as ethics, regulation, and privacy.

Preoperative planning

Deep learning has been widely adopted in preoperative planning for tasks ranging from anatomical classification, detection, and segmentation to image registration. The results seem to suggest that deep learning-based methods can outperform those which rely on conventional approaches. However, data-driven approaches often suffer from inherited limitations, making deep learning-based approaches less generalizable, explainable, and more data-demanding.

To overcome these issues, close collaborations between multidisciplinary teams should be encouraged, particularly between surgeons and AI researchers, to generate large-scale annotated data that will provide more training data for AI algorithms. An alternative solution is to develop AI

Preoperative planning	Intraoperative guidance
• More public available large-scale datasets for training DCNNs • Federated learning, meta-learning and explainable AI • Early detection and diagnosis based on multimodal information	• Shift from static image displays to show dynamic organ function • Advanced AR and VR technologies for surgical training and teaching • Remote surgical cooperation between multidisciplinary teams
Surgical robotic	Ethical and legal issues
• More versatile, lighter, and cheaper robots • Increased level of robotic autonomy with LfD and RL • Nanorobot for diagnosis and drug delivery	• Protect the privacy of patients' data • Cybercrime—supportive system when failures happen in AI • Ethics—trust between patients and AI systems

Fig. 4 An outlook of the future of surgery in preoperative planning, intraoperative guidance, surgical robotics, and potential ethical and legal issues.

techniques, such as meta-learning, or learning to learn, that enable generalizable systems to perform diagnosis with limited data sets yet improved explainability.

Although many state-of-the-art machine-learning and deep-learning algorithms have made breakthroughs in the field of general computer vision, the differences between medical and natural images can be massive and thus may impede their clinical applicability. In addition, the underlying models and the derived results may not be easily interpretable by humans, a condition that raises important issues, such as potential risks and uncertainty in surgery. Potential solutions to these problems would be to explore different transfer learning techniques to mitigate the differences between image modalities and develop more explainable AI algorithms to enhance its decision-making performance.

Furthermore, utilizing personalized multimodal patient information, including omics-data and lifestyle information, in AI development can be useful in early detection and diagnosis, thereby leading to personalized treatment. These improvements also allow early treatment options that result in minimal trauma, low surgical risks, and short recovery time.

Intraoperative guidance

AI techniques have already contributed to more accurate and robust intraoperative guidance for MIS. 3D shape instantiation, camera pose estimation, and dynamic environment tracking and reconstruction have been tackled to assist various surgical interventions.

The key points in developing computer-assisted guidance from visual observations should be improving the localization and mapping performance with textureless surfaces, variation in illumination, and limited field of view.

Another major challenge is organ/tissue deformation that complicates surgery with a dynamic and uncertain environment despite extensive preoperative planning. Although AI technologies have succeeded in detection, segmentation, tracking, and classification, further studies are warranted to extend these processes to more sophisticated 3D applications. Additionally, during a surgery, an important requirement for an AI algorithm is its ability to assist surgeons in real time efficiently. Such demands have been encountered in developing AR or VR where frequent interactions are required either between surgeons and autonomous guidance systems or during remote surgery involving multidisciplinary teams located in different geographical locations.

Aside from visual information, future AI technologies must fuse multimodal data from various sensors to achieve more precise perception of the complicated environment. Furthermore, increasing the use of micro- and nanorobotics in surgery will generate new guidance issues.

Surgical robotics

With the integration of AI, surgical robotics would be able to perceive and understand complicated surroundings, conduct real-time decision-making, and perform surgical tasks with increased precision, safety, automation, and efficiency. For instance, current robots can already automatically perform some simple surgical tasks, such as suturing and knot tying [118,119]. Nevertheless, the increased level of robotic autonomy for more complicated tasks can be achieved by advanced LfD and RL algorithms, especially when considering interactions with dynamic environments. Owing to the diversity of surgical

robotic platforms, generalized learning for accurate modeling and control is also required.

Most of the current surgical robots are expensive, bulky, and can only perform master–slave operations. We emphasize that a more versatile, lighter, and probably cheaper robotic system should be developed so that it can access more constrained regions during MIS [2]. Certainly, it also needs to be easily integrated in well-developed surgical workflows so that the robot can seamlessly collaborate with human operators. To date, the current technologies in RAS are still far from achieving full autonomy; human supervision would remain to ensure safety and high-level decision-making.

In the near future, intelligent micro- and nanorobots for noninvasive surgeries and drug delivery could be realized. Furthermore, with the data captured during preoperative examinations, robots could also assist in the manufacturing of personalized 3D bioprinted tissues and organs for transplant surgery.

Ethical and legal considerations of AI in surgery

Beyond precision, robustness, safety, and automation, we must also carefully consider the legal and ethical issues related to AI in surgery. These issues include the following: (1) privacy: patients' medical records, genetic data, illness prediction data, and operation process data must be protected with high security. (2) Cybercrime: the negative effects on patients should be minimized when failures happen in AI-based surgical systems, which should be verified and certificated while considering all possible risks. (3) Concerned parties should adhere to a code of ethics to ensure that new technologies, such as gene editing and bioprinted organ transplant for long-term human reproduction, are used responsibly and to build trust between humans and AI techniques gradually.

In conclusion, we still have a long way to go to replicate and match in robotic surgery the level of intelligence that surgeons display. AIs that can learn complex tasks on their own and with a minimum of initial training data will prove critical for next-generation systems [120]. Here we quote some of the questions raised by Yang *et al.* in their article on Medical Robotics [121]: *"As the capabilities of medical robotics following a progressive path represented by various levels of autonomy evolve, most of the role of the medical specialists will shift toward diagnosis and decision-making. Could this shift also mean that medical specialists will be less skilled in terms of dexterity and basic surgical skills as the technologies are introduced? What would be the implication on future training and accreditation? If robot performance proves to be superior to that of humans, should we put our trust in fully autonomous medical robots?"* Clearly, various issues must be addressed before AI can be more seamlessly integrated in the future of surgery.

Compliance with ethics guidelines

References

1. Vitiello V, Lee SL, Cundy TP, Yang GZ. Emerging robotic platforms for minimally invasive surgery. IEEE Rev Biomed Eng 2013; 6: 111–126

2. Troccaz J, Dagnino G, Yang GZ. Frontiers of medical robotics: from concept to systems to clinical translation. Annu Rev Biomed Eng 2019; 21(1): 193–218

3. Yang GZ. Body Sensor Networks. New York: Springer, 2014

4. Yang GZ. Implantable Sensors and Systems: from Theory to Practice. New York: Springer, 2018

5. Shortliffe E. Computer-Based Medical Consultations: MYCIN. Amsterdam: Elsevier, 2012. Vol. 2

6. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems (NIPS). Lake Tahoe. 2012: 1097–1105

7. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, van der Laak JAWM, van Ginneken B, Sánchez CI. A survey on deep learning in medical image analysis. Med Image Anal 2017; 42: 60–88

8. Khosravi P, Kazemi E, Imielinski M, Elemento O, Hajirasouliha I. Deep convolutional neural networks enable discrimination of heterogeneous digital pathology images. EBioMedicine 2018; 27: 317–328

9. Chilamkurthy S, Ghosh R, Tanamala S, Biviji M, Campeau NG, Venugopal VK, Mahajan V, Rao P, Warier P. Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study. Lancet 2018; 392(10162): 2388–2396

10. Meyer A, Zverinski D, Pfahringer B, Kempfert J, Kuehne T, Sündermann SH, Stamm C, Hofmann T, Falk V, Eickhoff C. Machine learning for real-time prediction of complications in critical care: a retrospective study. Lancet Respir Med 2018; 6(12): 905–914

11. Li X, Zhang S, Zhang Q, Wei X, Pan Y, Zhao J, Xin X, Qin C, Wang X, Li J, Yang F, Zhao Y, Yang M, Wang Q, Zheng Z, Zheng X, Yang X, Whitlow CT, Gurcan MN, Zhang L, Wang X, Pasche BC, Gao M, Zhang W, Chen K. Diagnosis of thyroid cancer using deep convolutional neural network models applied to sonographic images: a retrospective, multicohort, diagnostic study. Lancet Oncol 2019; 20(2): 193–201

12. Rubinstein E, Salhov M, Nidam-Leshem M, White V, Golan S, Baniel J, Bernstine H, Groshar D, Averbuch A. Unsupervised tumor detection in dynamic PET/CT imaging of the prostate. Med Image Anal 2019; 55: 27–40

13. Winkels M, Cohen TS. Pulmonary nodule detection in CT scans with equivariant CNNs. Med Image Anal 2019; 55: 15–26

14. Maicas G, Carneiro G, Bradley AP, Nascimento JC, Reid I. Deep

reinforcement learning for active breast lesion detection from DCE-MRI. In: Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). New York: Springer, 2017: 665–673

15. Lee H, Yune S, Mansouri M, Kim M, Tajmir SH, Guerrier CE, Ebert SA, Pomerantz SR, Romero JM, Kamalian S, Gonzalez RG, Lev MH, Do S. An explainable deep-learning algorithm for the detection of acute intracranial haemorrhage from small datasets. Nat Biomed Eng 2019; 3(3): 173–182

16. Kamnitsas K, Ledig C, Newcombe VFJ, Simpson JP, Kane AD, Menon DK, Rueckert D, Glocker B. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Med Image Anal 2017; 36: 61–78

17. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston. 2015: 3431–3440

18. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). New York: Springer, 2015: 234–241

19. Cicek O, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). New York: Springer, 2016: 424–432

20. Zhou XY, Yang GZ. Normalization in training U-Net for 2D biomedical semantic segmentation. IEEE Robot Autom Lett 2019; 4(2): 1792–1799

21. Gibson E, Giganti F, Hu Y, Bonmati E, Bandula S, Gurusamy K, Davidson B, Pereira SP, Clarkson MJ, Barratt DC. Automatic multi-organ segmentation on abdominal CT with dense V-networks. IEEE Trans Med Imaging 2018; 37(8): 1822–1834

22. Wang G, Li W, Zuluaga MA, Pratt R, Patel PA, Aertsen M, Doel T, David AL, Deprest J, Ourselin S, Vercauteren T. Interactive medical image segmentation using deep learning with image-specific fine tuning. IEEE Trans Med Imaging 2018; 37(7): 1562–1573

23. Laina I, Rieke N, Rupprecht C, Vizca'ıno JP, Eslami A, Tombari F, Navab N. Concurrent segmentation and localization for tracking of surgical instruments. In: Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). New York: Springer, 2017: 664–672

24. Feng X, Yang J, Laine AF, Angelini ED. Discriminative localization in CNNs for weakly-supervised segmentation of pulmonary nodules. In: Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). New York: Springer, 2017: 568–576

25. Bai W, Chen C, Tarroni G, Duan J, Guitton F, Petersen SE, Guo Y, Matthews PM, Rueckert D. Self-supervised learning for cardiac MR image segmentation by anatomical position prediction. In: International Conference on Medical Image Computing and Computer Assisted Intervention. New York: Springer, 2019: 541–549

26. Balakrishnan G, Zhao A, Sabuncu MR, Guttag J, Dalca AV. VoxelMorph: a learning framework for deformable medical image registration. IEEE Trans Med Imaging 2019: 38(8): 1788–1800

27. Shen Z, Han X, Xu Z, Niethammer M. Networks for joint affine and non-parametric image registration. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach. 2019: 4224–4233

28. Hu Y, Modat M, Gibson E, Li W, Ghavami N, Bonmati E, Wang G, Bandula S, Moore CM, Emberton M, Ourselin S, Noble JA, Barratt DC, Vercauteren T. Weakly-supervised convolutional neural networks for multimodal image registration. Med Image Anal 2018; 49: 1–13

29. Miao S, Piat S, Fischer P, Tuysuzoglu A, Mewes P, Mansi T, Liao R. Dilated FCN for multi-agent 2D/3D medical image registration. In: Proceedings of AAAI Conference on Artificial Intelligence. New Orleans. 2018

30. Sokooti H, de Vos B, Berendsen F, Lelieveldt BP, I˘sgum I, Staring M. Nonrigid image registration using multi-scale 3D convolutional neural networks. In: Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). New York: Springer, 2017: 232–239

31. Liao R, Miao S, de Tournemire P, Grbic S, Kamen A, Mansi T, Comaniciu D. An artificial agent for robust image registration. In: Proceedings of AAAI Conference on Artificial Intelligence. San Francisco. 2017

32. Cool D, Downey D, Izawa J, Chin J, Fenster A. 3D prostate model formation from non-parallel 2D ultrasound biopsy images. Med Image Anal 2006; 10(6): 875–887

33. Zhou X, Yang G, Riga C, Lee S. Stent graft shape instantiation for fenestrated endovascular aortic repair. In: The Hamlyn Symposium on Medical Robotics. London. 2017

34. Zhou XY, Lin J, Riga C, Yang GZ, Lee SL. Real-time 3D shape instantiation from single fluoroscopy projection for fenestrated stent graft deployment. IEEE Robot Autom Lett 2018; 3(2): 1314–1321

35. Zheng JQ, Zhou XY, Riga C, Yang GZ. Real-time 3D shape instantiation for partially deployed stent segments from a single 2D fluoroscopic image in fenestrated endovascular aortic repair. IEEE Robot Autom Lett 2019; 4(4): 3703–3710

36. Zhou XY, Riga C, Lee SL, Yang GZ. Towards automatic 3D shape instantiation for deployed stent grafts: 2D multiple-class and class-imbalance marker segmentation with equally-weighted focal U-Net. In: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Madrid: IEEE, 2018: 1261–1267

37. Zheng JQ, Zhou XY, Riga C, Yang GZ. Towards 3D path planning from a single 2D fluoroscopic image for robot assisted fenestrated endovascular aortic repair. In: 2019 International Conference on Robotics and Automation (ICRA). Montreal: IEEE, 2019: 8747–8753

38. Lee SL, Chung A, Lerotic M, Hawkins MA, Tait D, Yang GZ. Dynamic shape instantiation for intra-operative guidance. In: Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). New York: Springer, 2010: 69–76

39. Zhou XY, Yang GZ, Lee SL. A real-time and registration-free framework for dynamic shape instantiation. Med Image Anal 2018; 44: 86–97

40. Zhou XY, Wang ZY, Li P, Zheng JQ, Yang GZ. One stage shape instantiation from a single 2D image to 3D point cloud. In:

International Conference on Medical Image Computing and Computer Assisted Intervention. New York: Springer, 2019: 30–38

41. Mahmood F, Durr NJ. Deep learning and conditional random fields-based depth estimation and topographical reconstruction from conventional endoscopy. Med Image Anal 2018; 48: 230–243

42. Mahmood F, Chen R, Durr NJ. Unsupervised reverse domain adaptation for synthetic medical images via adversarial training. IEEE Trans Med Imaging 2018; 37(12): 2572–2581

43. Turan M, Ornek EP, Ibrahimli N, Giracoglu C, Almalioglu Y, Yanik MF, Sitti M. Unsupervised odometry and depth learning for endoscopic capsule robots. In: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Madrid: IEEE, 2018: 1801–1807

44. Shen M, Gu Y, Liu N, Yang GZ. Context-aware depth and pose estimation for bronchoscopic navigation. IEEE Robot Autom Lett 2019; 4(2): 732–739

45. Zhou T, Brown M, Snavely N, Lowe DG. Unsupervised learning of depth and ego-motion from video. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Hawaii. 2017

46. Zhan H, Garg R, Saroj Weerasekera C, Li K, Agarwal H, Reid I. Unsupervised learning of monocular depth estimation and visual odometry with deep feature reconstruction. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Salt Lake City. 2018

47. Ye M, Johns E, Handa A, Zhang L, Pratt P, Yang GZ. Selfsupervised siamese learning on stereo image pairs for depth estimation in robotic surgery. In: The Hamlyn Symposium on Medical Robotics. London. 2017: 27

48. Zhu JY, Park T, Isola P, Efros AA. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV). Venice. 2017: 2223–2232

49. Turan M, Almalioglu Y, Araujo H, Konukoglu E, Sitti M. Deep endovo: a recurrent convolutional neural network (RCNN) based visual odometry approach for endoscopic capsule robots. Neurocomputing 2018; 275: 1861–1870

50. Sganga J, Eng D, Graetzel C, Camarillo D. Offsetnet: deep learning for localization in the lung using rendered images. In: 2019 International Conference on Robotics and Automation (ICRA). Montreal: IEEE, 2019: 5046–5052

51. Mountney P, Stoyanov D, Davison A, Yang GZ. Simultaneous stereoscope localization and soft-tissue mapping for minimal invasive surgery. In: Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). New York: Springer, 2006: 347–354

52. Davison AJ, Reid ID, Molton ND, Stasse O. MonoSLAM: real-time single camera SLAM. IEEE Trans Pattern Anal Mach Intell 2007; 29(6): 1052–1067

53. Mountney P, Yang GZ. Motion compensated SLAM for image guided surgery. In: Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). New York: Springer, 2010: 496–504

54. Grasa OG, Bernal E, Casado S, Gil I, Montiel JM. Visual SLAM for handheld monocular endoscope. IEEE Trans Med Imaging 2014; 33(1): 135–146

55. Turan M, Almalioglu Y, Araujo H, Konukoglu E, Sitti M. A non-rigid map fusion-based direct SLAM method for endoscopic capsule robots. Int J Intell Robot Appl 2017; 1(4): 399–409

56. Song J, Wang J, Zhao L, Huang S, Dissanayake G. MISSLAM: real-time large-scale dense deformable SLAM system in minimal invasive surgery based on heterogeneous computing. IEEE Robot Autom Lett 2018; 3(4): 4068–4075

57. Zhou XY, Ernst S, Lee SL. Path planning for robot-enhanced cardiac radiofrequency catheter ablation. In: 2016 IEEE international conference on robotics and automation (ICRA). Stockholm: IEEE, 2016: 4172–4177

58. Shi C, Giannarou S, Lee SL, Yang GZ. Simultaneous catheter and environment modeling for trans-catheter aortic valve implantation. In: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Chicago: IEEE, 2014: 2024–2029

59. Zhao L, Giannarou S, Lee SL, Yang GZ. SCEM +: real-time robust simultaneous catheter and environment modeling for endovascular navigation. IEEE Robot Autom Lett 2016; 1(2): 961–968

60. Zhao L, Giannarou S, Lee SL, Yang GZ. Registration-free simultaneous catheter and environment modelling. In: Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). New York: Springer, 2016: 525–533

61. Mountney P, Yang GZ. Soft tissue tracking for minimally invasive surgery: learning local deformation online. In: Proceedings of International Conference on Medical Image Computing and ComputerAssisted Intervention (MICCAI). New York: Springer, 2008: 364–372

62. Ye M, Giannarou S, Meining A, Yang GZ. Online tracking and retargeting with applications to optical biopsy in gastrointestinal endoscopic examinations. Med Image Anal 2016; 30: 144–157

63. Wang R, Zhang M, Meng X, Geng Z, Wang FY. 3D tracking for augmented reality using combined region and dense cues in endoscopic surgery. IEEE J Biomed Health Inform 2018; 22(5): 1540–1551

64. Bernhardt S, Nicolau SA, Soler L, Doignon C. The status of augmented reality in laparoscopic surgery as of 2016. Med Image Anal 2017; 37: 66–90

65. Wang J, Suenaga H, Hoshi K, Yang L, Kobayashi E, Sakuma I, Liao H. Augmented reality navigation with automatic marker-free image registration using 3-D image overlay for dental surgery. IEEE Trans Biomed Eng 2014; 61(4): 1295–1304

66. Pratt P, Ives M, Lawton G, Simmons J, Radev N, Spyropoulou L, Amiras D. Through the HoloLens^{TM} looking glass: augmented reality for extremity reconstruction surgery using 3D vascular models with perforating vessels. Eur Radiol Exp 2018; 2(1): 2

67. Zhang X, Wang J, Wang T, Ji X, Shen Y, Sun Z, Zhang X. A markerless automatic deformable registration framework for augmented reality navigation of laparoscopy partial nephrectomy. Int J CARS 2019; 14(8): 1285–1294

68. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019; 25(1): 44–56

69. Mirnezami R, Ahmed A. Surgery 3.0, artificial intelligence and the next-generation surgeon. Br J Surg 2018; 105(5): 463–465

70. Bouget D, Benenson R, Omran M, Riffaud L, Schiele B, Jannin P. Detecting surgical tools by modelling local appearance and global shape. IEEE Trans Med Imaging 2015; 34(12): 2603–2617

71. Shvets AA, Rakhlin A, Kalinin AA, Iglovikov VI. Automatic instrument segmentation in robot-assisted surgery using deep learning. In: Proceedings of IEEE International Conference on Machine Learning and Applications (ICMLA). Stockholm: IEEE, 2018: 624–628

72. Islam M, Atputharuban DA, Ramesh R, Ren H. Real-time instrument segmentation in robotic surgery using auxiliary supervised deep adversarial learning. IEEE Robot Autom Lett 2019; 4(2): 2188–2195

73. Sznitman R, Richa R, Taylor RH, Jedynak B, Hager GD. Unified detection and tracking of instruments during retinal microsurgery. IEEE Trans Pattern Anal Mach Intell 2013; 35(5): 1263–1273

74. Zhang L, Ye M, Chan PL, Yang GZ. Real-time surgical tool tracking and pose estimation using a hybrid cylindrical marker. Int J CARS 2017; 12(6): 921–930

75. Ye M, Zhang L, Giannarou S, Yang GZ. Real-time 3D tracking of articulated tools for robotic surgery. In: Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). New York: Springer, 2016: 386–394

76. Zhao Z, Voros S, Weng Y, Chang F, Li R. Tracking-by-detection of surgical instruments in minimally invasive surgery via the convolutional neural network deep learning-based method. Comput Assist Surg (Abingdon) 2017; 22(sup1): 26–35

77. Nwoye CI, Mutter D, Marescaux J, Padoy N. Weakly supervised convolutional LSTM approach for tool tracking in laparoscopic videos. Int J CARS 2019; 14(6): 1059–1067

78. Sarikaya D, Corso JJ, Guru KA. Detection and localization of robotic tools in robot-assisted surgery videos using deep neural networks for region proposal and detection. IEEE Trans Med Imaging 2017; 36(7): 1542–1549

79. Kurmann T, Neila PM, Du X, Fua P, Stoyanov D, Wolf S, Sznitman R. Simultaneous recognition and pose estimation of instruments in minimally invasive surgery. In: Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). New York: Springer, 2017: 505–513

80. Padoy N, Hager GD. 3D thread tracking for robotic assistance in tele-surgery. In: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). San Francisco: IEEE, 2011: 2102–2107

81. Hu Y, Gu Y, Yang J, Yang GZ. Multi-stage suture detection for robot assisted anastomosis based on deep learning. In: Proceedings of IEEE International Conference on Robotics and Automation (ICRA). Brisbane: IEEE, 2018: 1–8

82. Gu Y, Hu Y, Zhang L, Yang J, Yang GZ. Cross-scene suture thread parsing for robot assisted anastomosis based on joint feature learning. In: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Madrid: IEEE, 2018: 769–776

83. Aviles AI, Alsaleh SM, Hahn JK, Casals A. Towards retrieving force feedback in robotic-assisted surgery: a supervised neuro-recurrent-vision approach. IEEE Trans Haptics 2017; 10(3): 431–443

84. Marban A, Srinivasan V, Samek W, Ferna'ndez J, Casals A. Estimation of interaction forces in robotic surgery using a semisupervised deep neural network model. In: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Madrid: IEEE, 2018: 761–768

85. Ahmidi N, Tao L, Sefati S, Gao Y, Lea C, Haro BB, Zappella L, Khudanpur S, Vidal R, Hager GD. A dataset and benchmarks for segmentation and recognition of gestures in robotic surgery. IEEE Trans Biomed Eng 2017; 64(9): 2025–2041

86. Fard MJ, Ameri S, Chinnam RB, Ellis RD. Soft boundary approach for unsupervised gesture segmentation in robotic-assisted surgery. IEEE Robot Autom Lett 2017; 2(1): 171–178

87. Krishnan S, Garg A, Patil S, Lea C, Hager G, Abbeel P, Goldberg K. Transition state clustering: unsupervised surgical trajectory segmentation for robot learning. Int J Robot Res 2017; 36(13–14): 1595–1618

88. Murali A, Garg A, Krishnan S, Pokorny FT, Abbeel P, Darrell T, Goldberg K. TSC-DL: unsupervised trajectory segmentation of multi-modal surgical demonstrations with deep learning. In: Proceedings of IEEE International Conference on Robotics and Automation (ICRA). Stockholm: IEEE, 2016: 4150–4157

89. Zappella L, Béjar B, Hager G, Vidal R. Surgical gesture classification from video and kinematic data. Med Image Anal 2013; 17(7): 732–745

90. Tao L, Zappella L, Hager GD, Vidal R. Surgical gesture segmentation and recognition. In: Proceedings o International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). New York: Springer, 2013: 339–346

91. Despinoy F, Bouget D, Forestier G, Penet C, Zemiti N, Poignet P, Jannin P. Unsupervised trajectory segmentation for surgical gesture recognition in robotic training. IEEE Trans Biomed Eng 2016; 63(6): 1280–1291

92. DiPietro R, Ahmidi N, Malpani A, Waldram M, Lee GI, Lee MR, Vedula SS, Hager GD. Segmenting and classifying activities in robot-assisted surgery with recurrent neural networks. Int J CARS 2019; 14(11): 2005–2020

93. Liu D, Jiang T. Deep reinforcement learning for surgical gesture segmentation and classification. In: Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). New York: Springer, 2018: 247–255

94. Padoy N, Hager GD. Human-machine collaborative surgery using learned models. In: Proceedings of IEEE International Conference on Robotics and Automation (ICRA). Shanghai: IEEE, 2011: 5285–5292

95. Calinon S, Bruno D, Malekzadeh MS, Nanayakkara T, Caldwell DG. Human-robot skills transfer interfaces for a flexible surgical robot. Comput Methods Programs Biomed 2014; 116(2): 81–96

96. Osa T, Sugita N, Mitsuishi M. Online trajectory planning in dynamic environments for surgical task automation. In: Robotics: Science and Systems. Berkeley. 2014: 1–9

97. Van Den Berg J, Miller S, Duckworth D, Hu H, Wan A, Fu XY, Goldberg K, Abbeel P. Superhuman performance of surgical tasks by robots using iterative learning from human-guided demonstrations. In: Proceedings of IEEE International Conference on Robotics and Automation (ICRA). Alaska: IEEE, 2010: 2074–2081

98. Murali A, Sen S, Kehoe B, Garg A, McFarland S, Patil S, Boyd WD, Lim S, Abbeel P, Goldberg K. Learning by observation for surgical subtasks: multilateral cutting of 3D viscoelastic and 2D orthotropic tissue phantoms. In: Proceedings of IEEE International

Conference on Robotics and Automation (ICRA). Seattle: IEEE, 2015: 1202–1209

99. Mayer H, Gomez F, Wierstra D, Nagy I, Knoll A, Schmidhuber J. A system for robotic heart surgery that learns to tie knots using recurrent neural networks. Adv Robot 2008; 22(13–14): 1521–1537

100. De Momi E, Kranendonk L, Valenti M, Enayati N, Ferrigno G. A neural network-based approach for trajectory planning in robot–human handover tasks. Front Robot AI 2016; 3: 34

101. Kober J, Bagnell JA, Peters J. Reinforcement learning in robotics: a survey. Int J Robot Res 2013; 32(11): 1238–1274

102. Abbeel P, Ng AY. Apprenticeship learning via inverse reinforcement learning. In: Proceedings of International Conference on Machine Learning (ICML). Beijing: ACM, 2004: 1

103. Tan X, Chng CB, Su Y, Lim KB, Chui CK. Robotassisted training in laparoscopy using deep reinforcement learning. IEEE Robot Autom Lett 2019; 4(2): 485–492

104. Ho J, Ermon S. Generative adversarial imitation learning. In: Proceedings of Advances in Neural Information Processing Systems (NIPS). Barcelona. 2016: 4565–4573

105. Levine S, Finn C, Darrell T, Abbeel P. End-to-end training of deep visuomotor policies. J Mach Learn Res 2016; 17(1): 1334–1373

106. Thananjeyan B, Garg A, Krishnan S, Chen C, Miller L, Goldberg K. Multilateral surgical pattern cutting in 2D orthotropic gauze with deep reinforcement learning policies for tensioning. In: Proceedings of IEEE International Conference on Robotics and Automation (ICRA). Singapore: IEEE, 2017: 2371–2378

107. Yang GZ, Dempere-Marco L, Hu XP, Rowe A. Visual search: psychophysical models and practical applications. Image Vis Comput 2002; 20(4): 291–305

108. Yang GZ, Mylonas GP, Kwok KW, Chung A. Perceptual docking for robotic control. In: International Workshop on Medical Imaging and Virtual Reality. New York: Springer, 2008: 21–30

109. Visentini-Scarzanella M, Mylonas GP, Stoyanov D, Yang GZ. I-brush: a gaze-contingent virtual paintbrush for dense 3D reconstruction in robotic assisted surgery. In: Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). New York: Springer, 2009: 353–360

110. Fujii K, Gras G, Salerno A, Yang GZ. Gaze gesture based human robot interaction for laparoscopic surgery. Med Image Anal 2018; 44: 196–214

111. Nishikawa A, Hosoi T, Koara K, Negoro D, Hikita A, Asano S, Kakutani H, Miyazaki F, Sekimoto M, Yasui M, Miyake Y, Takiguchi S, Monden M. Face mouse: a novel human-machine interface for controlling the position of a laparoscope. IEEE Trans Robot Autom 2003; 19(5): 825–841

112. Hong N, Kim M, Lee C, Kim S. Head-mounted interface for intuitive vision control and continuous surgical operation in a surgical robot system. Med Biol Eng Comput 2019; 57(3): 601–614

113. Graves A. Mohamed Ar, Hinton G. Speech recognition with deep recurrent neural networks. In: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing. Vancouver: IEEE, 2013: 6645–6649

114. Zinchenko K, Wu CY, Song KT. A study on speech recognition control for a surgical robot. IEEE Trans Industr Inform 2017; 13 (2): 607–615

115. Jacob MG, Li YT, Akingba GA, Wachs JP. Collaboration with a robotic scrub nurse. Commun ACM 2013; 56(5): 68–75

116. Wen R, Tay WL, Nguyen BP, Chng CB, Chui CK. Hand gesture guided robot-assisted surgery based on a direct augmented reality interface. Comput Methods Programs Biomed 2014; 116(2): 68–80

117. Oyedotun OK, Khashman A. Deep learning in vision-based static hand gesture recognition. Neural Comput Appl 2017; 28(12): 3941–3951

118. Hu Y, Zhang L, Li W, Yang GZ. Robotic sewing and knot tying for personalized stent graft manufacturing. In: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Madrid: IEEE, 2018: 754–760

119. Hu Y, Li W, Zhang L, Yang GZ. Designing, prototyping, and testing a flexible suturing robot for transanal endoscopic micro-surgery. IEEE Robot Autom Lett 2019; 4(2): 1669–1675

120. Yang GZ, Bellingham J, Dupont PE, Fischer P, Floridi L, Full R, Jacobstein N, Kumar V, McNutt M, Merrifield R, Nelson BJ, Scassellati B, Taddeo M, Taylor R, Veloso M, Wang ZL, Wood R. The grand challenges of science robotics. Sc Robot 2018; 3(14): eaar7650

121. Yang GZ, Cambias J, Cleary K, Daimler E, Drake J, Dupont PE, Hata N, Kazanzides P, Martel S, Patel RV, Santos VJ, Taylor RH. Medical roboticsregulatory, ethical, and legal considerations for increasing levels of autonomy. Sci Robot 2017; 2(4): 8638