



Automatic adjustment of laparoscopic pose using deep reinforcement learning

Lingtao Yu¹, Yongqiang Xia¹, Pengcheng Wang¹, and Lining Sun²

¹College of Mechanical and Electrical Engineering, Harbin Engineering University, Harbin, 150001, China

²College of Mechanical and Electrical Engineering, Soochow University, Suzhou, 215031, China

Correspondence: Yongqiang Xia (xyq2017@hrbeu.edu.cn)

Received: 3 November 2021 – Accepted: 1 June 2022 – Published: 28 June 2022

Abstract. Laparoscopic arm and instrument arm control tasks are usually accomplished by an operative doctor. Because of intensive workload and long operative time, this method not only causes the operation not to be flow, but also increases operation risk. In this paper, we propose a method for automatic adjustment of laparoscopic pose based on vision and deep reinforcement learning. Firstly, based on the Deep Q Network framework, the raw laparoscopic image is taken as the only input to estimate the Q values corresponding to joint actions. Then, the surgical instrument pose information used to formulate reward functions is obtained through object-tracking and image-processing technology. Finally, a deep neural network adopted in the Q -value estimation consists of convolutional neural networks for feature extraction and fully connected layers for policy learning. The proposed method is validated in simulation. In different test scenarios, the laparoscopic arm can be well automatically adjusted so that surgical instruments with different postures are in the proper position of the field of view. Simulation results demonstrate the effectiveness of the method in learning the highly non-linear mapping between laparoscopic images and the optimal action policy of a laparoscopic arm.

1 Introduction

In recent years, minimally invasive surgery (MIS) has become more and more important. MIS has been applied in various surgeries, including brain, heart, and liver (Chang and Rattner, 2019). This is due to its many benefits in medical practice. In addition to obvious less invasiveness, there are many advantages such as much less postoperative pain and blood loss. MIS also brings advantages of shorter recovery time and less infective rate, which make them beneficial to both inpatients and clinicians (Davies, 1995).

In a typical celiac minimally invasive robot-assisted surgery procedure, the surgeon is needed to control two or three instrument arms (Pandya et al., 2014). A laparoscopic arm is usually controlled by the doctor themselves or by an assistant. When the surgeon controls both instrument arm and laparoscopic arm, the surgeon must frequently stop to change the laparoscopic viewpoint, which causes the operation to be unstable and not smooth. While directing an assistant to control the laparoscopic arm leads to increased op-

eration time, with a laparoscopic pose automatic adjustment system, the surgeon will not have to manually move the laparoscopic arm in a robot-assisted surgery, and the expert assistant can be eliminated.

At present, many researchers have studied that in the method of laparoscopic pose adjustment, which can be classified as follows.

1. The eye-tracking-based method captures human eye movements through the eyeball positioning device, controls camera movement, and adjusts the camera's post and field of vision. Ali et al. (2008) developed an autonomous eye-gaze-based laparoscopic positioning system, which can keep the user's gaze point region at the centre of the laparoscope viewpoint. Cao et al. (2016) developed a laparoscopic post-adjustment system based on pupil variation using a support vector machine classifier (SVM) and a probabilistic neural network classifier (PNN).
2. The kinematics-based method acquires pose information and vector information such as velocity and accel-

eration of surgical instrument through a sensor and then adjusts the laparoscopic pose using the kinematic model relationship. Sandoval et al. (2021) made the laparoscopic posture able to be automatically adjusted based on the kinematics model and motion data of the surgical instruments. Yu et al. (2017) proposed a method for automatically adjusting the position of the laparoscopic window based on the kinematics model of the laparoscopic arm and surgical field parameters.

3. The vision-based method obtains post and trajectory information of the surgical instrument from laparoscopic imaging through image-processing and visual-tracking technology and then deduces the moving image relationship between the laparoscope and the laparoscopic arm. Shin et al. (2014) proposed a 3D instrument-tracking method using a single camera, which can obtain the 3D positions, roll angle, and grasper angle of the laparoscopic instrument with markers. Zhao et al. (2017) proposed a 2D/3D tracking-by-detection framework, which uses line features to describe the shaft via the RANSAC scheme and uses special image features to depict the end-effector based on deep learning.
4. Other methods: Franz et al. (2014) describe the basic working principles of electromagnetic (EM) tracking systems and summarize the future potential and limitations of EM tracking for medical use. Zinchenko et al. (2015) proposed using combined motion of surgical instruments to suggest robotics arm movements so that the camera can be repositioned, which is called “flag language”.

We present a novel method for laparoscopic pose automatic adjustment based on machine vision and deep reinforcement learning (Kober et al., 2013; Sekkat et al., 2021; Bohez et al., 2017; Zhang et al., 2016) in this paper. The laparoscopic arm will be automatically adjusted according to the condition of the surgical instrument to ensure that the surgical instruments maintain the proper position and size in the operation field.

2 Preliminaries

A celiac minimally invasive robot-assisted surgery system generally consists of a robotic arm (laparoscopic arm) with a laparoscope and two or three robotic arms (instrument arm) with a surgical instrument at the end, as shown in Fig. 1a.

The structure of the surgical robotic arm used in this paper is shown in Fig. 1b. The first four joints form the passive parts that are used for robotic arm pre-operative placement. The last four joints are active parts. During the operation, the doctor controls the active part through a master manipulator to adjust the posture of the end surgical instrument or laparoscope.

3 Laparoscopic pose automatic adjustment method

In this section, we give a total introduction of our method frame for laparoscopic pose automatic adjustment. The method is composed of two parts: the image processing and laparoscopic pose adjustment, as shown in Fig. 2. The former is divided into object detection before operation and tracking to obtain the position and size of the surgical instrument as pose parameters and then determine whether the laparoscopic pose needs to be adjusted through movement decision. Finally, we train a motion controller which is composed of a deep neural network (DNN) and reinforcement learning, taking the laparoscopic image as input and the outputting laparoscopic arm action. It controls the adaptive motion of the laparoscopic arm so that the surgical instrument is in the proper position of laparoscopic vision.

4 Movement decision

In this section, we first describe how to extract features of surgical instruments before operation, and then we use a scale-adaptive algorithm to track surgical instruments. Finally, we describe the movement decision module to determine whether the laparoscope requires movement regulation.

As shown in Fig. 3, we add a simple and specific marker in the surgical instrument. We can indirectly obtain the pose of the surgical instrument by detecting the pose characteristics of the marker, which can avoid the disadvantages of more time and workload for identifying the whole surgical instrument. The design of the marker takes advantage of the rotation invariance of the circular feature. According to the mapping relationship, the circle becomes an ellipse during the rotation, and its long axis size remains unchanged. So, the position information of the surgical instruments is replaced with the coordinates of the circle centre, and the size of the surgical instruments is replaced with the long axis of the circular markers.

Through the relevant image-processing technology, the pose parameters of the surgical instruments can be obtained. First, we obtain the greyscale probability density map of the image by back projection. Then, we can calculate the elliptic equation using the method of ellipse fitting by image inertia moment and obtain centre coordinates and the elliptic long axis value.

To quickly extract the post parameters of the surgical instrument, we need to quickly track the surgical instruments to obtain the adaptive size region of interest. The continuously adaptive mean shift (CAMShift) (Comaniciu and Meer, 2002) algorithm through OpenCV implementation for tracking was used in this paper. The CAMShift algorithm skilfully exploits the mean-shift algorithm by the adaptive region size adjustment step.

Before controlling the laparoscopic arm movement, it is judged based on the visual field state of the surgical instrument and the set motion decision conditions for whether ex-

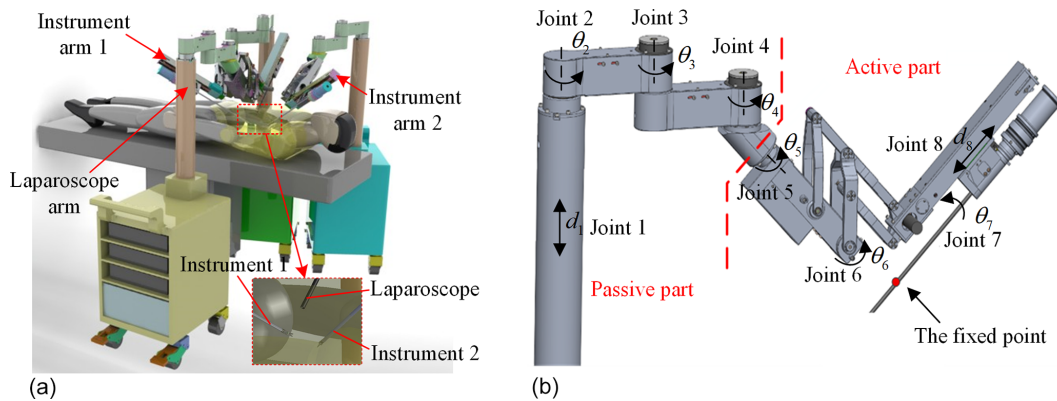


Figure 1. The MIS robotic system and its robotic arm structure. (a) Celiac minimally invasive robot-assisted surgery system. (b) The structure of the surgical robotic arm.

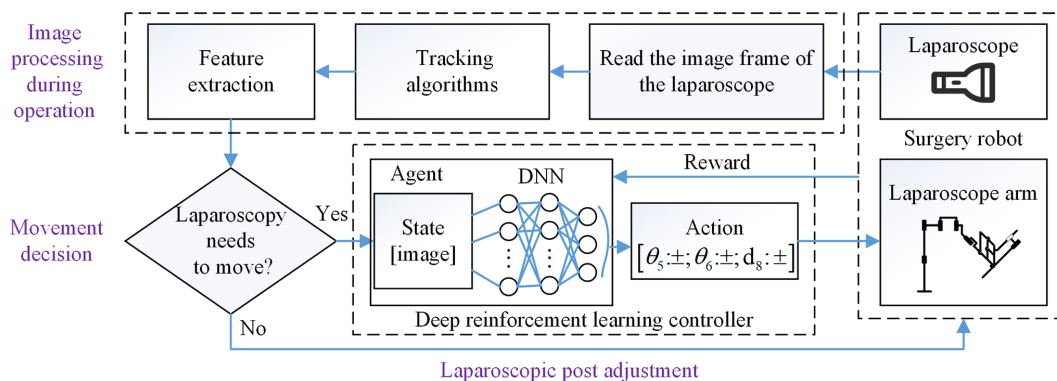


Figure 2. Laparoscopic pose automatic adjustment method.

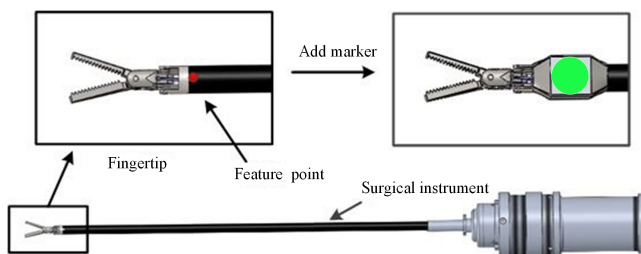


Figure 3. Surgical instrument marker setting.

ercise is required. Specific parameters are as shown in Fig. 4. In this paper, in order to simplify the simulation environment, it is assumed that there is only one surgical instrument in the laparoscopic field of view.

Where the elliptic long-axis value d is used to indicate the relative size of the surgical instrument, the centre coordinate (x, y) shows the position of the surgical instrument in the field of view. T_l and T_h are the lower threshold and upper threshold of the surgical instrument size. T_x and T_y are thresholds of the surgical instrument position. When any one of these thresholds is exceeded, the laparoscopic arm moves.

Setting these thresholds can effectively avoid laparoscopic sensitivity to surgical instrument movement.

5 Laparoscopic motion control based on deep reinforcement learning

In this work, we use deep reinforcement learning to map the laparoscopic image end to end to the joint motion of the laparoscopic arm. In Sect. 5.1 the Markov model of laparoscopic visual field adjustment is introduced, in Sect. 5.2 the network structure of the Deep Q Network (DQN) (Mnih et al., 2015) is introduced, and in Sect. 5.3 the reward function and network training algorithm are introduced.

5.1 Markov model of laparoscopic visual field adjustment

The laparoscopic arm control system makes action decisions based only on the laparoscopic field of view at the current moment, making the active joints of the laparoscopic arm move and obtaining a new laparoscopic field of view for the next step of motion control until the automatic adjustment task of the laparoscopic field of view

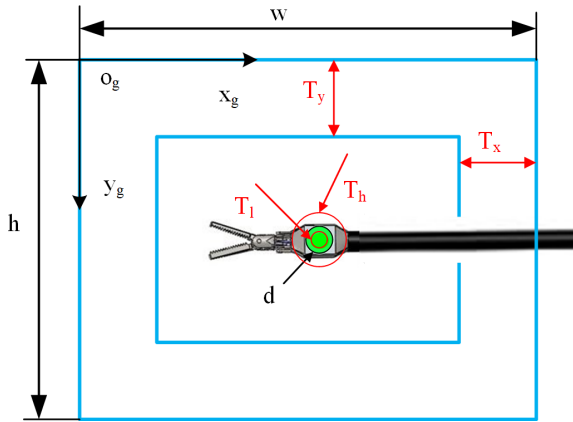


Figure 4. Threshold setting.

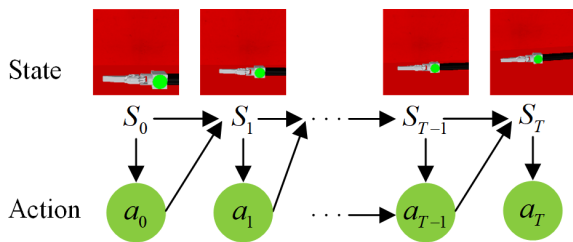


Figure 5. Markov model of laparoscopic visual field adjustment.

is completed, as shown in Fig. 5. Therefore, the problem of automatic adjustment of laparoscopic posture can be expressed as the Markov decision process (MDP), a common model of reinforcement learning (RL) (Mnih et al., 2013; Silver et al., 2014; Gu et al., 2016; Mnih et al., 2016). An MDP comprises a five-tuple $\{S, A, R, P, \gamma\}$. Here S is a finite set of states, A is a finite set of actions, P is the transition probability of states, $R: S \times A \times S \rightarrow R$ is the reward function, and γ is a discount factor. The goal of an RL agent is to learn the strategy $\pi = p(a|s)$ to maximize the total discounted reward $G(t) = \sum_{k=1}^{\infty} \gamma^k R_{t+k+1}$. The action-value function predicts all future rewards when performing the action a at the state s according to the strategy $\pi: Q_{\pi}(s, a) = E_{\pi}[G_t | S_t = s, A_t = a]$. The optimal strategy can be obtained by solving the optimal action-value function: $Q_*(s, a) = R_s^a + \gamma \sum_{s' \in S} P_{ss'}^a \max_{a'} Q_*(s', a')$.

5.2 Deep Q Network

In this work, we use a deep neural network to calculate the $Q(s, a; \theta_i)$ value for performing action a in a specific state s , and the optimal $Q_*(s, a; \theta_i)$ value is obtained by optimizing the neural network parameter θ . This algorithm that combines deep learning and reinforcement learning to achieve end-to-end learning from state to action is the DQN, which can be described as a pioneering work in deep reinforcement learning. When the state s' and R_s^a value obtained by performing action a in state s are unique, the target Q value

output by the neural network in the current iteration i can be expressed as $y_i = r + \gamma \max_{a'} Q(s', a'; \theta_i)$. DeepMind released an article (Mnih et al., 2015) in *Nature*, introducing the concept of a target network to further break the association of the target Q value and the current Q value. The concept of a target network with weight θ^- uses the old deep neural network to get the target Q value. Here are the optimization goals for the DQN with a loss function $L_i(\theta_i) = E[(r + \gamma \max_{a'} Q(s', a'; \theta_i^-) - Q(s, a; \theta_i))^2]$.

The DQN uses the evaluation network to calculate the $Q(s, a; \theta_i)$ and the target network to calculate the $Q(s', a'; \theta_i^-)$. The two networks have the same structure, as shown in Fig. 6. The network uses images obtained from laparoscopy as input, uses convolutional neural networks (CNNs) (Long et al., 2015) to extract features, and uses fully connected (FC) layers for strategy learning. To reduce the size of the feature map and increase the non-linearity, each convolutional layer is connected with a maximum pooling layer and a Relu layer. The image is resized to 128×128 before being forwarded into the network, and it finally obtains a $4 \times 4 \times 64$ visual feature map through a series of convolution and pooling operations.

Then the flattened visual features are fed into policy learning networks on the upper side which consist of two fully connected layers of 256 neurons and 6 neurons respectively. The six neurons in the output layer are the Q values of the six actions of the laparoscopic arm. The six actions are the joint variables of rotary joints 5 and 6 and prismatic joint 8 increasing and decreasing. The rotary joint angle increasing/decreasing step is constant at 2° , and the prismatic joint displacement is 10 mm. The policy learning network on the lower side determines whether the laparoscopic arm needs to be moved during the testing. The specific network configuration is shown in Table 1.

5.3 Design of the reward function and the network training algorithm

The reward of each action is determined according to the coordinates (x, y) and relative size of the surgical instrument (z) , as shown in Eqs. (1) and (2):

$$\text{reward} = \begin{cases} 3 & \text{if } |x - x_0| \leq \frac{t_x}{2} \text{ and } |y - y_0| \leq \frac{t_y}{2} \\ & \text{and } |z - z_0| \leq \frac{t_z}{2}, \\ -3 & \text{if } |x - x_0| \leq \frac{m_x}{2} \text{ or } |y - y_0| \leq \frac{m_y}{2}, \\ & \text{or } |z - z_0| \leq \frac{m_z}{2} \\ \text{reward}_{\text{buff}} & \text{others,} \end{cases} \quad (1)$$

$$\text{reward}_{\text{buff}} = \sum_{i=x,y,z} \lambda_i \left(\frac{2}{m_i - t_i} \times \max(\text{abs}(i - i_0), \frac{t_i}{2}) + \frac{t_i}{(m_i - t_i)} \right), \quad (2)$$

where x_0 and y_0 are the coordinates of the centre point of the image, and t_0 is the set standard size of the surgical instrument. t_x, t_y, t_z represents the tolerable error. When x, y, z is within the tolerable error range, the surgical instrument is

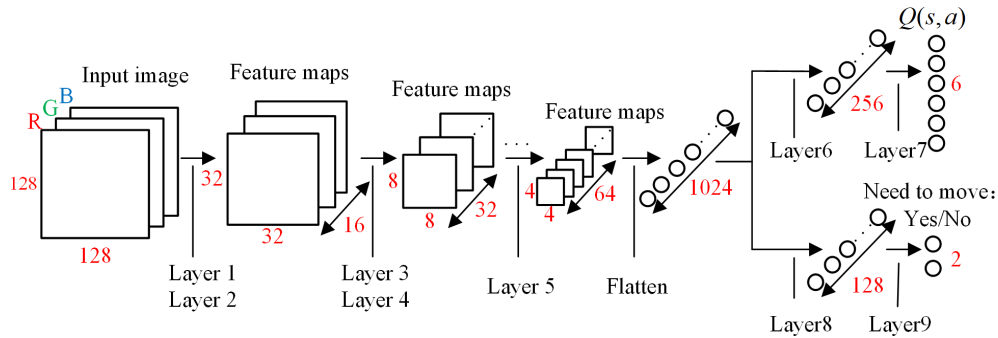


Figure 6. The network structure for the actor estimation.

Table 1. Network architecture configuration.

Layer	Type	Size	Input size	Filter number	Output size
Input	Resize	–	$512 \times 512 \times 3$	–	$128 \times 128 \times 3$
Layer 1	Conv+pool+relu	3×3	$128 \times 128 \times 3$	16	$64 \times 64 \times 16$
Layer 2	Conv+pool+relu	3×3	$64 \times 64 \times 16$	16	$32 \times 32 \times 16$
Layer 3	Conv+pool+relu	3×3	$32 \times 32 \times 16$	32	$16 \times 16 \times 32$
Layer 4	Conv+pool+relu	3×3	$16 \times 16 \times 32$	32	$8 \times 8 \times 32$
Layer 5	Conv+pool+relu	3×3	$8 \times 8 \times 32$	64	$4 \times 4 \times 64$
Layer 6	FC+relu	–	1024	256	256
Layer 7	FC	–	256	6	6
Layer 8	FC	–	1024	128	128
Layer 9	FC	–	128	2	2

in the correct position. m_x, m_y, m_z represents the maximum error. When x, y, z is outside the maximum error range, the surgical instrument is in the wrong position. If the surgical instrument is in the correct position and the wrong position, the task is terminated.

We train the DQN model in a simulation environment to obtain the best network parameters. The training process is shown in Algorithm 1. At the beginning of each epoch, we get the initial state. According to the state, we can calculate the reward and judge whether the task is terminated. When the task continues to execute, the model will give a specific action based on the current state: $action = \arg \max_a Q(s, a, \theta_i)$. The simulated environment has $1 - e$ -greedy probability of executing this specific action and e -greedy probability of randomly choosing an action to execute to explore more states. The simulation environment gets a new state after completing an action.

The current state, action, new state, and reward corresponding to the new state are stored in the memory pool as a set of data. When the simulation environment performs several actions, we will train the model, and when the model performs training several times, the parameters of the target network are replaced with the parameters of the evaluation network.

6 Simulation results and discussion

To verify the feasibility of the DQN-based method in learning laparoscopic pose automatic adjustment, we performed some experiments in simulation scenarios. The simulation settings are described in Sect. 6.1, and the simulation results are reported in Sect. 6.2.

6.1 Simulation settings

In the paper, we used the crossed-platform robotics simulator V-REP (Rohmer et al., 2013) to set up the simulation and test scenarios to validate our method. As shown in Fig. 7, the scene consists of a laparoscopic arm and a surgical instrument. A visual sensor is attached to the laparoscopic arm manipulator and can capture RGB images. The laparoscopic arm structure is shown in Sect. 2, in which the active joints 5, 6, and 8 are in the position control mode. We turn the background colour to red in the software to simulate a real surgical environment. The laparoscopic arm is initialized to the same pose and the surgical instrument is also initialized to the centre of the laparoscopic vision at the beginning of each training round. The surgical instrument moves randomly in the three-dimensional space, and the laparoscopic arm performs a series of actions to make the surgical instrument return to the correct position in the field of view.

Algorithm 1: Training process

Input: Simulation environment: env

Output: neural network parameters: θ

1

Randomly initialize the networks parameters: θ, θ^-

2

Initialize the experience pool: Memory

3

steps = 0

4

while epoch < total_epoch **then**

5

state, reward, terminate \leftarrow env. Random_initialize # Randomly initialize the simulation environmen

6

while not terminate **then**

7

action = DQN. Choose_action (state, θ) # with probab

8

state_new, reward, terminate \leftarrow env. Step (action)

9

Memory. Store (state, action, reward, state_new)

10

every few steps: DQN. Train (Memory, batch_size) # Model learning and update network parameters θ

11

state \leftarrow state_new

12

steps \leftarrow steps +1

13

every some steps: $\theta^- \leftarrow \theta$

14

end

15

end

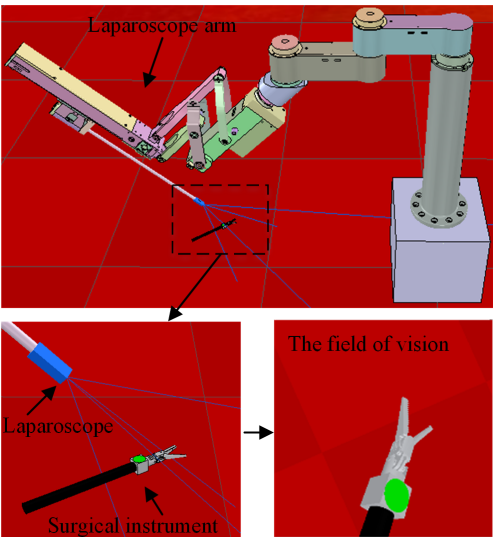


Figure 7. V-REP scene.

V-REP offers a remote application programming interface (API) allowing us to control a simulation from an external application. The remote API on the client side is available for many different programming languages. We implement our method in Python using the TensorFlow library.

The network weight parameters are trained using RM-Sprop optimization with the initial learning rate of 0.002. The

Table 2. Success rate in different test scenarios (%).

Test scenario	Surgical instrument trajectory		
	Random	Line	Curve
Scenario A	100	100	100
Scenario B	100	100	100
Scenario C	100	100	100

reward discount factor γ is set to 0.9, *e-greedy* is set to 0.9 and the batch size is set to 64. The experience replay memory size is set to 20 000. In the training process, the memory pool first collects the experience of the first 200 steps, and then evaluated network parameters are updated every 5 steps. Whenever the evaluation network is updated 200 times, the parameter value of the evaluated network is assigned to the target network. The agent is trained for 8000 epochs.

6.2 Simulation results

In this section, first, we give some parameter curves in the training phases. Secondly, we tested the model in different states to verify the robustness of the model. Finally, we analysed some failure cases.

The success rate in every 100 epochs and the average value of *R* (total reward per epoch) for every 100 epochs are used as indicators to detect the training situation of the model, as

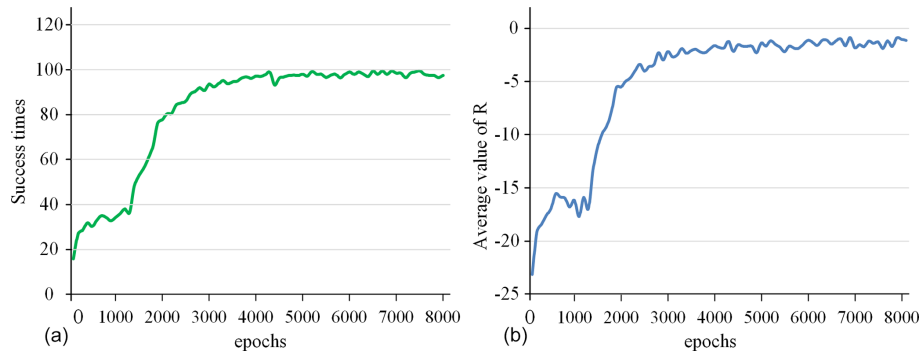


Figure 8. The success rate and the average value of R for all 100 epochs. (a) Success times per 100 epochs. (b) Average value of R per 100 epochs.

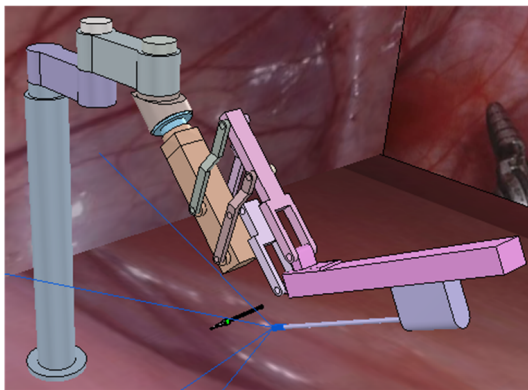


Figure 9. The real surgical image as the background.

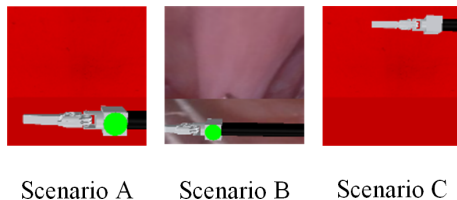


Figure 10. Laparoscopic field of view in different scenarios.

shown in Fig. 8. This curve is drawn based on the average value of five experiments. After 5000 epochs of training, the success rate is close to 100 %, and the average value of R converges to around -2 . This shows that the model parameters have converged, and the task of tracking surgical instruments and adjusting the visual field has been completed. After the model training is completed, the model parameters are fine-tuned in a simulation environment with different backgrounds and the marker removed to adapt to the complex surgical environment.

To simulate the complex surgical environment, three different test scenarios are set up. Scenario A uses the training scene as the test scene. Scenario B uses the real surgical image as the background, as shown in Fig. 9. Scenario C

uses the training scene but removes the marker on the surgical instruments. Figure 10 shows the laparoscopic field of view in different scenarios. We conducted 5000 experiments in each scenario. In each experiment, the position and size of surgical instruments in the laparoscopic field of view were random. Scenarios A, B, and C achieved 99.88 %, 99.78 %, and 100 % success rates respectively. Figure 11 shows the change process of the laparoscopic field of view in the three scenarios. The experimental results show that the method is less affected by the experimental environment, and when the marker is blocked, it can also achieve the expected goal well. Therefore, our model can be well adapted to the complicated surgical environment during the surgery.

It can be seen from Fig. 12a that the method can adjust the pose of the laparoscopic arm so that the surgical instruments in different postures are in the appropriate field of vision for the doctor. The successful pose adjustment results from various random movements of the surgical instrument demonstrate the performance of the learning method and its generalization capability. We also analysed some failed cases in Scenario A: when the initial surgical instrument position is set to the extreme condition of the visual field, the automatic adjustment task of the laparoscopic visual field will fail, as shown in Fig. 12b. In the initial posture of the surgical instrument shown on the left-hand side of Fig. 12b, the laparoscopic arm needs more than 40 steps to make the surgical instrument return to the correct position in the field of view. On the right-hand side of Fig. 12b, the laparoscope is moved directly to the failure state to avoid the total reward value being too small.

To simulate the actions of doctors in real surgery, we set up three different trajectories of surgical instruments: random motion, line motion, and curve motion. We conducted 500 experiments on each trajectory in each test scenario. In each experiment, if the surgical instrument is always in the correct position in the laparoscopic view, the experiment is considered successful. Table 2 shows the success rate of the task. In these scenarios, the laparoscopic arm can move the surgical instrument into a suitable field of view. It proves that

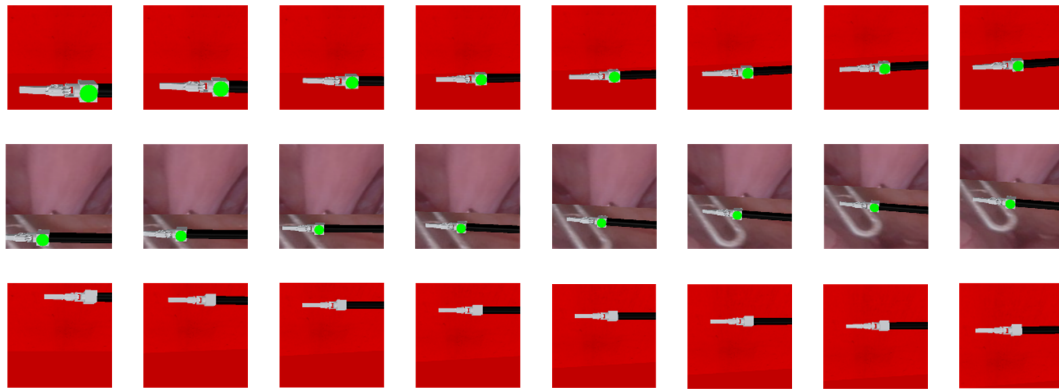


Figure 11. The adjustment process of the laparoscopic visual field.

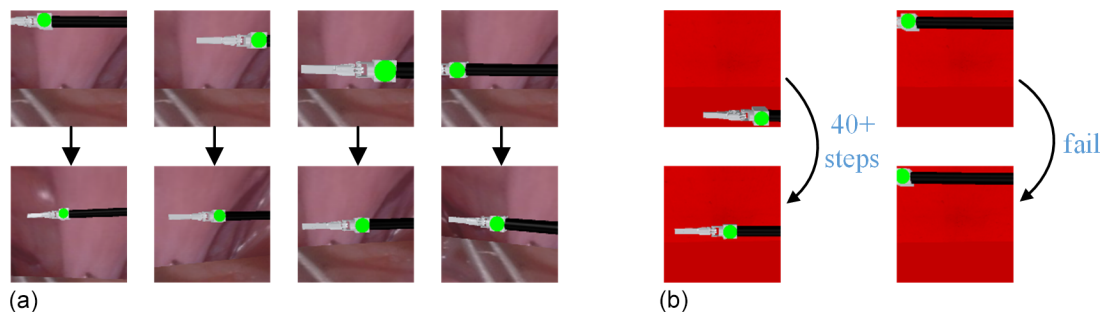


Figure 12. Visualization of test results. **(a)** Different pose testings of the surgical instrument. **(b)** Results of mission failure.

the method is widely effective and can meet the needs of doctors during surgery.

7 Conclusions

In this paper, the effectiveness of the proposed methods is demonstrated by the simulation results. This method based on machine vision and deep reinforcement learning uses convolutional neural networks to extract visual features. Then the related reward function is set up to train a policy learning network to complete the non-linear mapping of a laparoscopic image to the optimal action of a laparoscopic arm. The proposed pose adjustment system does not require camera calibration because it does not need a robot and camera mathematical model.

In addition, there are many aspects to be developed in the future. (1) This article did not experiment in the actual environment because of the expensive cost of a real experiment, so the next step is to develop a model-based data-efficient robot-tracking system to use in the actual surgical procedure. (2) The best view of laparoscopy is simple in this paper. During the operation, the surgeon must not only pay attention to surgical instruments, but also focus on the human body. So, the related reward function needs further design. (3) This article examines the issue of single surgical instrument tracking, but multiple surgical instruments may be required during

surgery. The next step is to study the issue of laparoscopic pose automatic adjustment for multiple surgical instruments.

Data availability. The data in this study can be requested from the corresponding author.

Author contributions. LY, YX, PW, and LS discussed and decided on the methodology in the study. The pre-operative planning algorithm, the reinforcement learning algorithm, and the simulations were performed by YX and PW. LY completed the literature review and the overall plan.

Competing interests. The contact author has declared that neither they nor their co-authors have any competing interests.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Acknowledgements. The authors would like to thank the National Natural Science Foundation of China for supporting this re-

search. We also greatly appreciate the efforts of the reviewers and our colleagues.

Financial support. This research has been supported by the Natural Science Foundation of Heilongjiang Province (grant no. LH2019F016).

Review statement. This paper was edited by Xichun Luo and reviewed by two anonymous referees.

References

- Ali, S. M., Reisner, L. A., King, B., Cao, A., Auner, G., Klein, M., and Pandya, A. K.: Eye Gaze Tracking for Endoscopic Camera Positioning: An Application of a Hardware/Software Interface Developed to Automate Aesop, in: *Studies in Health Technology and Informatics, 16th Conference on Medicine Meets Virtual Reality, Long Beach, United States, 30–1 January 2008, 4–7*, <https://pubmed.ncbi.nlm.nih.gov/18391246/> (last access: 8 March 2021), 2008.
- Bohez, S., Verbelen, T., De Coninck, E., Vankeirsbilck, B., Simoens, P., and Dhoedt, B.: Sensor Fusion for Robot Control through Deep Reinforcement Learning, in: *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems, Vancouver, Canada, 24–28 September 2017, 2365–2370*, <https://doi.org/10.1109/IROS.2017.8206048>, 2017.
- Cao, Y., Miura, S., Kobayashi, Y., Kawamura, K., Sugano, S., and Fujie, M. G.: Pupil Variation Applied to the Eye Tracking Control of an Endoscopic Manipulator, *IEEE Robotics and Automation Letters*, 1, 531–538, <https://doi.org/10.1109/Lra.2016.2521894>, 2016.
- Chang, J. and Rattner, D. W.: History of Minimally Invasive Surgical Oncology, *Surg. Oncol. Clin. N. Am.*, 28, 1–9, <https://doi.org/10.1016/j.soc.2018.07.001>, 2019.
- Comaniciu, D. and Meer, P.: Mean shift: A robust approach toward feature space analysis, *IEEE T. Pattern Anal.*, 24, 603–619, <https://doi.org/10.1109/34.1000236>, 2002.
- Davies, B.: Robotics in minimally invasive surgery, in: *IEE Colloquium on Through the Keyhole: Microengineering in Minimally Invasive Surgery, London, Britain, 6–6 June 1995, 5/1–5/2*, <https://doi.org/10.1049/ic:19950810>, 1995.
- Franz, A. M., Haidegger, T., Birkfellner, W., Cleary, K., Peters, T. M., and Maier-Hein, L.: Electromagnetic Tracking in Medicine-A Review of Technology, Validation, and Applications, *IEEE T. Med. Imaging*, 33, 1702–1725, <https://doi.org/10.1109/tmi.2014.2321777>, 2014.
- Gu, S., Lillicrap, T., Sutskever, I., and Levine, S.: Continuous Deep Q-Learning with Model-based Acceleration, in: *33rd International Conference on Machine Learning, New York, United States, 19–24 June 2016, 4135–4148*, <https://dl.acm.org/doi/abs/10.5555/3045390.3045688> (last access: 12 May 2021), 2016.
- Kober, J., Bagnell, J. A., and Peters, J.: Reinforcement learning in robotics: A survey, *Int. J. Robot. Res.*, 32, 1238–1274, <https://doi.org/10.1177/0278364913495721>, 2013.
- Long, J., Shelhamer, E., and Darrell, T.: Fully Convolutional Networks for Semantic Segmentation, in: *2015 IEEE Conference on Computer Vision and Pattern Recognition, Boston, United States, 7–12 June 2015, 3431–3440*, <https://doi.org/10.1109/cvpr.2015.7298965>, 2015.
- Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., and Riedmiller, M.: Playing atari with deep reinforcement learning, *arXiv [preprint]*, arXiv:1312.5602 (last access: 19 December 2020), 2013.
- Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., and Hassabis, D.: Human-level control through deep reinforcement learning, *Nature*, 518, 529–533, <https://doi.org/10.1038/nature14236>, 2015.
- Mnih, V., Badia, A. P., Mirza, M., Graves, A., Harley, T., and Lillicrap, T. P.: Asynchronous Methods for Deep Reinforcement Learning, in: *33rd International Conference on Machine Learning, New York, United States, 19–24 June 2016, 2850–2869*, <https://dl.acm.org/doi/abs/10.5555/3045390.3045594> (last access: 15 August 2020), 2016.
- Pandya, A., Reisner, L., King, B., Lucas, N., Composto, A., Klein, M., and Ellis, R. D.: A Review of Camera Viewpoint Automation in Robotic and Laparoscopic Surgery, *Robotics*, 3, 310–329, <https://doi.org/10.3390/robotics3030310>, 2014.
- Rohmer, E., Singh, S. P. N., and Freese, M.: V-REP: a Versatile and Scalable Robot Simulation Framework, in: *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems, Tokyo, Japan, 3–8 November 2013, 1321–1326*, <https://doi.org/10.1109/IROS.2013.6696520>, 2013.
- Sandoval, J., Laribi, M. A., Faure, J. P., Breque, C., Richer, J. P., and Zeghloul, S.: Towards an Autonomous Robot-Assistant for Laparoscopy Using Exteroceptive Sensors: Feasibility Study and Implementation, *IEEE Robotics and Automation Letters*, 6, 6473–6480, <https://doi.org/10.1109/Lra.2021.3094644>, 2021.
- Sekkat, H., Tigani, S., Saadane, R., and Chehri, A.: Vision-Based Robotic Arm Control Algorithm Using Deep Reinforcement Learning for Autonomous Objects Grasping, *Appl. Sci.-Basel*, 11, 7917, <https://doi.org/10.3390/app11177917>, 2021.
- Shin, S., Kim, Y., Cho, H., Lee, D., Park, S., Kim, G. J., and Kim, L.: A Single Camera Tracking System for 3D Position, Grasper Angle, and Rolling Angle of Laparoscopic Instruments, *Int. J. Precis. Eng. Man.*, 15, 2155–2160, <https://doi.org/10.1007/s12541-014-0576-6>, 2014.
- Silver, D., Lever, G., Heess, N., Degris, T., Wierstra, D., and Riedmiller, M.: Deterministic policy gradient algorithms, in: *31st International Conference on Machine Learning, Beijing, China, 21–26 June 2014, 605–619*, <https://dl.acm.org/doi/abs/10.5555/3044805.3044850> (last access: 24 December 2020), 2014.
- Yu, L. T., Wang, Z. Y., Sun, L. Q., Wang, W. J., Wang, L., and Du, Z. J.: A new forecasting kinematic algorithm of automatic navigation for a laparoscopic minimally invasive surgical robotic system, *Robotica*, 35, 1192–1222, <https://doi.org/10.1017/s0263574715001137>, 2017.
- Zhang, T. H., Kahn, G., Levine, S., and Abbeel, P.: Learning Deep Control Policies for Autonomous Aerial Vehicles with MPC-Guided Policy Search, in: *2016 IEEE International Conference on Robotics and Automation, Stockholm, Sweden, 16–21 May 2016, 528–535*, <https://doi.org/10.1109/ICRA.2016.7487175>, 2016.

- Zhao, Z. J., Voros, S., Weng, Y., Chang, F. L., and Li, R. J.: Tracking-by-detection of surgical instruments in minimally invasive surgery via the convolutional neural network deep learning-based method, *Computer Assisted Surgery*, 22, 26–35, <https://doi.org/10.1080/24699322.2017.1378777>, 2017.
- Zinchenko, K., Huang, W. S. W., Liu, K. C., and Song, K. T.: A Novel Flag-Language Remote Control Design for a Laparoscopic Camera Holder Using Image Processing, in: 15th International Conference on Control, Automation and Systems, Busan, South Korea, 13–16 October 2015, 959–963, <https://doi.org/10.1109/ICCAS.2015.7364763>, 2015.