Automatic Microsurgical Skill Assessment Based on Cross-Domain Transfer Learning

Dandan Zhang, *Student Member, IEEE*, Zicong Wu, Junhong Chen, Anzhu Gao, Xu Chen, Peichao Li, Zhaoyang Wang, Guitao Yang, Benny Lo, *Senior Member, IEEE*, and Guang-Zhong Yang, *Fellow, IEEE*

***Abstract*—The assessment of microsurgical skills for Robot-Assisted Microsurgery (RAMS) still relies primarily on subjective observations and expert opinions. A general and automated evaluation method is desirable. Deep neural networks can be used for skill assessment through raw kinematic data, which has the advantages of being objective and efficient. However, one of the major issues of deep learning for the analysis of surgical skills is that it requires a large database to train the desired model, and the training process can be time-consuming. This letter presents a transfer learning scheme for training a model with limited RAMS datasets for microsurgical skill assessment. An in-house Microsurgical Robot Research Platform Database (MRRPD) is built with data collected from a microsurgical robot research platform (MRRP). It is used to verify the proposed cross-domain transfer learning for RAMS skill level assessment. The model is fine-tuned after training with the data obtained from the MRRP. Moreover, microsurgical tool tracking is developed to provide visual feedback while task-specific metrics and the other general evaluation metrics are provided to the operator as a reference. The method proposed has shown to offer the potential to guide the operator to achieve a higher level of skills for microsurgical operation.**

***Index Terms*—Microsurgical skill analysis, transfer learning.**

I. Introduction

WITH the recent advances in robotics, micro-surgical robotic technologies have developed rapidly which significantly improves the accuracy and dexterity in microsurgery. Robot-Assisted Microsurgery (RAMS) can reduce blood loss and lead to fewer complications, shorter operating time, and lower treatment costs. To realise the full potential of RAMS, microsurgical training is important for surgical trainees to master the skills required to operate on patients.

Manuscript received December 13, 2019; accepted April 7, 2020. Date of publication April 20, 2020; date of current version May 12, 2020. This letter was recommended for publication by Associate Editor E. De Momi and Editor P. Valdastri upon evaluation of the reviewers' comments. *(Corresponding authors: Dandan Zhang; Benny Lo; Guang-Zhong Yang).*

Dandan Zhang, Zicong Wu, Junhong Chen, Xu Chen, Peichao Li, Zhaoyang Wang, and Benny Lo are with the The Hamlyn Centre for Robotic Surgery, Imperial College London, London SW7 2AZ, U.K. (e-mail: d.zhang17@imperial.ac.uk; ZXW658@gmail.com; junhong.chen16@imperial.ac.uk; xu.chen18@imperial.ac.uk; p.li18@imperial.ac.uk; zhaoyang.wang18@imperial.ac.uk; benny.lo@imperial.ac.uk).

Anzhu Gao and Guang-Zhong Yang are with the The Hamlyn Centre for Robotic Surgery, Imperial College London, London SW7 2AZ, U.K., and also with the The Institute of Medical Robotics, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: anzhu_gao@sjtu.edu.cn; gzyang@stju.edu.cn).

Guitao Yang is with the The Electrical and Electronic Engineering Department, Imperial College London, London SW7 2BU, U.K. (e-mail: guitao.yang16@imperial.ac.uk).

Digital Object Identifier 10.1109/LRA.2020.2989075

Thus far, most of the assessment techniques are performed via outcome-based analysis [1], structured checklists, and rating scales [2]. These can be subjective and only provide an overall assessment of the outcome measures. The Objective Structured Assessment of Technical Skills (OSATS) [3] is a validated tool commonly used for the Structured Assessment of Robotic Microsurgical Skills [4]. Some other scoring methods have been proposed, including the Global Rating Scales [5], Structured Assessment of Microsurgery [6] and Modified Global Rating Scales [7]. Subjective methods require an expert surgeon to score the performance of the trainee subjectively, the grading process of which is time-consuming, expensive and inconsistent due to the inherent biases in human interpretations [8].

For objective skill assessment methods, standard and specific evaluation metrics can be calculated based on motion data for the quantitative analysis. These includes the tooltip trajectory [9], clutching number and velocity [10]. Neurosurgery skill assessment has been conduced based on the tooltip acceleration and angular velocity in a simple pick and place task [11]. Other evaluation metrics have been proposed based on motion data analysis [12]. For example, the variants of trajectory [13], curvature [14], motion jerk value [15], energy based metrics [16], semantic labels [17] and other extracted features that have high correlations with surgical skills can be utilized as well. Though descriptive analysis based on evaluation metrics can assist automatic skill assessment, it is difficult to determine which kinds of metrics should be used and how to weigh the metrics for overall performance assessment.

Raw motion data can be transformed to intermediate interpretations with advanced feature selection techniques. The high-level features can then be used for skill level classification with the emerging machine learning methods. Novel feature fusion has been explored for surgical skill level classification, which fuses four different types of holistic features from robot kinematic data based on Sequential Motion Texture, Discrete Fourier Transform, Discrete Cosine Transform and Approximate Entropy [18]. For traditional machine learning methods, the skill assessment results depend significantly on the feature extraction process. However, critical information may be inadvertently discarded during feature extraction process, and complex motion profiles are difficult to be fully explored.

Deep Learning techniques can learn discriminative features efficiently and perform feature extraction progressively to discover abstract representations during the training process. This kind of end-to-end learning method has been applied in [19] for skill assessment. In order to decode skill information from raw motion profiles via end-to-end learning, the availability of database is significant.

Emerging platforms for Robot-Assisted Minimally Invasive Surgery (RAMIS) are able to collect multi-modality data from surgical robots or simulators [20], such as vision and kinematic data, which can be used for skill assessment [21]. For example, JIGSAWS (JHU-ISI Gesture and Skill Assessment Working Set) [22] is a publicly available database, which includes three typical tasks performed by surgeons with various skill levels, i.e., suturing, knot tying and needle passing. JIGSAWS has been used extensively for surgical gesture segmentation, classification and skill assessment. However, there is so far no publicly available database for RAMS.

Though deep learning techniques can be used for skill assessment with the advantage of automatic feature decoding, insufficient data has limited its widespread applications. Since there are various types of robot-assisted surgery, the data is collected based on different surgical robotic platforms with different types of features and sensing modalities. It's difficult to collect large databases based on various situation and train a new model using raw data. In this case, transfer learning is a promising method. To address the issues due to the lack of sufficient reference datasets, we investigate the use of transfer learning, which enables the translation of deep learning model learnt from existing surgical datasets to new surgical tasks or systems. It helps to remove the need for data recollection for different surgical tasks based on different types of surgeries performing on different surgical robotic platforms.

Transfer learning has been widely used in medical imaging. As for RAMS applications, transfer learning has been explored for microsurgical tool segmentation. For example, in the work about binary segmentation of neurosurgical instrument [23], transfer learning was used to generate architectures with pre-learned knowledge from [24]. However, cross-domain transfer learning has not been explored in automatic microsurgical skill assessment. The aim of this paper is to introduce a novel automatic microsurgical skill assessment method based on cross-domain transfer learning.

The main contributions of this paper are listed as follows.

1) Transfer learning is used to realize automatic skill assessment with data in the Microsurgical Robot Research Platform Database (MRRPD). The pre-trained model is obtained via the JIGSAWS and the efficacy of the model is evaluated prior to transferring the model for microsurgical skill assessment for RAMS based on kinematic data.
2) Microsurgical tool tracking is used for task-specific features estimation. Visual feedback is provided, which can assist the operators to grasp the higher level of microsurgical skills for RAMS based on the vision data. General skill evaluation metrics can be deduced and provide feedback to the operators.
3) Interpretable feedback obtained from the neural network model could be used as a reference to explain which period of the operation can be used to distinguish the operator as a novice/intermediate or expert, which elucidates the underlying mechanism of the classification.

The details of the proposed framework are described as follows. Firstly, the system overview is introduced in Section II, where the microsurgical research platform, the microsurgical tasks and the database information are described. Secondly, the architecture of the neural network and the proposed transfer learning method are detailed in Section III. Thirdly, the visual feedback and other evaluation metrics are described to assist the explainability of the assessment results for microsurgical skills in Section IV. Conclusions are drawn in Section V along with the relative merits and potential future improvements of the method.

II. System Overview

In this section, the database construction process is introduced. The microsurgical research platform, the microsurgical tasks and the database information are described.

A. Experimental Platform

Most of the surgeons for microsurgery acquire skills by watching microsurgical videos, observing an expert's operation in the operating room and practise using phantoms or animals. This is a time-consuming process, while advanced surgical assessment systems can benefit microsurgical skill training by using recording devices and computational technologies for automatic skill assessment [8]. Therefore, a robotic platform with operation data collection is significant. With proper robot-assisted surgical training system, the automatic skill assessment turns to be possible and will benefit the microsurgical training.

A Microsurgical Robot Research Platform (MRRP) is used to collect data for RAMS skill level evaluation, while a pair of Phamtom Omni device is used as the master manipulators to simulate teleoperation [25], [26]. In this system, the hand motion of the operator measured with a master manipulator is scaled down with a fixed motion-scaling ratio using the master-slave control.

The features of the software architecture are illustrated as follows.

- The software architecture is developed using C++ and Python as the programming language and can be implemented in a Linux system.
- Robot Operating System (ROS) serves as the middleware for the control of the MRRP, which is convenient for researchers to implement high-level control framework.
- Real-time microsurgical robot control performance can be achieved.
- A QT5-based GUI is constructed to provide convenience for users to interact with the robot.

The operator can manipulate the master manipulators to control the slave robot, as shown in Fig. 1(a). Fig. 1(b) illustrates the scenario for visual feedback on the monitor, which is a part of the control console. Fig. 1(c) shows the microscopic views of the data collected during the path following task and positioning task respectively.

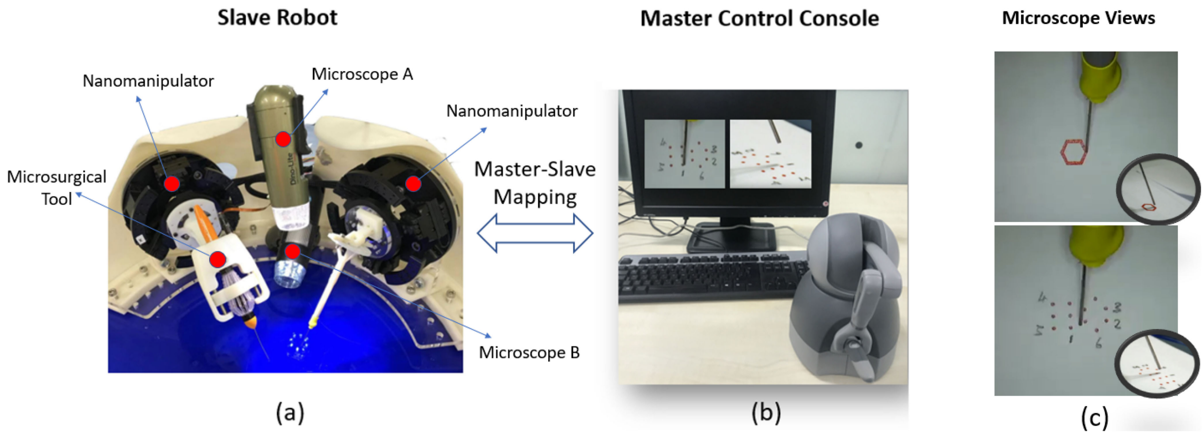

Fig. 1. Overview of the scenario for the master-slave remote control. (a) The MRRP slave robot and visual system. (b) Master control console. The left image on the monitor is the top-to-down view; the right image on the monitor is for depth monitoring. (c) Microscopic views.

B. Microsurgical Tasks

RAMS can benefit from the automation of common, repetitive or well-defined but ergonomically difficult sub-tasks. The microsurgical training can start from practising several microsurgical tasks before conducting more complex microsurgical procedures.

Although there are many different kinds of microsurgical procedures, the basic requirements remain similar among all of them. Three types of microsurgical tasks are defined as exemplars, which include a positioning task, a path following task and a needle insertion task.

- **Positioning Task**: The purpose of this task is to examine the ability of the operator's precision in targeting. The positioning task requires the subject to aim and place the needle tip at the targeted points. More specifically, this task can be used to evaluate whether the user can reach the specified targets within tolerable errors or not.
- **Path Following Task**: This task is designed to examine the ability of the operator in terms of the general position-changing maneuvers. The operator is required to trace a pre-defined trajectory, which is a simple geometric contour printed on a piece of paper or drawn on a fabric. The operator is asked to place the needle tip as close as possible to the paper or the fabric.
- **Needle Insertion Task**: This task is aimed to examine the operators' skill in reaching the specific points with high precision. The operator is required to adjust the tip of a needle on the surface of a phantom and simulate the needle insertion process by pointing the needle tip to the pre-defined targets.

Two experimental phantoms designed for the microsurgical tasks are shown in Fig. 2.

C. Database

The Microsurgical Robot Research Platform Database (MR-RPD) was built up with the data collected during the tasks mentioned above. The two nanomanipulators (SmarAct, Germany) with parallel kinematic structure for microsurgical tasks can be defined as the PSMs (Patient Side Manipulators), while the two Phantom Omni are regarded as the MTMs (Master Tool Manipulators).

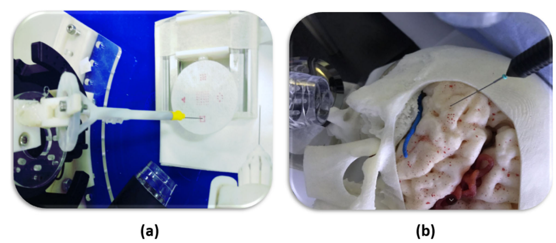

Fig. 2. Experiment set up for data collection. (a) Experimental platform designed for the positioning task and the path following task. (b) Experimental platform designed for the needle insertion task.

The data collected during microsurgical operation include the vision data from the top-to-down view of Microscope A as well as the side view of Microscope B, and the kinematic data of the movements of the manipulators of MRRP are also collected.

Eight subjects were recruited to join in the experimental data collection. For each operation task, the operator performs the same trial for multiple times. For the same task, operation procedures are the same, which makes the skill assessment to be fair. All the subjects had five minutes to practise and get familiarised with the tasks before the actual experiments. They were asked to go through the whole procedure twice to get accustomed to the experimental protocols. Afterwards, the participants were asked to perform the designed tasks and the data was then collected. 20-24 trials were collected for different tasks respectively.

III. The Transfer Learning Framework

In this section, the deep neural model architecture is briefly explained, demonstrating the concept of end-to-end learning. Transfer learning method is explored for the skill assessment for RAMS.

A. Database for Deep Learning

Knowledge transfer is challenging due to the differences in data distribution and context. A deep neural network model can be trained based on the surgical task for laparoscopic surgery,

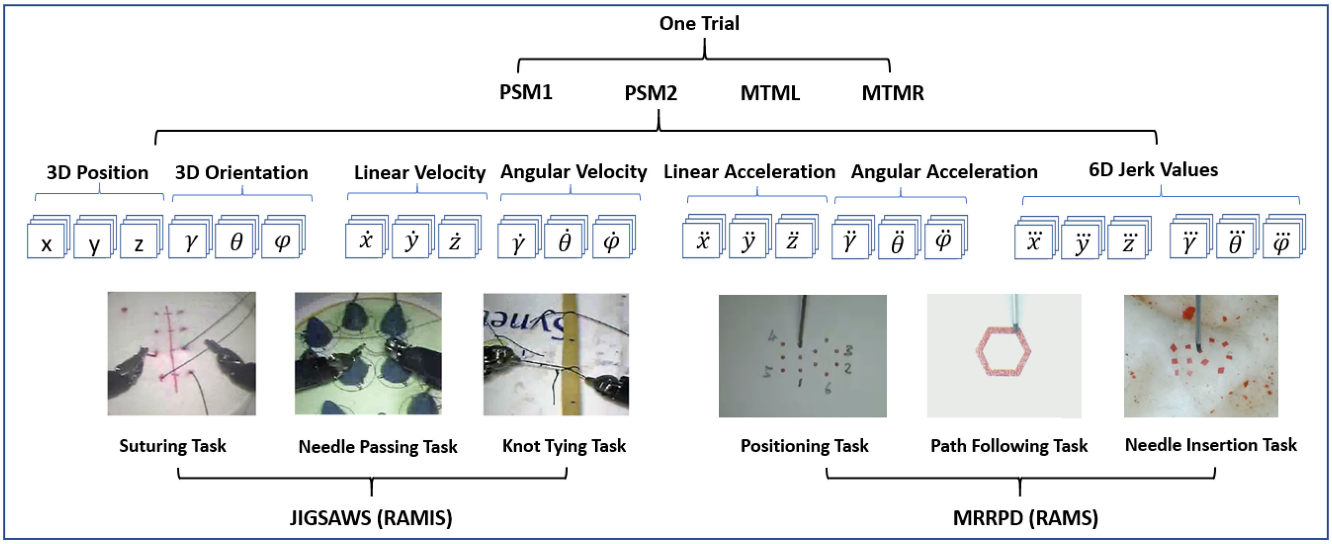

Fig. 3. The re-organized kinematic data structure of the JIGSAWS and the MRRP.

while the model can be fine-tuned to be adaptive to the database for microsurgical operation.

The JIGSAWS dataset has been collected from eight right-handed subjects with three different skill levels by performing three surgical tasks (suturing, needle passing and knot tying) using the da Vinci surgical system. 103 trials were collected in total. The kinematic data collected from four manipulators, including the left and right masters and slave robots. 76 kinematic variables are captured at a frequency of 30 frames per second for each trial [22].

In order to simplify the process of transfer learning, the kinematic data of the original JIGSAWS database is reorganized. The arrangement of the kinematic data in JIGSAWS database can be viewed in Fig. 3. The original rotation matrix of each data frame is transformed to the representation of Euler angle, which reduces the data dimension. The acceleration and jerk values are obtained and added to each data frame. The 3D position data of the end-effector, the linear and angular velocities are preserved.

The overall kinematic data include the operation information of the right Patient Side Manipulator (PSM1), the left Patient Side Manipulator (PSM2), the left Master Tool Manipulator (MTML) and the right Master Tool Manipulator (MTMR).The targeted classes include three surgical levels, i.e. Novice (N), Intermediate (I) and Expert (E). The annotation of skill level labels follows the same method provided by JIGSAWS, which is determined by the operators' experience. The features used for both database are re-organized to be the same.

B. Architecture

A hierarchical structure is utilized to construct the deep learning model. For sequential data based deep neural network, 1D CNN can be used for raw data feature extraction. The Rectified Linear Unit (ReLU) activation function can be employed after each convolution operation.

For sequential data processing or modeling, recurrent neural networks, in particular long short-term memory (LSTM), have been verified to be effective. Except for using one dimensional convolution operation, LSTM can be used.

The raw data layer included the original kinematic data for each arm. The feature cluster layer groups the columns of kinematic data into clusters, which includes the position, velocity, acceleration and jerk values.

One-arm characteristic layer can be obtained. The characteristic layer is formed by merging all the information from different arms after two convolution operations followed by the ReLU activation function. A Global Average Pooling layer is used, followed by a fully-connected layer with softmax activation that produces the classification result to estimate the microsurgical skill level.

The multinomial cross-entropy is used as our objective cost function for training the network, and Adam is used to optimize the parameters [27]. The architecture for the Deep Neural Network Model training is shown in Fig. 4. Data is normalized by scaling the values between 0 and 1, which means that each value of the same type of feature is subtracted with the minimum value among the same feature and is divided by the difference between the maximum and minimum value before training for both JIGSAWS and MRRPD.

The network can be re-organized based on different configuration of data structure collected during experiments. For example, if no master manipulators were used during the experiments, the one-arm characteristic layer merging process can only consist of PSM1 and PSM2. If bimanual operation is not necessary, the one-arm characteristic layer merging process can only include the kinematic data of MTMR and PSM1, or MTML and PSM2. To accelerate the training process, the jerk values information can be excluded to reduce the feature dimensions. This means that the raw data layer can be flexibly defined, with different combinations of feature clusters and information from different arms.

To summarize, the raw data layer and the feature cluster layer are targeted at extracting features in latent low-level variables, while the third layer captures the global information related to the surgical skill level. After a preliminary test using the 1D CNN

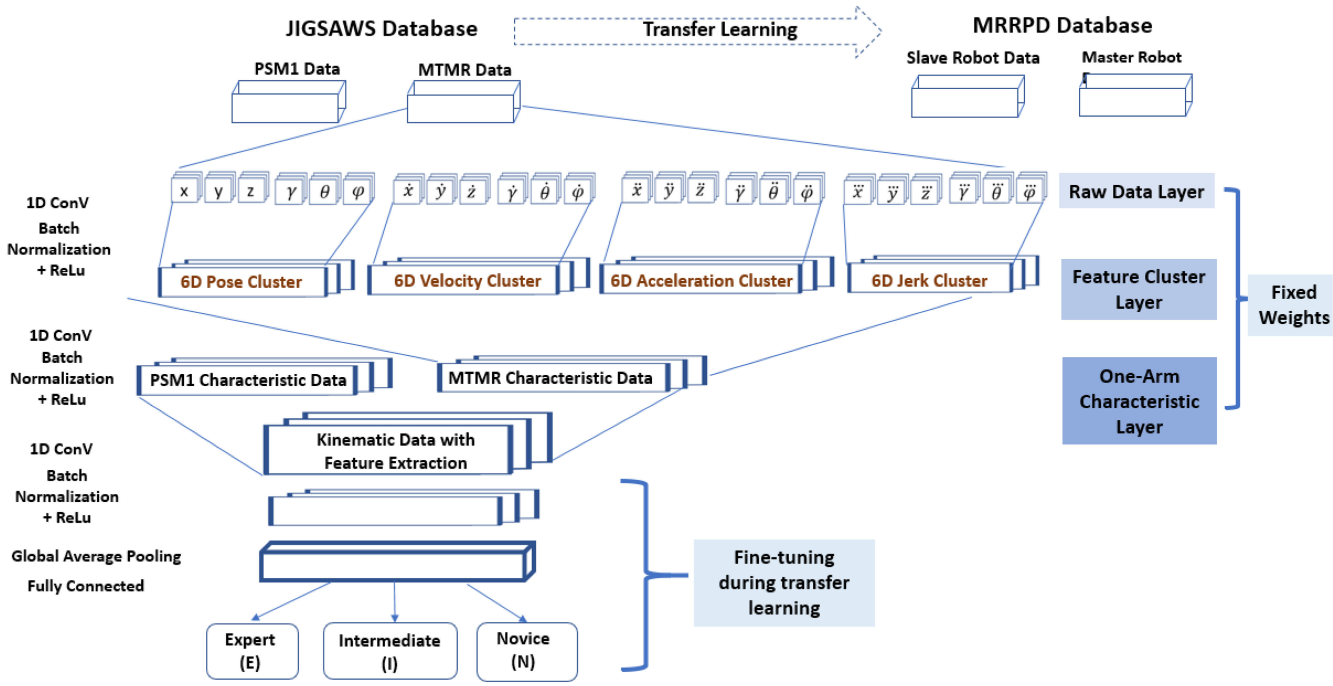

Fig. 4. The architecture for the deep neural network model and the implementation of transfer learning.

and LSTM, we can observe that the training time of using the 1D CNN is much shorter than LSTM, while the testing accuracy is higher. Therefore, 1D CNN is chosen for the proposed approach.

C. Transfer Learning

With a small-scale dataset, over-fitting is difficult to avoid when training deep neural network. Unlike other domains, it is not feasible to collect a large amount of data for surgical motions and label them objectively for training purposes. Transfer learning is an effective tool, especially when the size of dataset is limited, through which the domain randomization can be used to improve generalization [28].

The goal here is to transfer the knowledge gained from the JIGSAWS to accelerate the learning on microsurgical skill level classification. This instance of transfer learning is known as domain adaptation, a scenario in transductive transfer learning [29], where the targeted classes for identification from the source and target tasks are the same, but the data distribution of target and source domain are different.

During the transfer learning, the parameters of the first and second layers are fixed. This means that we don't need to train a complete new model from scratch. We assumed that the feature extraction mode can be similar. The fixed layer and the trainable layer are illustrated in Fig. 4.

D. Results

The rationality of the deep neural network model is verified based on the three surgical tasks separately to see the classification accuracy. After that, the pretrained model obtained based on the JIGSAWS database is transferred and evaluated on the RAMS database.

TABLE I
RESULTS OF THE DEEP NEURAL NETWORK MODEL

Surgical Tasks	Macro Average	Micro Average
Suturing	0.9917±0.0236	0.9844±0.0442
Needle Passing	0.9097±0.0643	0.8685±0.0847
Knot Tying	0.8472±0.0813	0.8080±0.0780

TABLE II
RESULTS FOR THE TRANSFER LEARNING

Transfer Type	Micro Average	Macro Average
Suturing → Path Following	0.8472±0.0802	0.8333±0.0642
Suturing → Positioning	0.9792±0.0417	0.9722±0.0556
Suturing → Needle Insertion	0.8889±0.1571	0.8750±0.1596

Cross validation is used to evaluate the performance of the deep neural network based on the leave-one-super-trial-out (LOSO) configuration for the deep neural network model. In order to remove the outlier, the best and the worst results were abandoned, while the mean value and standard deviation of 8 tests were calculated. The learning rate we used is 0.0001, while 500 epoches are used to train the model.

Table II shows the cross-validation results, the performance of each technique is evaluated with the Micro and Macro average accuracy respectively. Suppose that there are n targeted classes, $C = [i, j]$ represents the confusion matrix with the dimension of $n \times n$. Each element of the confusion matrix represents that the sample from class i is predicted as class j. Given the confusion matrix, Micro average is computed as the average of total correct predictions across all classes, while Macro represents the average true positive rates for each class. Micro and Macro Average can be calculated as follows.

$$\begin{cases} Micro = \frac{\sum_{i=1}^{n} C[i,i]}{\sum_{i,j=1}^{n} C[i,j]} \\ Macro = \frac{1}{n}\sum_{i=1}^{n} \frac{C[i,i]}{\sum_{j=1}^{n} C[i,j]} \end{cases} \quad (1)$$

Table I shows the training results based on the JIGSAWS database. It can be concluded that the deep neural network has high testing accuracy for the suturing task and the needle passing task, while the testing accuracy for knot-tying task is acceptible.

Table II includes the results of transfer learning. Since the testing accuracy for suturing task is the highest, the learned model is transfer for the path following task, the positioning task and the needle insertion task. Results indicated that the transfer learning is effective for microsurgical skill assessment.

IV. Evaluations With Interpretability

In this section, the relevant evaluation metrics for skill assessments are summarized and served as part of the systematic analysis for microssurgical skill level assessment. The visualization of discriminative and interpretable skill assessment process is important for providing helpful feedback to trainees. Therefore, we try to provide explainable results along with the transfer learning method to pave a way for guiding the trainees to achieve higher level of skills.

A. Interpretability of the Transfer Learning Results

It is important for the trainee to gain valuable feedback, which guides them to gain higher level of skills. Class activation map can be generated from a global activation map layer, which is able to mitigate the black-box effect of the deep neural model by visualizing the critical parts that contributes the most for the classification results [30]. By constructing a heatmap from the class activation map, the main reason behind the subject's classification results can be identified.

Fig. 5(a) is an example of using the class activation map to identify the discriminative behaviors that is specific to a novice or an expert through visualization. The red area in Fig. 5 represents the main reason behind the subject's classification results. Based on the visual feedback, the operators can pay more attention to the practise of path following in that specific segments to accelerate the improvement of the overall operation performance.

B. Task-Specific Metrics Visualization

Some task-specific features are required to evaluate skills for specific types of surgery, while genetic metrics are not sufficient enough for evaluation of the surgical skill for different surgery types [31], [32]. Therefore, several metrics can be considered based on the specific requirements from different specific tasks. Precision Degree (P) can be used to evaluate the positioning task or needle insertion task while Root Mean Square Error ($RMSE$) can be used to evaluate the path following task. Two evaluation metrics are chosen and listed as follows.

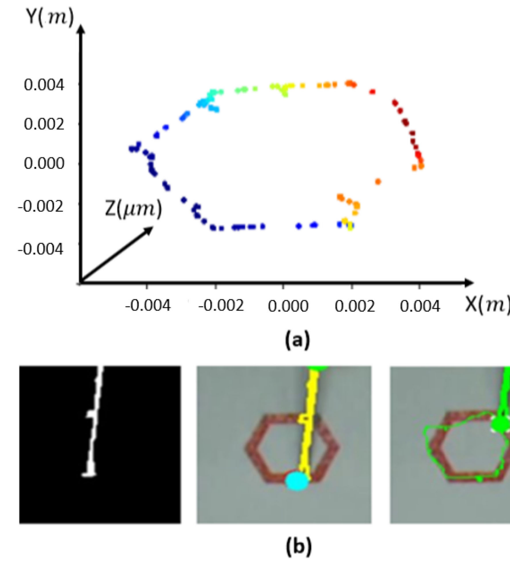

Fig. 5. Vision based skill analysis results. (a) Class activation map. (b) Visual tracking feedback.

- **Precision Degree** (P): P calculates the average error of reaching the pre-defined targeted points. It is used to evaluate whether the operator can reach the targeted place without significant deviation.
- **Root Mean Square Error** ($RMSE$): RMSE is the square root of the variance of the trajectory, known as the standard error between the desired trajectory and the real trajectory. It is used to compare the difference between the actual trajectory performed by the users and the optimized trajectory [33].

Suppose that there are m points for positioning during one microsurgical task, the position of the targeted points are $\boldsymbol{P_t(i)} = [X_t(i), Y_t(i)](i = ,1, 2...m)$. Suppose that $\boldsymbol{Pg_{1:T}} = [Xg_{1:T}, Yg_{1:T}]$ represents the desired trajectory for following, While $\boldsymbol{Pr_{1:T}} = [Xr_{1:T}, Yr_{1:T}]$ represents the position profile obtained by the tracking algorithm. P and $RMSE$ can be calculated as follows.

$$\begin{cases} P = \frac{1}{m}\sum_{i=1}^{i=m} min(||\boldsymbol{P_t(i)} - \boldsymbol{Pg(t)}||, (t \in 1:T) \\ RMSE = \frac{1}{T}\sum_{t=1}^{t=T} ||\boldsymbol{Pr(t)} - \boldsymbol{Pg(t)}|| \end{cases} \quad (2)$$

The tooltip position of the microsurgical tools mounted on the MRRP can be tracked automatically for further analysis. During the user studies, the real-time microscope images were collected during the operation. A vision-based tracking method was developed to automatically evaluate the performance of the operators.

Fig. 5(b) shows the process of visual data processing. The video data were collected through the microscopes. The RGB data of microsopic images can be obtained and converted to greyscale image. Gaussian blurring is used to remove noises for stable tracking. The microsurgical tool is segmented. A threshold-based method is used to extract the boundary of the needle. By searching the extreme points among the extracted boundary, the position of the tooltip can be tracked online, which is the end of the line segment. The original trajectory of the

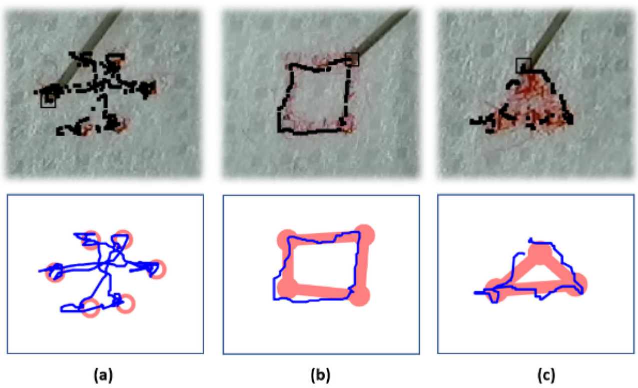

Fig. 6. Vision based real-time tracking results and analysis for (a) Positioning of the vertices of a hexagon. (b) Path following of a square. (c) Path following of a triangle.

TABLE III
EVALUATION RESULTS

	Positioning Task	p-value	Path Following	p-value
T_s	37.2s	0.0008	41.1s	0.0048
P_s	509.4	0.1859	538.6	0.0090
A_s	$15.2s^{-1}$	0.0002	$14.1s^{-1}$	0.0401
$RMSE$	/	/	7.0	0.0068
P	2.3	0.0219	/	/

position of the tooltip can be generated while Kalman filtering is used to get a smooth trajectory. In this way, the stable and accurate tooltip tracking can be realized.

Fig. 6 shows the real-time tracking results, which demonstrates the example of analysis results for task-specific feedback to all the trials after the video data processing.

Except for task-specific features, other general evaluation metrics can be automatically calculated as a reference. For example, the movement time (T_s) [34], the operative speed (A_s) [35], instrument traveled length (P_s) [14]. Table III summarizes the task-specific evaluation metrics for 22 trials as an example. The basic functions of the MRRP can be verified, which indicated that MRRP has the potential to be applied for microsurgical skills training.

Normality test was conducted to verify that the data we used for statistic analysis is normally distributed. An ANOVA significance test between performance metrics and subjects' skill levels is used to verify the significance of the evaluation results, which are shown in Table III. 0.05 is chose as the significance level. Among all the evaluation metrics, P_s(instrument traveled length) is not significantly different among skill levels in the positioning task only ($p = 0.1859$). This may be due to the fact that the main focus of the positioning task is to reach the targeted point precisely, while the trajectory of moving is not optimized during the operations for all the subjects during the experiments.

C. Discussions

In this paper, an automatic microsurgical skill assessment method is developed based on transfer learning to classify the microsurgical skills of the trainees into expert level, intermediate level and novice level. Class activation map is used to provide interpretability for the neural network results, while microsurgical tool tracking is used to obtain the trajectory based on the video data for task-specific feature calculation. In the mean time, the general evaluation metrics can be automatically calculated and provided to the operators as a reference.

For real microsurgery application, the micro-forcep may need to follow a pre-defined trajectory to reach the targeted operation area, which requires proficient path following skill. Micro cannulation and microvascular anastomosis are important microsurgical tasks, which requires robotic assistance for high precision positioning.

In the future, more clinical data can be collected to evaluate the effectiveness of the proposed method. Moreover, surgical tasks can often be represented by certain repetitive and specific gestures, while they can be further decomposed into basic surgemes. To provide more effective feedback to the operator, skill analysis based on specific surgemes can be conducted. Therefore, the operators have clearer targets for improving their operational skills. Furthermore, Explainable AI (XAI) can be explored to enable the end-to-end learning methods to be better understood by surgeons.

V. CONCLUSION

Due to the limitation of available data of RAMS, cross-domain transfer learning is investigated and validated for automatic microsurgical skill assessment. The data distribution of the source and target domains are considerably different while the data structure and the class labels remain the same. We trained a deep learning model for skill assessment based on data collected from laparoscopic surgical tasks. The pre-trained model is tuned and transferred for RAMS applications.

The proposed method is validated on the JIGSAWS dataset and achieved competitive results among the existing aprroaches. From our studies, we demonstrated that the transfer learning method is effective, since the model is trained based on the data of RAMIS while the skill level classification layer is fine-tuned after feeding the data obtained from the MRRP.

The visualization of the interpretable features are employed to explain the underlying mechanism of the skill classification results to the trainees based on class activation map. Moreover, the task-specific features and several general features are automatically calculated and visualized based on reliable tooltip tracking to provide the trainee as the personalized discriminative feedback, which paves a way to guide the novices/intermediate to become an expert.

ACKNOWLEDGMENT

The authors would like to acknowledge all the subjects involved in this study.

REFERENCES

[1] B. Ben *et al.*, "Surgeon specific mortality in adult cardiac surgery: Comparison between crude and risk stratified data," *Brit. Med. J.*, vol. 327, no. 7405, pp. 13–17, 2003.

[2] M. A. Aghazadeh *et al.*, "External validation of global evaluative assessment of robotic skills (gears)," *Surgical Endoscopy*, vol. 29, no. 11, pp. 3261–3266, 2015.

[3] J. Martin, *et al.*, "Objective structured assessment of technical skill (osats) for surgical residents," *Brit. J. Surgery*, vol. 84, no. 2, pp. 273–278, 1997.

[4] P. A. Liverneaux, S. Hendriks, J. C. Selber, and S. J. Parekattil, "Robotically assisted microsurgery: Development of basic skills course," *Arch. Plastic Surgery*, vol. 40, no. 4, pp. 320–326, 2013.

[5] P. U. Kalu, J. Atkins, D. Baker, C. J. Green, and P. E. Butler, "How do we assess microsurgical skill?" *Microsurgery: Official J. Int. Microsurgical Soc. Eur. Federation Soc. Microsurgery*, vol. 25, no. 1, pp. 25–29, 2005.

[6] J. C. Selber *et al.*, "Tracking the learning curve in microsurgical skill acquisition," *Plastic Reconstructive Surgery*, vol. 130, no. 4, pp. 550e–557e, 2012.

[7] D. G. Ezra *et al.*, "Skills acquisition and assessment after a microsurgical skills course for ophthalmology residents," *Ophthalmology*, vol. 116, no. 2, pp. 257–262, 2009.

[8] C. E. Reiley, H. C. Lin, D. D. Yuh, and G. D. Hager, "Review of methods for objective surgical skill evaluation," *Surgical Endoscopy*, vol. 25, no. 2, pp. 356–366, 2011.

[9] D. Zhang, J. Liu, L. Zhang, and G.-Z. Yang, "Hamlyn CRM: A compact master manipulator for surgical robot remote control," *Int. J. Comput. Assisted Radiol. Surgery*, vol. 15, pp. 503–514, 2020.

[10] D. Zhang, B. Xiao, B. Huang, L. Zhang, J. Liu, and G.-Z. Yang, "A self-adaptive motion scaling framework for surgical robot remote control," *IEEE Robot. Autom. Lett.*, vol. 4, no. 2, pp. 359–366, Apr. 2018.

[11] Z. Lin *et al.*, "Objective skill analysis and assessment in neurosurgery by using an ultra-miniaturized inertial measurement unit wb-3—Pilot tests," in *Proc. Annu. Int. Conf. IEEE Eng. Medicine Biol. Soc.*, 2009, pp. 2320–2323.

[12] D. Zhang, J. Liu, A. Gao, and G.-Z. Yang, "An ergonomic shared workspace analysis framework for the optimal placement of a compact master control console," *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 2995–3002, Apr. 2020.

[13] G. Forestier *et al.*, "Surgical motion analysis using discriminative interpretable patterns," *Artif. Intell. Medicine*, vol. 91, pp. 3–11, 2018.

[14] T. N. Judkins, D. Oleynikov, and N. Stergiou, "Objective evaluation of expert and novice performance during robotic surgical training tasks," *Surgical Endoscopy*, vol. 23, no. 3, pp. 590–597, 2009.

[15] K. Liang, Y. Xing, J. Li, S. Wang, A. Li, and J. Li, "Motion control skill assessment based on kinematic analysis of robotic end-effector movements," *Int. J. Med. Robot. Comput. Assisted Surgery Mrcas*, vol. 14, no. 1, p. e1845, 2017.

[16] B. Poursartip, M. E. Lebel, R. Patel, M. Naish, and A. L. Trejos, "Analysis of energy-based metrics for laparoscopic skills assessment," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 7, pp. 1532–1542, Jul. 2018.

[17] M. Ershad, Z. Koesters, R. Rege, and A. Majewicz, *Meaningful Assessment of Surgical Expertise: Semantic Labeling With Data and Crowds*. Berlin, Germany:Springer-Verlag 2016.

[18] A. Zia and I. Essa, "Automated surgical skill assessment in rmis training," *Int. J. Comput. Assisted Radiol. Surgery*, vol. 13, no. 5, pp. 731–739, 2018.

[19] Z. Wang and A. M. Fey, "Deep learning with convolutional neural network for objective skill evaluation in robot-assisted surgery," *Int. J. Comput. Assisted Radiology Surgery*, 2018.

[20] D. Zhang, F. Cursi, and G.-Z. Yang, "Wsrender: A workspace analysis and visualization toolbox for robotic manipulator design and verification," *IEEE Robot. Autom. Lett.*, vol. 4, no. 4, pp. 3836–3843, Oct. 2019.

[21] D. Zhang, J. Liu, L. Zhang, and G.-Z. Yang, "Design and verification of a portable master manipulator based on an effective workspace analysis framework," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2019, pp. 417–424.

[22] Y. Gao *et al.*, "JHU-ISI gesture and skill assessment working set (JIGSAWS): A surgical activity dataset for human motion modeling," in *Proc. MICCAI Workshop: M2CAI*, 2014, vol. 3, p. 3, [Online]. Available: https://cirl.lcsr.jhu.edu/research/hmm/datasets/jigsaws_release/

[23] N. Kalavakonda, B. Hannaford, Z. Qazi, and L. Sekhar, "Autonomous neurosurgical instrument segmentation using end-to-end learning," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit. Workshops*, 2019.

[24] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Represent.*, 2015.

[25] D. Zhang, J. Chen, W. Li, D. B. Salinas, and G.-Z. Yang, "A microsurgical robot research platform for robot-assisted microsurgery research and training," *Int. J. Comput. Assisted Radiol. Surgery*, vol. pp. 15–25, 2020.

[26] D. Zhang, Y. Guo, J. Chen, J. Liu, and G.-Z. Yang, "A handheld master controller for robot-assisted microsurgery," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2019, pp. 394–400.

[27] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Represent.*, Y. Bengio and Y. LeCun. Eds., San Diego, CA, USA, May 7–9, 2015.

[28] J. Tobin, R. Fong, A. Ray, J. Schneider, W. Zaremba, and P. Abbeel, "Domain randomization for transferring deep neural networks from simulation to the real world," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2017, pp. 23–30.

[29] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.

[30] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2016, pp. 2921–2929.

[31] M. J. Fard, S. Ameri, E. R. Darin, R. B. Chinnam, A. K. Pandya, and M. D. Klein, "Automated robot-assisted surgical skill evaluation: Predictive analytics approach," *Int. J. Med. Robot. Comput. Assisted Surgery Mrcas*, vol. 14, no. 1, p. e1850, 2018, doi: 10.1002/rcs.1850.

[32] D. Stefanidis, D. J. Scott, and K. J. Jr, "Do metrics matter? Time versus motion tracking for performance assessment of proficiency-based laparoscopic skills training," *Simul. Healthcare J. Soc. Simul. Healthcare*, vol. 4, no. 4, pp. 104–108, 2009.

[33] K. Harada *et al.*, "Microsurgical skill assessment: toward skill-based surgical robotic control," in *Proc. Int. Conf. IEEE Eng. Medicine Biol. Soc.*, 2011, pp. 6700–6703.

[34] A. L. Trejos, R. V. Patel, R. A. Malthaner, and C. M. Schlachta, "Development of force-based metrics for skills assessment in minimally invasive surgery," *Surgical Endoscopy*, vol. 28, no. 7, pp. 2106–2119, 2014.

[35] Y. Kassahun *et al.*, "Surgical robotics beyond enhanced dexterity instrumentation: A survey of machine learning techniques and their role in intelligent and autonomous surgical actions," *Int. J. Comput. Assisted Radiol. Surgery*, vol. 11, no. 4, pp. 553–568, 2016.