Medical & Biological Engineering & Computing
https://doi.org/10.1007/s11517-019-02002-0

ORIGINAL ARTICLE

A CNN-based prototype method of unstructured surgical state perception and navigation for an endovascular surgery robot

Yan Zhao[1] · Shuxiang Guo[1,2] · Yuxin Wang[1] · Jinxin Cui[1] · Youchun Ma[1] · Yuwen Zeng[1] · Xinke Liu[3] · Yuhua Jiang[3] · Youxinag Li[3] · Liwei Shi[1] · Nan Xiao[1]

Received: 13 September 2018 / Accepted: 5 June 2019

Abstract

Performance of robot-assisted endovascular surgery (ES) remains highly dependent on an individual surgeon's skills, due to common adoption of master-slave robotic structure. Surgeons' skill modeling and unstructured surgical state perception pose prohibitive challenges for an autonomous ES robot. In this paper, a novel convolutional neural network (CNN)-based framework is proposed to address these challenges for navigation of an ES robot based on surgeons' skill learning. An operating action probability estimator is proposed by integrating a two-dimensional CNN, with which the features of a surgical state image are extracted and then directly mapped to the action probability. A one-dimensional CNN with multi-input is developed to recognize the guide wire operating force condition. An eye-hand collaborative servoing algorithm is proposed to combine the outputs of these two networks and to control the robot under a closed-loop architecture. A real-world ES robot is employed for data collection and task performance evaluation in laboratory condition. Compared with the state of the art, the CNN-based method shows its capability of adapting to different situations and achieves similar success rate and average operating time. Robotic operation performs similar operating trajectory and maintains similar level of operating force with manual operation. The CNN-based method can be easily extended to many other surgical robots.

Keywords Unstructured surgical state perception · Surgeons' operating skill learning · Autonomous surgical robot · Deep convolutional neural network

1 Introduction

Endovascular surgery (ES) is widely employed to treat thrombus, tumor, embolism, and so on, due to its advantages compared with traditional thoracotomy and craniotomy, such as less bleeding, fewer complications, small trauma, and quick recovery [1]. An ES robot is considered to be a promising technology to further improve the operation accuracy, efficiency, and safety [2]. The current studies on the ES robot mainly focus on the master control device with haptic feedback [3–8], the mechanism and control of the slave manipulator [9–11], the active catheter [12–15], the sensing system for surgical state acquirement [9, 16–19], and the forewarning [20–22]. Unlike an autonomous robot system for surgical tasks on static rigid tissue [23] or soft tissue [24] with rigid instruments, current ES robot systems always adopt a master-slave paradigm [25, 26], as shown in Fig. 1. As a result, execution of surgical tasks with such ES robot is entirely controlled by the surgeon and limited to an individual surgeon's skills.

✉ Shuxiang Guo
guoshuxiang@bit.edu.cn

✉ Liwei Shi
shiliwei@bit.edu.cn

✉ Nan Xiao
xiaonan@bit.edu.cn

[1] Key Laboratory of Convergence Biomedical Engineering System and Healthcare Technology, The Ministry of Industry and Information Technology, School of Life Science, Beijing Institute of Technology, No.5, Zhongguancun South Street, Haidian District, Beijing 100081, China

[2] Faculty of Engineering, Kagawa University, 2217-20 Hayashi-cho, Takamatsu, Kagawa 760-8521, Japan

[3] Department of Interventional Neuroradiology, Beijing Neurosurgical Institute and Beijing Tiantan Hospital, Capital Medical University, Beijing 100050, China

Published online: 20 June 2019

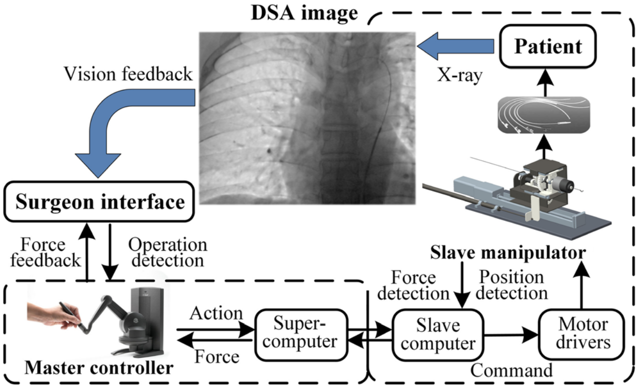


Fig. 1 Schematic diagram of the robotic-assisted ES

Surgeons acquire their skills through a long period of study in the form of explicit learning and training in the form of implicit learning [27]. During ES, the surgeon estimates the spatial relationship between the guide wire and vessel wall in a digital subtraction angiography (DSA) image. Meanwhile, the surgeon recognizes the operating force condition of the guide wire according to haptic sense. Then, the surgeon makes decision of the operating action with their surgical experience. In this paper, the action decision-making procedure is defined as the surgeons' skills. The surgical state information shown in the DSA image is dynamic and non-structured. The inference procedure from the surgical state to the operating action is complex and highly non-linear. In addition, given the large variability of patients' tissues and the difference between surgeons' operating methods, explicit modeling of the surgeons' skills is practically limited or even impossible.

Machine learning (ML) methods have been employed for surgeons' skill learning, as well as autonomous surgical robot [28], such as recurrent neural networks (RNN) [29], deep neural network (DNN) [30], Gaussian mixture model (GMM) [31], and hidden Markov model (HMM) [32–34]. However, these methods have limitations due to the fact that their inputs are the coordinate position of the gripper [29], the base position, the rotation vector, the tool position [30], and guide wire tip motion data obtained by an electromagnetic position sensor [31–33]. These methods do not have the capability of understanding the unstructured surgical state image. A method combining non-rigid registration and GMM was proposed by Yang et al. [35], which is the state of the art for an autonomous ES robot. Despite the fact that it has the capability to adjust new vascular anatomy, only the vessel centerline is mapped to the actions but not the vessel contour morphology. Moreover, this method is limited by generating a fixed action sequence from a fixed starting point.

A deep convolutional neural network (CNN) provides a promising way to deal with the challenges of unstructured surgical state perception and surgeons' skill learning. Deep CNN has demonstrated its outstanding capability of understanding a raw pixel image in image classification [36], diagnosing a tumor, playing games, even robot grasping, and so on. It successfully overcame the challenge of an enormous search space and the difficulty in evaluating board positions and moves in the game of Go [37]. A deep CNN-based model successfully learned the control policies for seven Atari 2600 games from image input and surpassed a human expert on three of them [38]. A novel method was proposed by employing three-dimensional CNNs for automated pulmonary nodule detection from volumetric computed tomography (CT) data [39]. Further, deep CNNs have been successfully used to map the visual input to control commands for robot grasping [40–42]. A deep Q-Network (DQN)-based system was proposed with the capability of autonomously learning robot controllers from image input without any prior knowledge of configuration [43]. However, to our knowledge, there is no reported study on an autonomous ES robot directly with image input based on deep CNN methods.

In this paper, a CNN-based framework is proposed for unstructured surgical state perception and surgeons' skill learning for an ES robot. A 2-D deep CNN is developed for unstructured surgical state perception with image input, which is defined as the action probability estimator. A multi-input one-dimensional (1-D) deep CNN is proposed to recognize the operating force mode of the guide wire, which is defined as the operating force mode recognizer. Then, an eye-hand collaborative servoing algorithm is proposed to combine those two networks for action execution under closed-loop control. Demonstrations of two kinds of representative surgical tasks, including vessel branch passing and obstacle passing, are collected under laboratory settings for network training.

2 Methodology

2.1 Overview of the CNN-based framework

During the clinical ES procedure, the surgeon applies pushing, pulling, and rotating actions on the guide wire end outside the sheath to deliver the guide wire tip to the target area along the vessel lumen. They make the action decision by estimating the surgical state that is presented mainly by two kinds of information: the visual information in the DSA image and operating force on their hands. They estimate the spatial relationship between the guide wire and vessel contours according to the visual information, while ensuring the surgery safety by adjusting the operating force. The spatial information between the guide wire and vessel contours is unstructured and dynamic because of their flexible and deformable physical property. Although non-rigid registration and GMM methods were proposed [35], only the vessel centerline is mapped to the actions. The important information including vessel contour morphology and operating force was not taken into account. For operating force mode recognition, a threshold method was adopted in Jian et al.'s research [21]. But, during the procedure that the guide wire is continuously inserted into the vessel lumen, the interaction force between the guide wire body and vessel wall is complicated and changeable. So, it is hard to classify the operating force mode by simply setting a safety threshold of the operating force that is always detected at the guide wire end.

A CNN-based framework is proposed in this paper as shown in Fig. 2. It mainly consists of two phases: phase I is data collection and phase II is automatic operation. In phase I, the master-slave robot is used for data collection, which is used to train the network for automatic operation. A human surgeon controls the slave robot through the master robot to manipulate the guide wire for specific surgical tasks, while the actions detected by the master robot and operating force detected by the slave robot, as well as images detected by the camera under laboratory conditions or by the DSA device under clinical conditions, are recorded as demonstration data. In phase II for automatic operation, the slave robot is controlled by the trained models (i.e., the action probability estimator and operating force mode recognizer). In each close control loop of automatic operation, the current surgical state image is taken as the input of the trained action probability estimator and the guide wire operating force is taken as the input of the trained operating force mode recognizer. The eye-hand collaborative mechanism infers the optimal action according to the output of these two learning models. Then, the inferred optimal action is executed by the slave robot to operate the guide wire for surgical tasks.

Four preconditions are set for this work. (1) The vessel contour is visible in the medical image. In clinical condition, the vessel contour can be obtained from the DSA image. (2) Camera images are used as a substitute for the DSA image in experimental condition for training data collection and evaluation experiments. The vessel contour and catheter contour are extracted using a Canny operator in the preprocess procedure. For the DSA image in clinical condition, the textures could be extracted with specific extracting and tracking methods for a medical image [44–49] in the preprocess procedure. (3) The target point of the catheter tip in certain tasks is set by human. (4) The guide wire operating force detected by the force sensor inside the slave manipulator is used to represent the surgeons' operating force on the guide wire.

2.2 Action probability estimator

An action probability estimator is proposed to predict the probability distribution of the candidate actions among given action spaces based on the current surgical state image. In order to make accurate predictions, the action probability estimator should have the capability not only to extract the features of spatial relationship between the guide wire and vessel contour but also to deduce the extracted features to the action probability. The action probability estimator is developed based on CNN, as shown in Fig. 3. The input of the network is the preprocessed current surgical state image, and the output is the estimated action probability distribution.

Firstly, the contour of the moving guide wire and the vessel wall in the surgical state image is extracted using a Canny operator. The surgical tasks could be represented by a marking target area and the vessel centerline from the starting area to the target area in the image.

Then, an 8-layer CNN is used to extract the features in the preprocessed images as shown in Fig. 3. It was proven that using more kernels with smaller size could achieve better performance than using less kernels with larger size [48]. At each convolutional layer, the kernels are used to sweep over the input (i.e., the output of the previous layer) step by step to extract a stack of higher-level features. A kernel with trained parameters detects a certain kind of meaningful feature

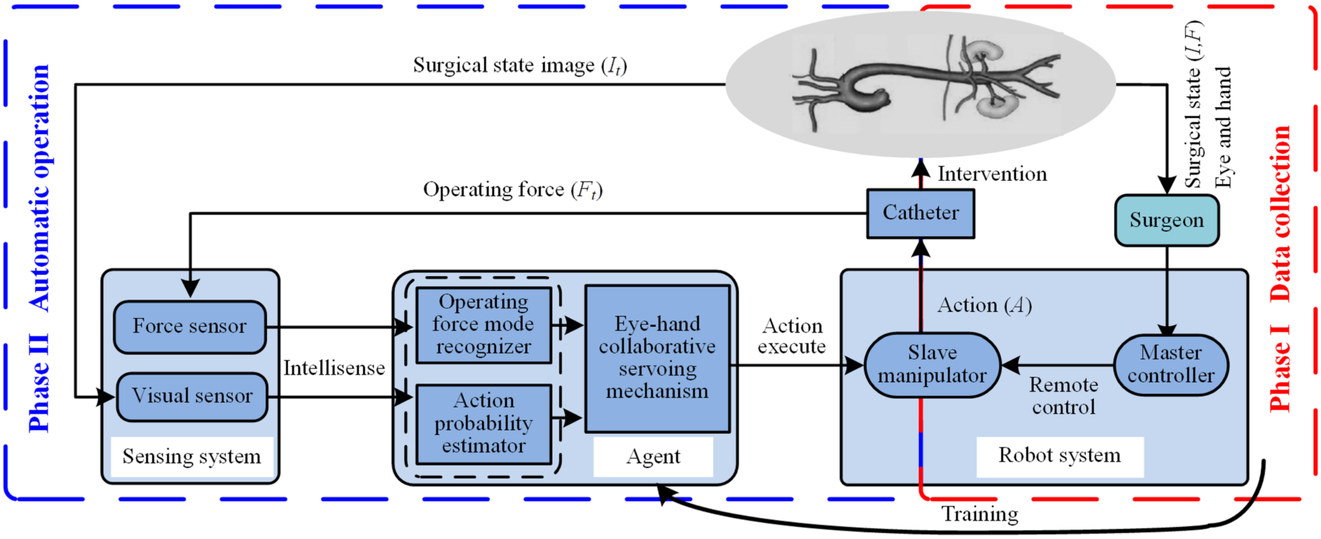


Fig. 2 Diagram of the proposed CNN-based framework

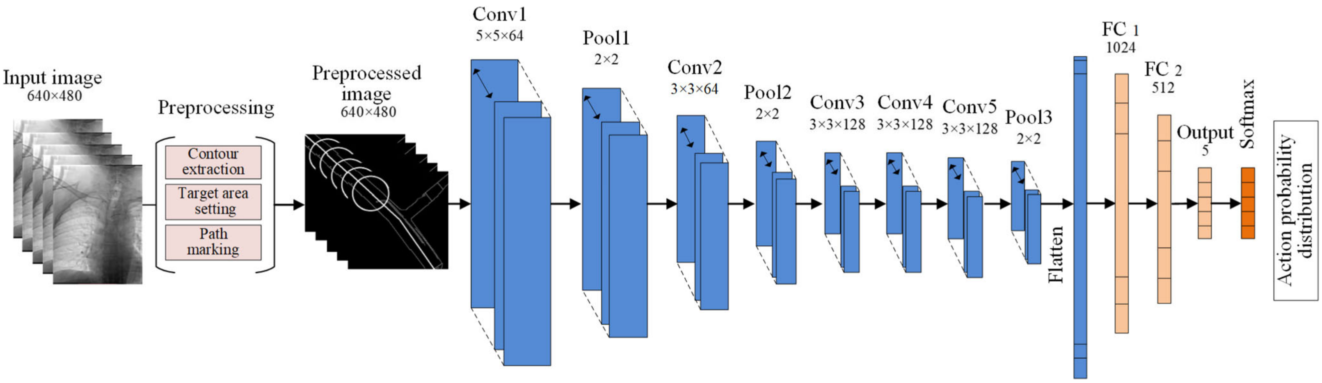


Fig. 3 Architecture of the action probability estimator

contained in the local information. Then, a bias term is added and a non-linear activation function is employed. The 2-D convolutional layer can be formulated as (1):

$$h_i^l(x,y) = \sigma\left(b_i^l + \sum_k \sum_{u,v} h_k^{l-1}(x-u, y-v) W_{ki}^l(u,v)\right) \quad (1)$$

where h_i^l and h_k^{l-1} are respectively the *i*th 2-D feature map in the *l*th layer and the *k*th 2-D feature map in the previous layer, $W_{ki}^l \in R^2$ is the *k*th 2-D convolutional kernel mapping h_k^{l-1} toh_i^l; $h_i^l(x,y)$, $h_k^{l-1}(x-u, y-v)$, and $W_{ki}^l(u,v)$ represent the element values with (x,y) being the coordinates of h_i^l and (u,v) being the coordinates of W_{ki}^l, b_i^l is the bias term, $\sigma(\cdot)$ is the non-linear activation function, and the rectified linear units (ReLU) ($\sigma(a) = \max(0,a)$) is used in this paper.

After the first, second, and fifth convolutional layers, max pooling layers are respectively adopted to further reduce the feature dimension. Max pooling is a form of non-linear downsampling operation. It is beneficial for avoiding overfitting during training the network by reducing the scale of parameters. Max pooling is performed over the whole feature map, so that the max pooled feature represents the global feature [30]. Surgeons' skills are mainly represented by the change procedure of the spatial relationship between the guide wire and vessel wall contained in the texture feature of the medical image sequence. So, the features reflecting surgeons' skills remained after every max pooling layer.

After the third max pooling layer, the feature maps are flattened and then connected with two fully connected layers. The fully connected layers are used to fit the non-linear relationship between the low-dimensional feature maps and the action probability distribution. The fully connected layer can be expressed as (2):

$$h^f = \sigma\left(b^f + W^f h^{f-1}\right) \quad (2)$$

where h^{f-1} is the input feature vector of the *f*th fully connected layer, h^f is the output feature vector, W^f is the weight matrix, b^f is the bias term, and $\sigma(\cdot)$ is the activation function ReLU.

By denoting the neuron vector in the output layer by h^0, the number of the neurons of h^0 equals to the number of actions in the given candidate action space $A_t[a_0, a_1, a_2, a_3, a_4]$ (i.e., pushing forward, pushing forward synchronously with rotating, rotating, pulling back synchronously with rotating, and pulling back; the displacements of the pushing and pulling action are set as 0.6 mm, and the angular displacement of the rotating action is set as 6°). Then, a Softmax layer is used to calculate the action probability of each action by the Softmax regression $p_a(h^o) = \exp\left(h_a^o\right)/\sum_{a=0}^{A-1}\exp\left(h_a^o\right)$, where h_a^o is the *a*th output value.

2.3 Operating force mode recognizer

Recognition of the operating force condition can be defined as a two-class classification problem. If the guide wire tip encounters an obstacle, the operating force condition is defined as an abnormal force mode. Otherwise, the operating force condition is defined as a normal force mode. As shown in Fig. 4, a multi-input 1-D CNN is proposed as the operating force mode recognizer. It takes both the current operating force values and several last operating force values as the multiple input. In this way, the variation tendency of the operating force is also taken into account to recognize the operating force mode.

The operating force value at current time is defined as F_{ti}. The operating force sequence from $F_{ti}$to M last operating force values is defined as F_t (M is set as 50 in this work). The sampling frequency of the force sensor is 90 Hz. F_{ti} is taken as the input of the first convolutional layer, which is followed by the second convolutional layer, two fully connected layers, and a Softmax layer as the output layer. The 1-D convolutional layer can be expressed as (3).

$$\eta_i^l(x) = \sigma\left(b_i^l + \sum_k \sum_u \eta_k^{l-1}(x) W_{ki}^l(u)\right) \quad (3)$$

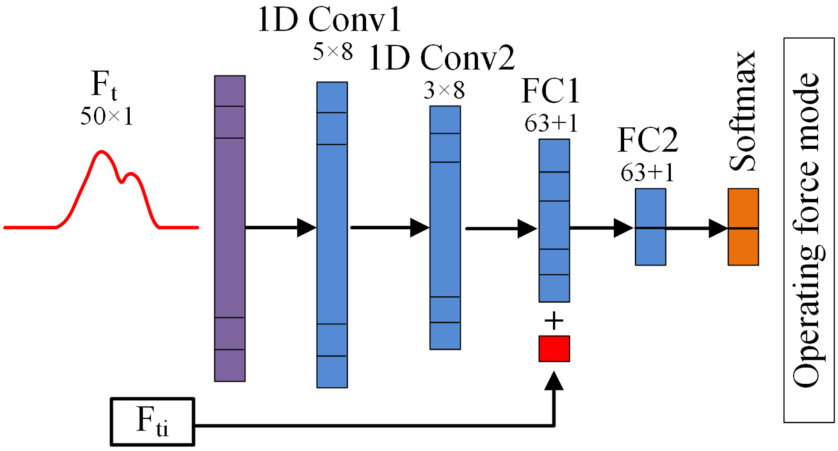


Fig. 4 Architecture of the operating force mode recognizer

where η_i^l is the ith output feature vector of the lth 1-D convolutional layer, η_k^{l-1} is the kth input feature vector of the lth 1-D convolutional layer, $W_{ki}^l \in R$ is the kth kernel of the lth 1-D convolutional layers, b_i^l is the bias term, and $\sigma(\cdot)$ is the activation function ReLU.

To avoid data submergence of the current operating force value F_i during convolution, F_i is directly taken as a part of the input feature of the first fully connected layer. The first and second fully connected layers can be expressed respectively as (4) and (5):

$$\eta^{f_1} = \sigma\left(b^{f_1} + W^{f_1}(F_t, F_{ti})\right) \tag{4}$$

$$\eta^{f_2} = \sigma\left(b^{f_2} + W^{f_2}\eta^{f_1}\right) \tag{5}$$

where (F_t, F_{ti}) and η^{f_1} are respectively the input feature vector and output feature vector of the first fully connected layer, η^{f_1} and η^{f_2} are respectively the input feature vector and output feature vector of the second fully connected layer, W^{f_1} and W^{f_2} are the weight matrixes of the two fully connected layers, b^{f_1} and b^{f2} are the bias terms, and $\sigma(\cdot)$ is the activation function ReLU.

2.4 Eye-hand collaborative servoing algorithm

In this section, an eye-hand collaborative servoing algorithm is proposed to realize autonomous control of the slave robot under closed-loop architecture. It combines the output of the action probability estimator and operating force mode recognizer. If the recognition result of the operating force mode recognizer is a normal force mode, the action with the maximum probability will be executed. If the recognition result of the operating force mode recognizer is an abnormal force mode, the servoing algorithm will control the slave manipulator to avoid the obstacle by the given avoiding actions (i.e., pulling back the guide wire by a certain distance and then rotating it by a certain angle). The eye-hand collaborative servoing algorithm is illustrated below:

2.5 Hardware setup and data collection

In order to collect data for network training, an ES robot [26] is used to develop the hardware setup, as shown in Fig. 5a. It mainly consists of a Geomagic Touch X device as the master controller, a slave manipulator, a vessel model, and a gray-scale monocular camera. During data collection, surgeons control the slave manipulator via the master controller to operate the guide wire inside the vessel model as shown in Fig. 5b. The master controller records the operating actions. The force sensor inside the slave manipulator detects the operating force, as shown in Fig. 5c. Meanwhile, the camera acquires the surgical state image. Two kinds of vessel models are used for data collection and evaluating experiments, as shown in Fig. 6. One branch in the medical model and three designed vessel models are used for data collection, while another branch in the medical model and the fourth designed vessel model are used for evaluating experiments. The proposed CNN-based method is developed on the basis of TensorFlow and a graphic workstation (with a GPU of NVIDIA Quadro K4000).

Three kinds of experiments in different cases are designed. The starting area and target area are given by an operator.

Case I The medical vessel model is used. Case I(a): the guide wire tip is moved from the starting area and is orientated towards the objective branch. The guide wire tip can be pushed directly into the objective branch. Case I(b): the guide wire tip is moved from the starting area and is orientated against the objective branch. The guide wire tip needs to be rotated firstly and then pushed into the objective branch. One

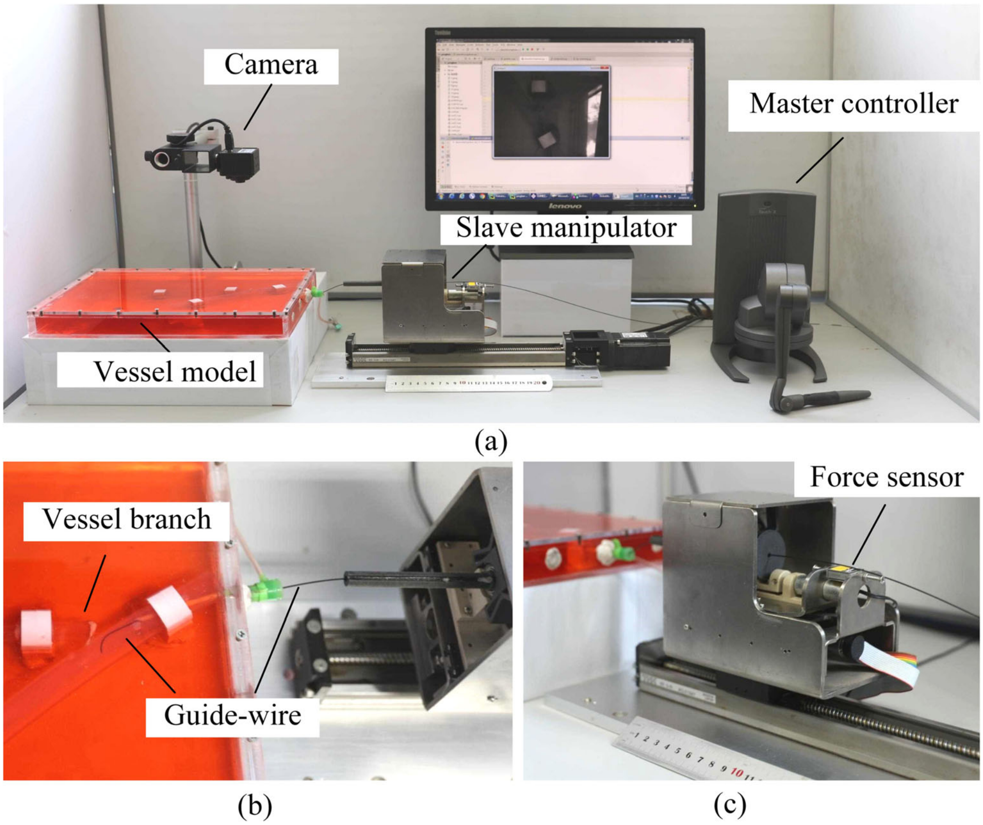


Fig. 5 Experimental setup. **a** Overview of the ES robot system and the vessel model. **b** The guide wire inserted into the vessel by the slave manipulator. **c** The force sensor inside the slave manipulator

hundred fifty demonstrations are conducted by 5 surgeons, and about 5500 samples are collected for each case.

Case II The designed vessel models are used. Case II(a): the guide wire tip is moved from the starting area and is orientated towards the objective branch. The guide wire tip can be pushed directly into the objective branch. Case II(b): the guide wire tip is moved from the starting area and is orientated against the objective branch. The guide wire tip needs to be rotated firstly and then pushed into the objective branch. Case

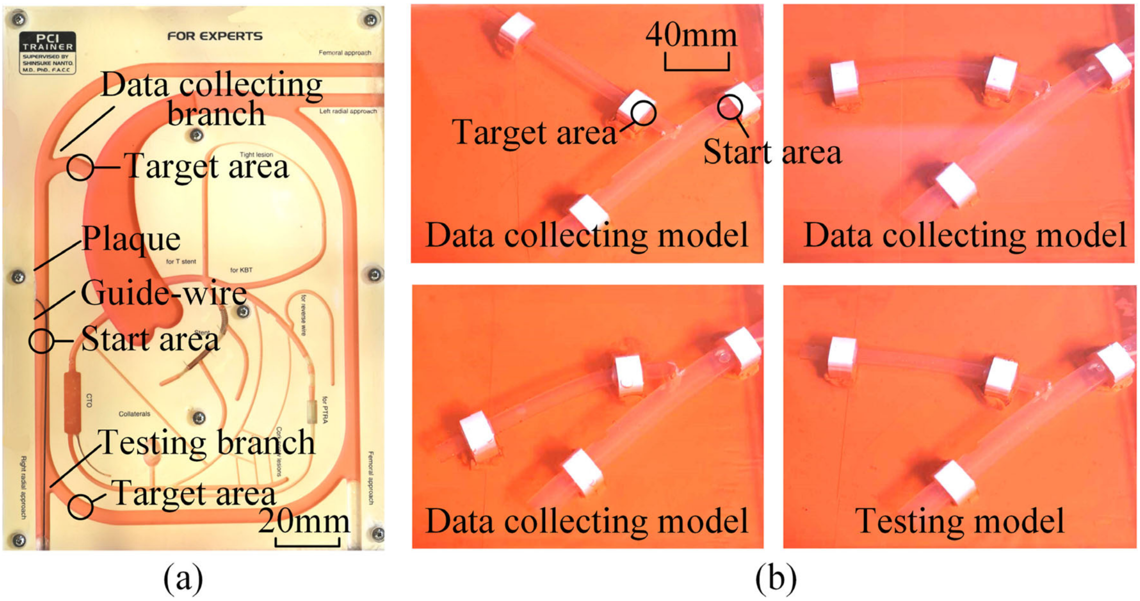


Fig. 6 Vessel models. **a** One branch in the medical vessel model for data collection and another for test. **b** Three designed vessel models for data collection and one for test

II(c): the guide wire tip has been moved into the wrong branch. The guide wire tip needs firstly to be pulled back to the entrance of the branch, then be rotated and pushed into the objective branch. Two hundred demonstrations are conducted by 5 surgeons, and about 11,000 samples are collected for each case.

Case III The medical vessel model is used. A simulated plaque is pasted on the vessel wall. The guide wire is pulled from the starting area and had an encounter with the plaque. Case III(a): the guide wire tip is not blocked by the obstacle, and the guide wire needs to be pushed forward directly and passes the obstacle. Case III(b): the guide wire tip is blocked by the obstacle, and the guide wire will impress the obstacle and bend gradually. Fifty times of demonstrations are conducted by 5 surgeons for each case.

2.6 Training the network

To avoid overfitting, the collected data are augmented by 15 times, adopting augmenting methods of flipping, rotating, and adding Gaussian noise. Finally, about 704,000 of samples are obtained. The lost function used to train the network can be obtained based on a cross-entropy method. By defining all the trainable parameters by θ, the lost function can be expressed as (6):

$$L(\theta) = -\frac{1}{N}\sum_{j=1}^{N}\sum_{a=0}^{A-1} 1\left\{y_A^{(j)} = c\right\}\ln P\left(\hat{y}_A^{(j)} = c|I^{(j)};\theta\right) \tag{6}$$

where $I^{(j)}$ is the input image batch, $y_A^{(j)}$ is the corresponding label sets, $\hat{y}_A^{(j)}$ is the predicted action probability, and $P(\cdot)$ is the probability of classifying $I^{(j)}$ as c class, which can be expressed as (7):

$$P\left(\hat{y}_A^{(j)} = c|I^{(j)};\theta\right) = \frac{e^{\theta_c^T I^{(j)}}}{\sum_{l=1}^{k} e^{\theta_l^T I^{(j)}}} \tag{7}$$

As for training the operating force mode recognizer, the loss function is also based on a cross-entropy method.

An Adam (adaptive moment estimation) optimizer is used to train the networks. The batch size is 32. The learning rate is initially set as 0.005, and it is reduced with a decay of 0.909 at each training step until it reaches 0.0001.

3 Evaluation experiments and results

In this section, the goal of the evaluation experiments is to answer the following questions: (1) how well does the proposed CNN-based method compared with the state of the art? (2) How well does the proposed eye-hand collaborative operation perform?

3.1 Comparison with the state of the art

Vessel branch passing tests are conducted for comparison between the proposed CNN-based method and Yang's non-rigid registration and GMM method [35]. Because the operating force is not taken into account into Yang's method, the evaluation tests do not consist of an obstacle passing task. For detailed comparison, evaluation tests in different cases with different training data are conducted, as shown in Table 1. For convenience, the data collected in each case are defined according to themselves. For instance, the data collected in case I is defined as case I data. In addition, the data randomly mixing the case I data and case II data is defined as multi-data. For quantitative comparison, evaluation metrics are defined as follows:

(1) Success rate (SR): for a given surgical task, if the guide wire tip achieves nearby the target area in a limit of 100 s, the test is counted as success; otherwise, it is counted as failure. Fifty times of tests are conducted for each case.

(2) Average operating time (AOT): the efficiency of task execution was measured by average operating time. Larger operating time indicates large X-ray radiation dose, which leads to more damage to the surgeon and patient.

Although Yang's method achieves relative higher SR and shorter AOT, our method shows its capability of adapting to different situations. As shown in Table 1, for the tests on the medical model in case I(a) and case I(b), Yang's method trained respectively with case I(a) data and case I(b) data achieves SR of 100% and 94%, respectively, which are relatively higher than those with our method. In addition, the AOT with Yang's method are shorter than those with our method, with the largest AOT difference of 28.5% in case I(a). However, Yang's method trained with case I(a) data cannot complete the task in case I(b). In fact, the guide wire tip directly moves into the wrong branch in all the tests in case I(a) by Yang's method trained with case I(b) data. Similar results are also observed from the tests in case II with Yang's method trained with non-corresponding data. The reason is that the GMM in Yang's method could only be trained with the data from a relatively fixed starting point and generates a fixed action command sequence. In this way, despite changes of the starting point, the robot is still controlled by the generated command sequence. As a result, the guide wire operated by the robot cannot reach the target area. In contrast, our method trained with case I data can perform the task in both case I(a) and case I(b). It shows the same performance in the tests in case II. This is because our method could learn the non-linear features in mixed data. Also, based on the closed-loop control architecture, our method estimates the current surgical state and then generates the optimal action command at

Table 1 Comparison between our method and Yang's method

Test model	Method	Training data	Case I(a)		Case I(b)			
			SR	AOT (s)	SR	AOT (s)		
Medical vessel model	Yang's method [35]	Case I(a) data	100%	21.4	0	–		
		Case I(b) data	0	–	94%	34.8		
	Our method	Case I data	84%	32.3	80%	46.4		
		Multi-data	94%	25.2	92%	39.7		
			Case II(a)		Case II(b)		Case II(c)	
			SR	AOT (s)	SR	AOT (s)	SR	AOT (s)
Designed vessel model	Yang's method [35]	Case II(a) data	98%	31.7	0	–	–	
	Our method	Case II (b) data	0	–	94%	39.2		
		Case II data	96%	43.3	94%	54.8	86%	55.7
		Multi-data	92%	48.9	90	57.2	84%	58.6

each time step, rather than preplanning and executing a fixed action command sequence.

AOT of our method trained by the multi-data are reduced respectively by 22.0% and 25.2% in case I(a) and case I(b), compared with those trained by case I data. The reason for the time reduction is the decrease in the wrong actions. Despite data augmentation, the original vessel shape is constant in the case I data. But there are more vessel shapes in the case II data. So, the sample diversity is increased by combining case I data with case II data. Then, the vessel shape tends more to be considered a kind of feature by the network. It is beneficial to improve the recognizing accuracy of the network. AOT with case II data and multi-data are at the same level in case II tests. The reason might be that only one kind of training data with wide difference cannot obviously improve the recognizing capability of the network. In some particular situation, reciprocating motion of the guide wire is observed, which prolongs the task completing time. In this situation, the spatial relationship between the guide wire and vessel wall is different with most of the training samples; as a result, the recognizing accuracy of the action probability estimator is low. The estimator might output wrong action in this situation. When the number of wrong actions is large enough, it could lead to a failure test due to moving into a wrong vessel branch or exceeding the time limitation. It also indicates the importance of sample diversity for recognizing capability of the action probability estimator.

In 61 of total 250 tests of our method trained with multi-data in case I and case II, the guide wire tip moves into the wrong vessel branch. And in 53 of these 61 tests, the guide wire tip is successfully pulled back, rotated to the suitable direction, and then pushed towards the target area. It indicates the capability of recovering from failure situation of our method (Fig. 7), which is important for the application in a dynamic surgical state. In contrast, Yang's method does not show this capability. Further, it should be pointed out that the demonstrations of recovery from a wrong vessel branch are not consisted in the case I data, but consisted in the case II(c) data. It indicates that the network trained with multi-data learns the skills of recovering from failure situation and shows the skills in the tests in case I. It also indicates that the network learns successfully to recognize the spatial relationship between the guide wire and vessel contours.

3.2 Performance evaluation of eye-hand collaborative operation

For evaluating the eye-hand collaborative servoing algorithm, evaluation tests of the task combining obstacle passing and vessel branch passing are conducted in the testing vessel branch of the medical vessel model. The guide wire could pass the obstacle only if with suitable orientation. In most of the evaluation tests, both of the robot and surgeon would try several times of attempts before the guide wire tip passes the obstacle. If the tasks are completed within 100 s, the test is counted as success; otherwise, it is counted as failure. Two hundred times of tests are conducted and finally achieve a

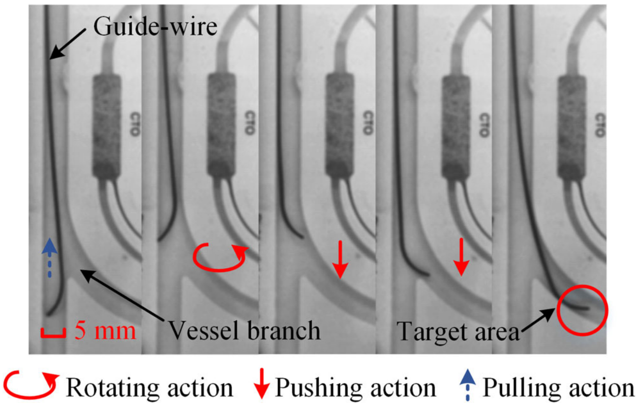


Fig. 7 A test demonstrates the capability of recovering from failure situation

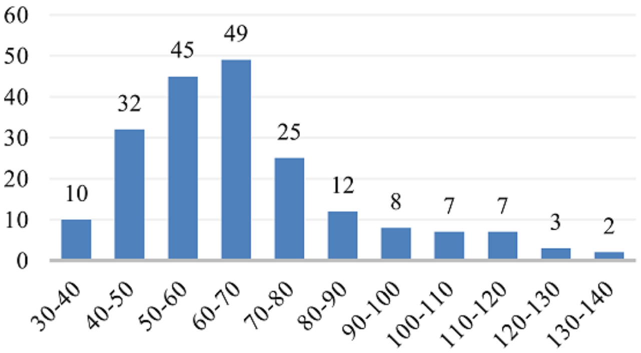


Fig. 8 Distribution of operating times

success rate of 90.5%. It can be seen from Fig. 8 that the operating times of the evaluation tests distribute mainly from 40 to 80 s. The AOT of eye-hand collaborative operation is 55.4 s, which is close to the surgeon's tests of 44.1 s. Although the maximum operating time is 139.2 s, which is larger than the surgeon's maximum operating time of 51.7 s.

Figure 9a, k shows respectively the guide wire tip trajectories of a representative robotic test and a manual test. It can be seen from the trajectories that the robot controlled by the eye-hand collaborative servoing algorithm performs the tasks through a similar procedure compared with the manual test. Owing to obstacle passing attempts, both of these two trajectories are relatively tanglesome near the simulated plaque. After passing the simulated plaque, the trajectory in the robotic test is as smooth as that in the manual test.

The detailed moving procedures of the guide wire in these two tests are respectively shown in panels b–j and l–t of Fig. 9. Three obstacle passing attempts are observed in these two tests. In all the robotic tests and manual tests, the times of obstacle passing attempts range from 0 to 6. Points B, D, F, L, N, and P in Fig. 10 are respectively corresponding to Fig. 9b, d, j, l, n, t. It can be seen that the operating force mode recognizer correctly recognizes the abnormal force mode at points B, D, and F, and then the obstacle avoiding actions are correctly executed. After that, the action probability estimator correctly understands the surgical state and rotating actions are executed near the junction to adjust the guide wire tip towards the objective branch.

Further, the operating forces at points B, D, F, L, N, and P are not of the same value as shown in Fig. 10. It indicates that the operating force mode recognizer recognizes the force mode not simply according to a threshold force value but according to both the variation tendency and instantaneous value of the operating force. It is similar to the surgeon's manner of operating force mode recognition. In addition, the maximum operating forces during obstacle avoiding attempts in all the evaluation tests range from 0.24 to 0.79 N, which are at the same level as those in manual tests (ranging from 0.29 to 0.83 N).

The results of operating force comparison indicate that the operating force mode recognizer learns the skills of the surgeon and shows a similar performance to manual tests. It is beneficial to assure the surgery safety. The overall results demonstrate that the task combining obstacle passing and vessel branch passing could be autonomously performed by the robot with the eye-hand collaborative servoing algorithm.

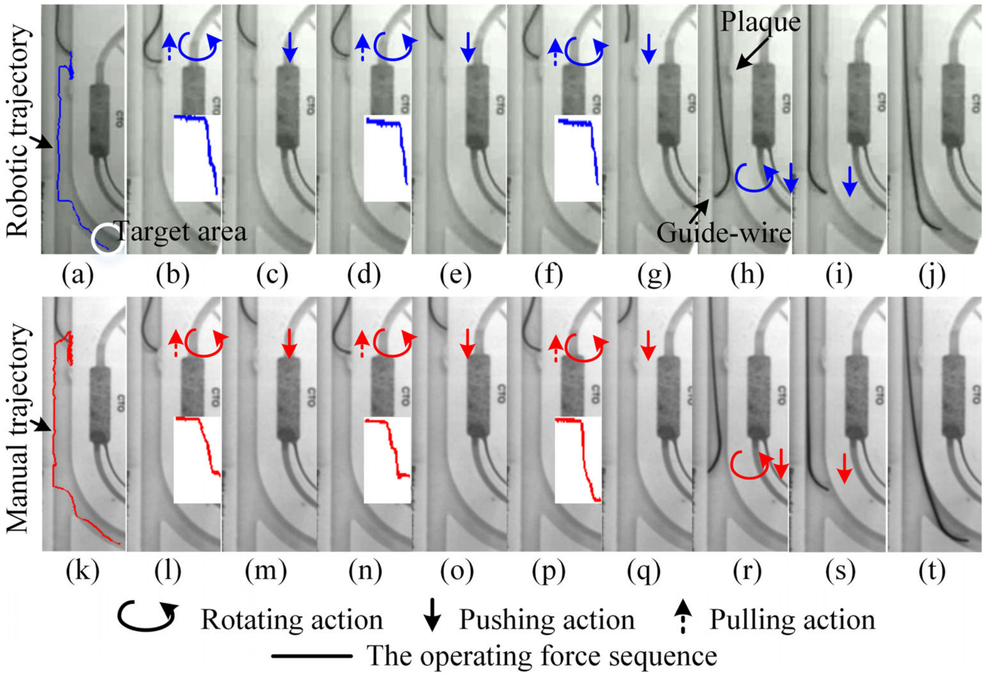


Fig. 9 Comparison of guide wire tip trajectory and detailed operating procedure. **a**–**j** Robotic test. **k**–**t** Manual test

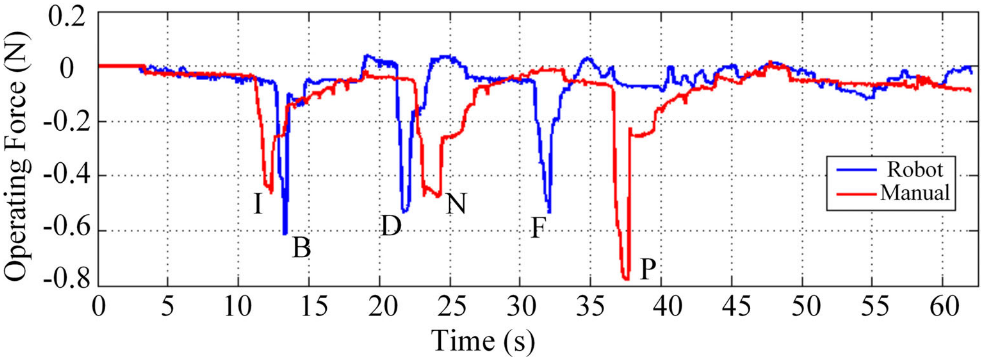


Fig. 10 Comparison of operating force. Points B, D, F, L, N, and P in this figure are respectively corresponding to panels b–j and l–t of Fig. 9

4 Discussion

Despite the unstructured surgical state and the surgeon's implicit skills, surgical tasks in ES requiring human cognition and experience can be executed autonomously without manually preprogramming for different situations. We demonstrate the feasibility and potential of the proposed CNN-based method to unstructured surgical state perception and the surgeon's skill learning in ES.

The experimental results show the CNN-based method's capability of adjusting to different situations. With the high non-linear mapping capacity, the CNN could be trained with the sample data collected at different conditions, compared with non-rigid registration and GMM methods [35]. As a result, it has the capability of adapting to different situations. Owing to its closed-loop control architecture and ability to perceive an unstructured surgical state, our method shows the capability of recovering from failure situation. Recovery from failure situation is a critical issue for application of the deep learning method in real-world robotics [50]. Our method achieves the capability of recovering from failure situation by combining samples of case II(c) with the training data, which consists of the samples of recovery from a wrong branch. It is interesting that the network trained with multi-data learns the skills of recovering from failure situation and shows the skills in the tests of case I that does not consist of samples of recovery from a wrong branch. It demonstrates the ability of the CNN-based method that it could transfer the learned skills to different situations. It is inspirational for the issue of experience sharing among multiple ES robots trained by the demonstration data collected by different surgeons with different skills.

Under a closed-loop control architecture, the eye-hand collaborative servoing algorithm effectively integrates the capabilities of the action probability estimator and operating force mode recognizer. The action probability estimator helps to choose the optimal action based on unstructured surgical state perception. The operating force mode recognizer maintains a similar operating force level to manual operation, which is beneficial to surgical safety assurance. These capabilities are obtained through learning from demonstrations rather than manually preprogramming. Further, the inputs to the network are an image and force signal, which are not related to the kinematics and dynamics of the surgical robot. It means that calibration of the camera and robot coordinates is not needed, which is necessary for many current methods [24].

The main reason for the failure cases is that diversity of the samples is not high enough. Because there are always some patterns for human demonstrations, some particular situations during the robotic test are not similar to those in the samples. In these situations, wrong actions might be chosen and executed. SR could be improved by enhancing generalization of the network in the future work, using more training data from various high-fidelity surgery scenarios.

The intent of the proposed CNN-based framework is not to replace human surgeons but to enhance the autonomy of the ES robot, since improving the autonomy of the surgical robot could expand human capacity and capability in human-robot collaborative surgery [51]. In human-robot collaborative surgery, the slave robot would be controlled by both the human surgeon (through a master robot) and trained network within specific collaborative architecture. For instance [52], repetitive and low-risk surgical tasks could be performed autonomously by a robot under a human surgeon's supervision, while the surgeon focuses on high-risk tasks. In this way, the burden on the surgeon could be reduced, and more attention of the surgeon could be turned to complex hand high-risk tasks.

5 Conclusion

In this paper, we report a CNN-based method of unstructured surgical state perception and the surgeon's skill learning for an ES robot. The results demonstrate the feasibility and potential of the CNN-based method for enhancing the autonomy of the ES robot. In addition, because the kinematic and dynamic models of the robot are unnecessary, the CNN-based method has the potential to be easily extended to many other kinds of surgical robots.

Future studies will include the following: enhancement of the network generalization through collecting more various demonstration data, 3-D perception of the surgical state by taking images from different views as input to the network, and a specific architecture for human-robot collaborative operation.

Funding This research is partly supported by the National High-tech R&D Program (863 Program) of China (No. 2015AA043202) and National Key Research and Development Program of China (2017YFB1304401).

Compliance with ethical standards

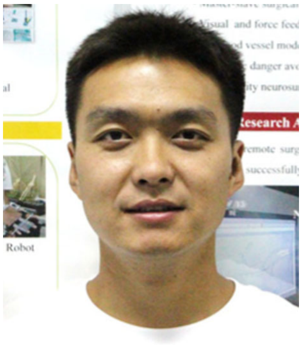
Conflict of interest The authors declare that they have no conflict of interest.

References

1. Jenny DJohn JVDD, Paul B (2010) Current technology on minimally invasive surgery and interventional techniques. IEEE International Conference on Instrumentation Control and Automation. https://doi.org/10.1109/ICA.2011.6130118
2. Hedyeh RT, Christopher JP, Yang GZ (2013) Current and emerging robot-assisted endovascular catheterization technologies: a review. Ann Biomed Eng 42:697–715. https://doi.org/10.1007/s10439-013-0946-8
3. Yu S, Shuxiang G, Xuanchun Y, Linshuai Z, Yu W, Hirata H, Ishihara H (2017) Design and performance evaluation of a haptic interface based on MR fluids for endovascular tele-surgery. Microsyst Technol 24(10):909–918. https://doi.org/10.1007/s00542-017-3404-y
4. Yu S, Shuxiang G, Xuanchun Y, Linshuai Z, Yu W, Hirata H, Ishihara H, Tamiya T (2018) Performance evaluation of a robot-assisted catheter operating system with haptic feedback. Biomed Microdevices 20:50. https://doi.org/10.1007/s10544-018-0294-4
5. Xuanchun Y, Shuxinag G, Nan X, Tamiya T, Hirata H, Ishihara H (2016) Safety operation consciousness realization of a MR fluids-based novel haptic interface for teleoperated catheter minimally invasive neuro surgery. IEEE/ASME Trans Mechatronics 21:1043–1054. https://doi.org/10.1109/TMECH.2015.2489219
6. Zhou C, Xie L, Shen X, Luo M, Wu Z, Gu L (2015) Cardiovascular interventional surgery virtual training platform and its preliminary evaluation. Int J Med Robotics Comput Assist Surg 11:375–387. https://doi.org/10.1002/rcs.1627
7. Xianqiang B, Shuxiang G, Nan X, Youxiang L, Cheng Y, Yuhua J (2018) A cooperation of catheters and guide-wires-based novel remote-controlled endovascular interventional robot. Biomed Microdevices 20:20. https://doi.org/10.1007/s10544-018-0261-0
8. Xianqiang B, Shuxiang G, Nan X, Li Y, Shi L (2018) Compensatory force measurement and multimodal force feedback for remote-controlled vascular interventional robot. Biomed Microdevices 20:74. https://doi.org/10.1007/s10544-018-0318-0
9. Jian G, Shuxiang G, Lin S, Peng W (2016) Design and performance evaluation of a novel robotic catheter system for endovascular interventional surgery. Microsyst Technol 22:2167–2176. https://doi.org/10.1007/s00542-015-2659-4
10. Yang X, Wang H, Sun L, Hongnian Y (2015) Operation and force analysis of the guide wire in a minimally invasive. Chin J Mech Eng 28:249–257. https://doi.org/10.3901/cjme.2014.1229.181
11. Wang K, Chen B, Xu X (2016) Design and control method of surgical robot for endovascular intervention operation. ROBIO. https://doi.org/10.1109/ROBIO.2016.7866331
12. Chaonan Z, Shuxiang G, Nan X, Jiaqing W, Youxiang L, Yuhua J (2018) Transverse microvibration-based guide wire drag reduction evaluation for vascular interventional application. Biomed Microdevices 20:69. https://doi.org/10.1007/s10544-018-0315-3
13. Yi L, Shuxiang G, Hirata H, Ishihara H, Tamiya T (2018) Development of a powered variable-stiffness exoskeleton device for elbow rehabilitation. Biomed Microdevices 20:64. https://doi.org/10.1007/s10544-018-0312-6
14. Kim Y, Cheng S, Diakite M, Gullapalli R, Simard J, Desai J (2017) Toward the development of a flexible mesoscale MRI-compatible neurosurgical continuum robot. IEEE Trans Robot 33:1386–1397. https://doi.org/10.1109/tro.2017.2719035
15. Kim Y, Cheng S, Desai J (2017) Active stiffness tuning of a spring-based continuum robot for MRI-guided neurosurgery. IEEE Trans Robot 34:18–28. https://doi.org/10.1109/TRO.2017.2750692
16. Jin G, Shuxiang G, Maoxun L, Tamiya T (2018) A marker-based contactless catheter-sensing method to detect surgeons' operations for catheterization training systems. Biomed Microdevices 20:76. https://doi.org/10.1007/s10544-018-0321-5
17. Shuxiang G, Yuan W, Nan X, Youxiang L, Yuhua J (2018) Study on real-time force feedback with a master-slave interventional surgical robotic system. Biomed Microdevices 20(37). https://doi.org/10.1007/s10544-018-0278-4
18. Linshuai Z, Shuxiang G, Huadong Yu YS (2017) Performance evaluation of a strain-gauge force sensor for a haptic robot-assisted catheter operating system. Microsyst Technol 23:5041–5050. https://doi.org/10.1007/s00542-017-3380-2
19. Jian G, Shuxiang G, Yang Y (2016) Design and characteristics evaluation of a novel teleoperated robotic catheterization system with force feedback for endovascular interventional surgery. Biomed Microdevices 18:76. https://doi.org/10.1007/s10544-016-0100-0
20. Vandini A, Bergeles C, Glocker B, Giataganas P, Guangzhong Y (2017) Unified tracking and shape estimation for concentric tube robots. IEEE Trans Robot 33:901–915 **1109/TRO.2017.2690977**
21. Jian G, Xiaolaing J, Shuxiang G (2018) Study of the operational safety of an endovascular interventional surgical robotic system. Micromachines 9:119. https://doi.org/10.3390/mi9030119
22. Linshuai Z, Shuxaing G, Huadong Y, Yu S, Tamiya T, Hirata H, Ishihara H (2018) Design and performance evaluation of collision protection-based safety operation for a haptic robot-assisted catheter operating system. Biomed Microdevices 20:22. https://doi.org/10.1007/s10544-018-0266-8
23. Murali A, Sen S, Kehoe B, Garg A, McFarland S, Patil S, Boyd WD, Lim S, Abbeel P, Goldberg K (2015) Learning by observation for surgical subtasks: multilateral cutting of 3D viscoelastic and 2D Orthotropic Tissue Phantoms. ICRA. https://doi.org/10.1109/ICRA.2015.7139344
24. Shademan A, Decker RS, Opfermann JD, Leonard S, Krieger A, Kim PCW (2018) Supervised autonomous robotic soft tissue surgery. Sci Transl Med 8:337ra64. https://doi.org/10.1126/scitranslmed.aad9398
25. Xianqiang B, Shuxiang G, Nan X, Youxiang L, Cheng Y, Rui S, Jinxin C, Yuhua J, Xinke L, Keyun L (2018) Operation evaluation in-human of a novel remote-controlled endovascular interventional robot. Biomed Microdevices 20:34. https://doi.org/10.1007/s10544-018-0277-5
26. Yan Z, Shuxiang G, Nan X, Yuxin W, Youxiang L, Yuhua J (2018) Operating force information on-line acquisition of a novel slave manipulator for vessel interventional surgery. Biomed Microdevices 20(33). https://doi.org/10.1007/s10544-018-0275-7
27. Hui L, Jianxin Y, Huchen L, Guangdong T (2018) Acquiring and sharing tacit knowledge based on interval 2-tuple linguistic

assessments and extended fuzzy petrinets. Int J Uncertain Fuzz 26: 43–65. https://doi.org/10.1142/S0218488518500034

28. Kassahun Y, Yu B, Tibebu A, Stoyanov D, Giannarou S, Metzen J, Poorten E (2016) Surgical robotics beyond enhanced dexterity instrumentation: a survey of machine learning techniques and their role in intelligent and autonomous surgical actions. Int J Comput Ass Rad 11:847–847. https://doi.org/10.1007/s11548-015-1340-9
29. Mayer H, Gomez F, Wierstra D, Nagy I (2006) A system for robotic heart surgery that learns to tie knots using recurrent neural networks. IROS 22(13):543–548. https://doi.org/10.1163/156855308X360604
30. Seita D, Krishnan S, Fox R, McKinley S, Canny J, Goldberg K (2018) Fast and reliable autonomous surgical debridement with cable-driven robots using a two-phase calibration procedure. ICRA. https://doi.org/10.1109/ICRA.2018.8460583
31. Kassahun Y, Yu B, Poorten E (2013) Learning catheter-aorta interaction model using joint probability densities. Joint Workshop on New Technologies for CRAS. https://lirias.kuleuven.be/handle/123456789/415899. Accessed 12 Nov 2016
32. Hedyeh RT, Jindong L, Su-Lin L, Colin B, Guangzhong Y (2013) Learning-based modeling of endovascular navigation for collaborative robotic catheterization. Med Image Comput Comput Assist Interv 16:369–377. https://doi.org/10.1007/978-3-642-40763-5_46
33. Hedyeh RT, Jindong L, Christopher JP, Colin B, Guangzhong Y (2014) Hierarchical HMM based learning of navigation primitives for cooperative robotic endovascular catheterization. Med Image Comput Comput Assist Interv 17:496–503. https://doi.org/10.1007/978-3-319-10404-162
34. Jeremy DB, Conor EOB, Sarah CL, Kristoffel RD, David IL, Katherine JK (2017) Using contact forces and robot arm accelerations to automatically rate surgeon skill at peg transfer. IEEE Trans Biomed Eng 64:2263–2675. https://doi.org/10.1109/TBME.2016.2634861
35. Wenqiang C, Jindong L, Hedyeh RT, Celia R, Colin B, Guangzhong Y (2018) Learning-based endovascular navigation through the use of non-rigid registration for collaborative robotic catheterization. Int J Comput Assist Radiol Surg 13:855–864. https://doi.org/10.1007/s11548-018-1743-5
36. Olga R, Jia D, Hao S, Jonathan K, Sanjeev S, Sean M, Zhiheng H, Andrej K, Aditya K, Michael B, Alexander CB, Feifei L (2015) ImageNet large scale visual recognition challenge. Int J Comput Vis 115:211–252. https://doi.org/10.1007/s11263-015-0816-y
37. David S, Aja H, Chris JM, Arthur G, Laurent S, George van den D, Julian S, Ioannis A, Veda P, Marc L, Sander D, Dominik G, John N, Nal K, Ilya S, Timothy L, Madeleine L, Koray K, Thore G, Demis H (2016) Mastering the game of Go with deep neural networks and tree search. Nature 529:484–489. https://doi.org/10.1038/nature16961
38. Volodymyr M, Koray K, David S, Alex G, Ioannis A, Daan W, Martin R (2013) Playing Atari with deep reinforcement learning. NIPS Deep Learning Workshop, arXiv:1312.5602
39. Qi D, Hao C, Lequan Y, Jing Q, Pheng Ann H (2017) Multilevel contextual 3-D CNNs for false positive reduction in pulmonary nodule detection. IEEE Trans Biomed Eng 64:1558–1567. https://doi.org/10.1109/TBME.2016.2613502
40. Sergey L, Chelsea F, Trevor D, Pieter A (2016) End-to-end training of deep visuomotor policies. J Mach Learn Res 17:1–40. https://doi.org/10.1016/j.dam.2015.09.011
41. Roberto C, Andrew O, Dinesh J, Justin L, Wenzhen Y, Jitendra M, Edward HA, Sergey L (2017) More than a feeling: learning to grasp and regrasp using vision and touch. IEEE Robotics and Automation Letters 3:3300–3307. https://doi.org/10.1109/LRA.2018.2852779
42. Sergey L, Peter P, Alex K, Julian I, Deirdre Q (2017) Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection. Int J Robot Res 37:421–436. https://doi.org/10.1177/0278364917710318
43. Fangyi Z, Jürgen L, Michael M, Ben U, Peter C (2015) Towards vision-based deep reinforcement learning for robotic motion control. In: Australasian Conference on Robotics and Automation, https://doi.org/10.1016/j.seppur.2006.09.006
44. Mingxin J, Rong L, Jian J, Binjie Q (2017) Extracting contrast-filled vessels in X-ray angiography by graduated RPCA with motion coherency constraint. Pattern Recogn 63:653–666. https://doi.org/10.1016/j.patcog.2016.09.042
45. Binjie Q, Mingxin J, Dongdong H, Yisong L, Qiegen L, Yueqi Z, Song D, Jun Z, Baowei F (2019) Accurate vessel extraction via tensor completion of background layer in X-ray coronary angiograms. Pattern Recognit 87:38–54. https://doi.org/10.1016/j.patcog.2018.09.015
46. Pierre A, Ihor S, Daniel R, Wiro JN, Adriaan M, Theo VW (2017) A hidden Markov model for 3D catheter tip tracking with 2D X-ray catheterization sequence and 3D rotational angiography. IEEE Trans Med Imaging 36:757–768. https://doi.org/10.1016/0.1109/TMI.2016.2625811
47. Li W, Xiaoliang X, Guilin B, Zengguang H, Xiaoran C, Pusit P (2017) Guide-wire detection using region proposal network for X-ray image-guided navigation. IEEE IJCNN:3169–3175. https://doi.org/10.1109/IJCNN.2017.7966251
48. Christian S, Vincent V, Sergey I, Jonathon S, Zbigniew W (2016) Rethinking the inception architecture for computer vision. IEEE CVPR:2818–2826. https://doi.org/10.1109/CVPR.2016.308
49. Nasr-Esfahania E, Karimia N, Jafaria MH, Soroushmehrbc SMR, Samaviabc S, Nallamothuc BK, Najarianbc K (2018) Segmentation of vessels in angiograms using convolutional neural networks. Biomedical Signal Processing and Control 40:240–251. https://doi.org/10.1016/j.bspc.2017.09.012
50. Kragic D From active perception to deep learning. Sci Robot 3: eaav1778 http://robotics.sciencemag.org/. Accessed 8 Dec 2016
51. Guangzhong Y, James C, Kevin C, Eric D, James D, Pierre ED, Nobuhiko H, Peter K, Sylvain M, Rajni VP, Veronica JS, Russell HT (2017) Medical robotics—regulatory, ethical, and legal considerations for increasing levels of autonomy. Sci. Robot. 2. https://doi.org/10.1126/scirobotics.aam8638
52. Pierre BR, Maura P, Hawkeye K, Guangzhong Y (2016) Hubot: a three state human-robot collaborative framework for bimanual surgical tasks based on learned models. ICRA. https://doi.org/10.1109/ICRA.2016.7487198



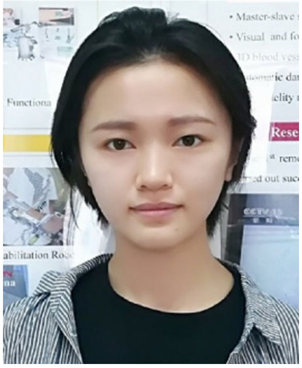
Yan Zhao is a Ph.D. student in biomedical engineering at the Beijing Institute of Technology. He researches on autonomous surgical robot and human-robot collaboration.

Jinxin Cui is a graduate student in biomedical engineering at the Beijing Institute of Technology. He is interested in the application of machine learning for a surgeon's skill analysis.

Shuxiang Guo is a professor in biomedical engineering at the Beijing Institute of Technology and Kagawa University. He has studied on surgical robotics, rehabilitation robotics, and endoscopic capsule robotics for more than 30 years. He is an editor in chief of IJMA (International Journal of Mechatronics and Automation).

Youchun Ma is a graduate student in biomedical engineering at the Beijing Institute of Technology. She is interested in the application of deep learning for a surgical robot.

Yuxin Wang is a graduate student in biomedical engineering at the Beijing Institute of Technology. He is interested in deep learning and autonomous surgical robot.



Yuwen Zeng is a graduate student in biomedical engineering at the Beijing Institute of Technology. She is interested in medical image processing and intelligent diagnosis of aneurysm.