



City Research Online

City St George's, University of London

Citation: Zhang, D., Wu, Z., Chen, J., Gao, A., Chen, X., Li, P., Wang, Z., Yang, G., Lo, B. & Yang, G-Z. (2020). Automatic Microsurgical Skill Assessment Based on Cross-Domain Transfer Learning. IEEE Robotics and Automation Letters, 5(3), pp. 4148-4155. doi: 10.1109/lra.2020.2989075

This is the accepted version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

Permanent repository link: <https://openaccess.city.ac.uk/id/eprint/38013/>

Link to published version: <https://doi.org/10.1109/lra.2020.2989075>

Copyright and Reuse: Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to [City Research Online policy](#).

Automatic Microsurgical Skill Assessment Based on Cross-Domain Transfer Learning

Dandan Zhang, *Student Member, IEEE*, Zicong Wu, Junhong Chen, Benny Lo 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, it requires a large available database to train the desired model, and the training process can be time-consuming. This paper presents a transfer learning scheme for training with limited RAMS datasets.

An in-house Microsurgical Robot Research Platform Database (MRRPD) is collected from a microsurgical robot research platform (MRRP). It is used for the verification of the proposed cross-domain transfer learning and the pre-trained model for RAMS skill level assessment is fine-tuned after feeding the data obtained from the MRRP. The skill level classification macro accuracy and micro accuracy based on transfer learning can reach 95.6% and 93.8% respectively. The black-box effect of the neural network is eliminated through the visualization of class activation map. Moreover, microsurgical tool tracking is developed to calculate task-specific features and provide visual feedback with intuitiveness while 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

Robotic systems for microsurgery have been developing rapidly, with appreciated benefits of improved accuracy and dexterity. Robot-Assisted Microsurgery (RAMS) also has the benefit of reduced blood loss, fewer complications, shorter operating time, and lower treatment costs. To realise the full potential of RAMS, microsurgical training is important for surgical trainees to acquire sufficient proficiency levels before conducting RAMS on patients.

Thus far, most of the assessment techniques are performed via outcome-based analysis [1], structured checklists, and rating scales [2], [3]. These can be subjective and only provide an overall assessment of the outcome measures.

Emerging platforms for Robot-Assisted Minimally Invasive Surgery (RAMIS) are able to collect multi-modality data from surgical robots or simulators [4], such as vision and kinematic data, which can be used for skill assessment. For example,

JIGSAW(JHU-ISI Gesture and Skill Assessment Working Set) [5] 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. This dataset consists of 39 trials of suturing, 36 trials of knot tying, and 28 trials of needle passing, the data of which was collected via the da Vinci's API software interface. It has been used extensively for surgical gesture segmentation, classification and skill assessment. However, there is no publicly available database for RAMS.

To address the issues due to the lack of sufficient reference datasets, transfer learning can be utilized. It can assist surgical data processing by obtaining knowledge from existing surgical datasets, which helps reduce the burden for data recollection for different surgical tasks based on different types of surgery performing on various surgical robotic platforms.

The main contributions of this paper are listed as follows.

(1) An in-house database is constructed based on a Microsurgical Robot Research Platform (MRRP), which consists of both video data and kinematic data.

(2) Transfer learning is used to realize automatic skill assessment through raw data based on the MRRPD database. The pre-trained model is obtained via the JIGSAWS database and the effectiveness is evaluated before the transfer learning.

(3) Microsurgical tool tracking is used for task-specific features estimation. Visual feedback is provided, which can enable the operators to achieve a higher level of microsurgical skills for RAMS. General evaluation metrics can be automatically calculated and provide explainable results to the operators as a reference.

The details of the proposed framework are described as follows. Firstly, the related work for microsurgical skill assessment is introduced in Section II. Secondly, the system overview is illustrated in Section III, where the microsurgical research platform, the microsurgical tasks and the database information are illustrated. The architecture of the neural network and the proposed transfer learning method is detailed in Section IV, while the visual feedback and other evaluation metric calculation are also described. Finally, conclusions are drawn in Section IV along with the relative merit and potential future improvement of the method.

II. RELATED WORK

In this section, the available microsurgical skill analysis methods are reviewed.

D. Zhang, Z. Wu, J. Chen, B. Lo and G.-Z. Yang are with the Hamlyn Centre for Robotic Surgery, Imperial College London, London, United Kingdom.

G.-Z. Yang is also with the Institute of Medical Robotics, Shanghai Jiao Tong University, China. corresponding author email: d.zhang17@imperial.ac.uk

A. Subjective Methods

The Objective Structured Assessment of Technical Skills (OSATS) [6] is a validated tool that allows scoring the test results, it has been used for the Structured Assessment of Robotic Microsurgical Skills [7]. Some other scoring methods have been proposed, including the Global Rating Scales [8], Structured Assessment of Microsurgery [9], [10] and Modified Global Rating Scales [11].

B. Evaluation Metrics

For objective skill assessment methods, standard [12] and specific evaluation metrics can be calculated based on motion data for the quantitative analysis [13]. The tooltip trajectory, excursion, and velocity has been used to evaluate microvascular anastomosis in [14]. Neurosurgery skill assessment has been conducted based on the tool tip acceleration and angular velocity in a simple pick and place task [15]. Other evaluation metrics have been proposed based on motion data analysis. For example, the variants of trajectory [16], curvature [17], motion jerk value [18]), energy based metrics [19], [20], semantic labels [21] and other extracted features that have high correlations with surgical skills can be utilized as well.

C. Learning Based Methods

With the advent of emerging machine learning techniques, raw motion data can be transformed to intermediate interpretations with advanced feature selection before being fed into learning models as an input for skill assessment. Novel feature fusion has been explored for 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 [22]. Other machine learning algorithms, such as logistic regression [23], Hidden Markov Model [24], Support Vector Machines [20], have been employed for skill assessment as well.

Deep Learning techniques can learn discriminative features efficiently and perform feature extraction progressively to discover abstract representations during the training process via hierarchical layers. This kind of end-to-end learning method has been employed in [25] for skill assessment. In order to decode skill information from raw motion profiles via end-to-end learning, a large amount of data are required.

D. Summary

Subjective methods require an expert surgeon to score the skills of the trainee subjectively, the grading process of which is time-consuming and may be inconsistent due to the biases in human interpretations [26].

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.

For traditional machine learning methods, the skill assessment results depend significantly on the feature extraction process. However, critical information has the potential to be

discarded during the feature extraction, and the complex nature of surgical motion profiles is difficult to be fully explored.

Though deep learning techniques can be utilized for skill assessment with the advantage of automatic feature decoding, the issues of insufficient data limit its widespread applications. Since there are various types of robot-assisted surgery, the data are collected based on different surgical robotic platforms with different types of features. It's difficult to collect large database based on various situation and train a new model using raw data. In this case, transfer learning is promising.

Transfer learning has been widely used in the research area of medical imaging. The work from [27] successfully accomplished the multiple DoF localization of endoscopic capsule robot by a deep neural network, with pre-trained weights from the ImageNet and the tuned network have been evaluated on the physiological model. Similar work has also been illustrated in [28]. However, cross-domain transfer learning has not been explored in the research area of automatic microsurgical skill assessment.

Therefore, an automatic microsurgical skill assessment based on cross-domain transfer learning method is developed in this paper.

III. 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 [26]. 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 Phantom Omni device is used as the master manipulators to perform teleoperation [29]. 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 used in Linux system.
- ROS serves as the middleware for the control of the MRRP, which is convenient for researchers to implement high-level control framework. The ability to integrate with other high-level robot components and developing environments can be ensured.

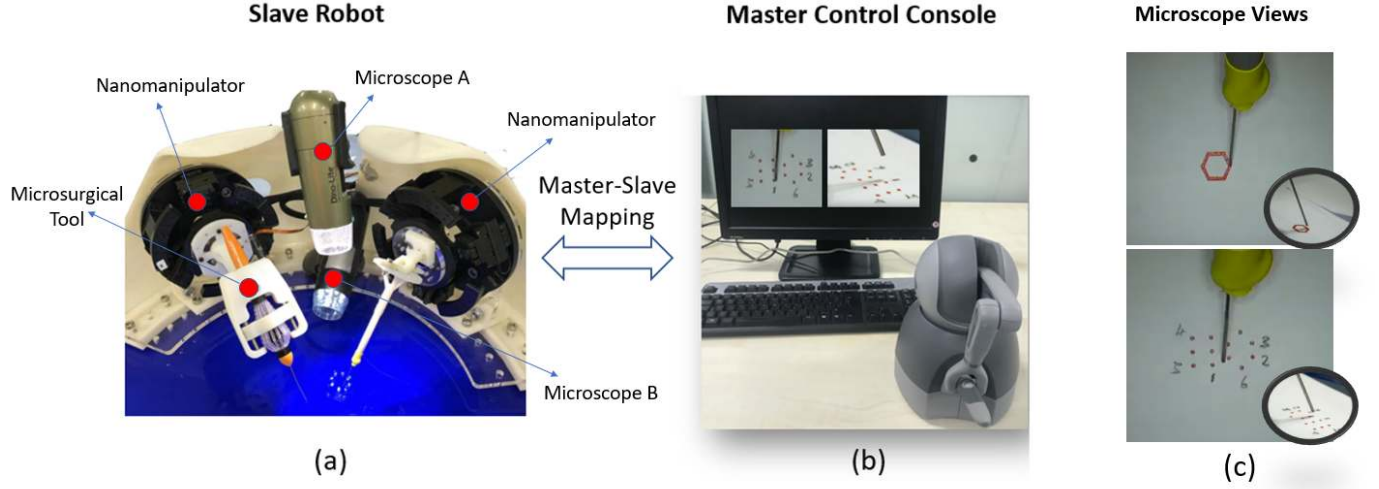


Fig. 1: Overview of the scenario for the master-slave remote control. (a) The MRRP slave robot and visual system. (b) Master control console visualization. The left image on the monitor is the top-to-down view for operational targets visualization; the right image on the monitor is for depth monitoring.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

- The real-time microsurgical robot control performance can be achieved by low-level control.
- A QT5-based GUI is constructed to provide convenience for users to interact with the robot.

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

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 microsurgical procedures.

Although there are many different kinds of microsurgical procedures, the basic requirements remain similar to all of them. Two types of microsurgical tasks are defined as examples, including a positioning task and a path following task.

- **Positioning Task:** The purpose of this task was to examine the ability of the operator in terms of precise targeting. The positioning task requires the subject to place the needle tip within a range 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 was 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 the textile fabric. The operator is asked to place the needle tip as close as possible to the paper or the textile fabric.

The experimental phantom used for microsurgical task operation is shown in Fig. 2 (a). Fig. 2 (b) shows an example

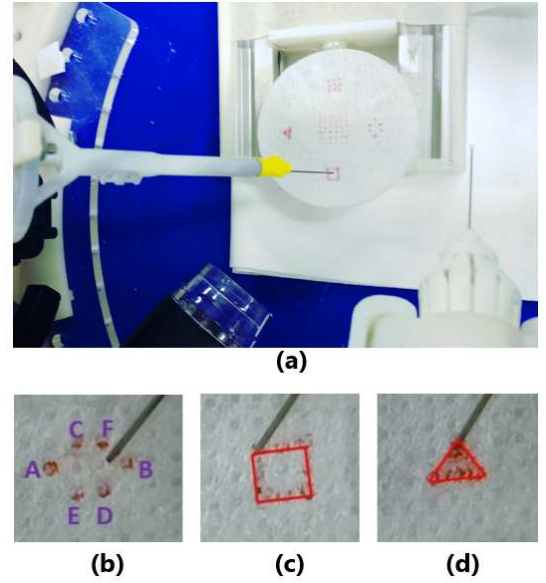


Fig. 2: Experimental scenes for user studies. a) Experimental platform for user studies; b) The positioning task; c) The square-based path following task; d) The triangle-based path following task.

for performing positioning task. The subjects were required to place the micro-needle tooltip in the vertices of a hexagon in a predefined order of A, B, C, D, E, F, A. As for the path following task, the desired trajectory is represented by thick light red lines. Fig. 2(c) and (d) demonstrates two examples of the path following task, which requires the subjects to trace the contour of a square with a side of 8 mm and a triangle with a side of 6mm respectively.

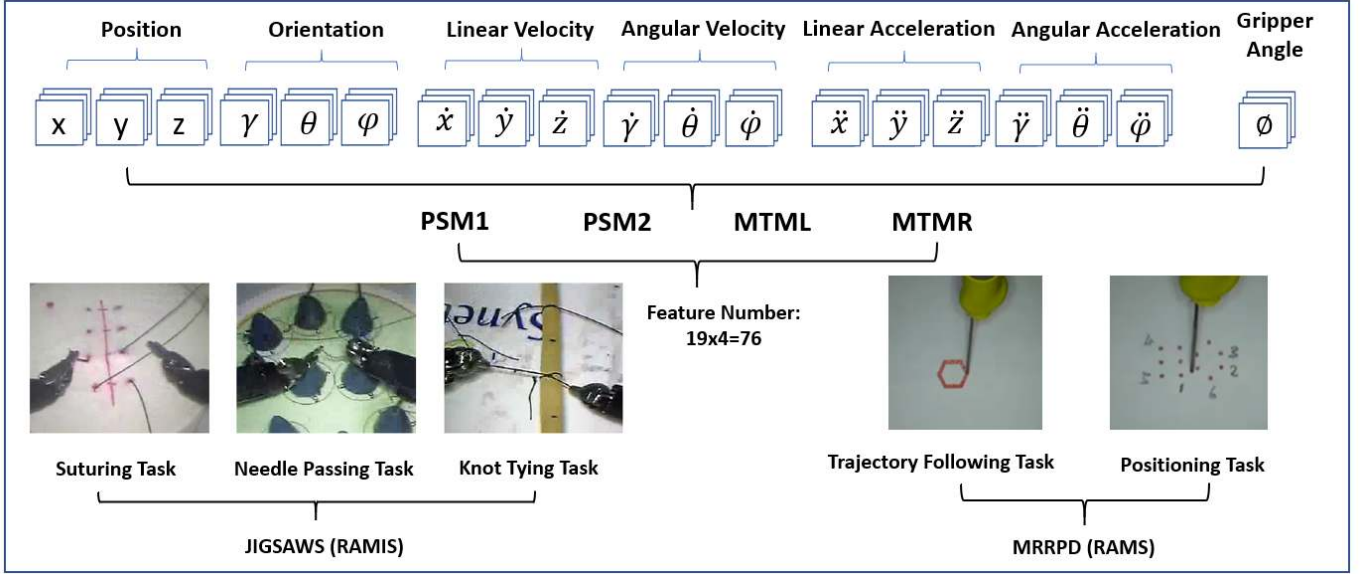


Fig. 3: The re-organized kinematic data structure of the JIGSAWS database and the MRRP.

C. Database

The Microsurgical Robot Research Platform Database (MRRPD) was built up based on 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).

The data collected during microsurgical operation include vision data for the top-to-down view from Microscope A as well as the side view from Microscope B. 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 several 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 get familiarised with the task before the actual experiment. They were asked to go through the whole procedure twice to get accustomed to the experimental protocols. After they felt confident with conducting the experimental assessment, the data collection began. All subjects had a practice session for basic skills acquisition. A total of 24 trials were collected finally.

IV. TRANSFER LEARNING

In this section, the deep neural model architecture is briefly illustrated to demonstrate the effectiveness of the end-to-end learning. Transfer learning method is explored for the skill assessment for RAMS. The relevant evaluation metrics for skill assessments are summarized and served as part of the systematic analysis for microsurge skill level assessment.

A. Overview

Deep surgical skill model can directly process multivariate time series via automatic learning to explore the intrinsic

motion characteristics for decoding skills. Since there is no public available database for RAMS and the database collected by the in-house MRRP is limited, transfer learning and can be considered.

Knowledge transfer is challenging due to the differences in data distribution. A deep neural network model can be trained based on the surgical task for laparoscopic surgery, while the model can be fine-tuned to be adaptive to the database for microsurgical operation.

B. Database for Deep Learning

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. 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 [5].

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 data are obtained and added to each data frame. The 3D position data of the end-effector, the linear and angular velocity and the gripper angle information 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).

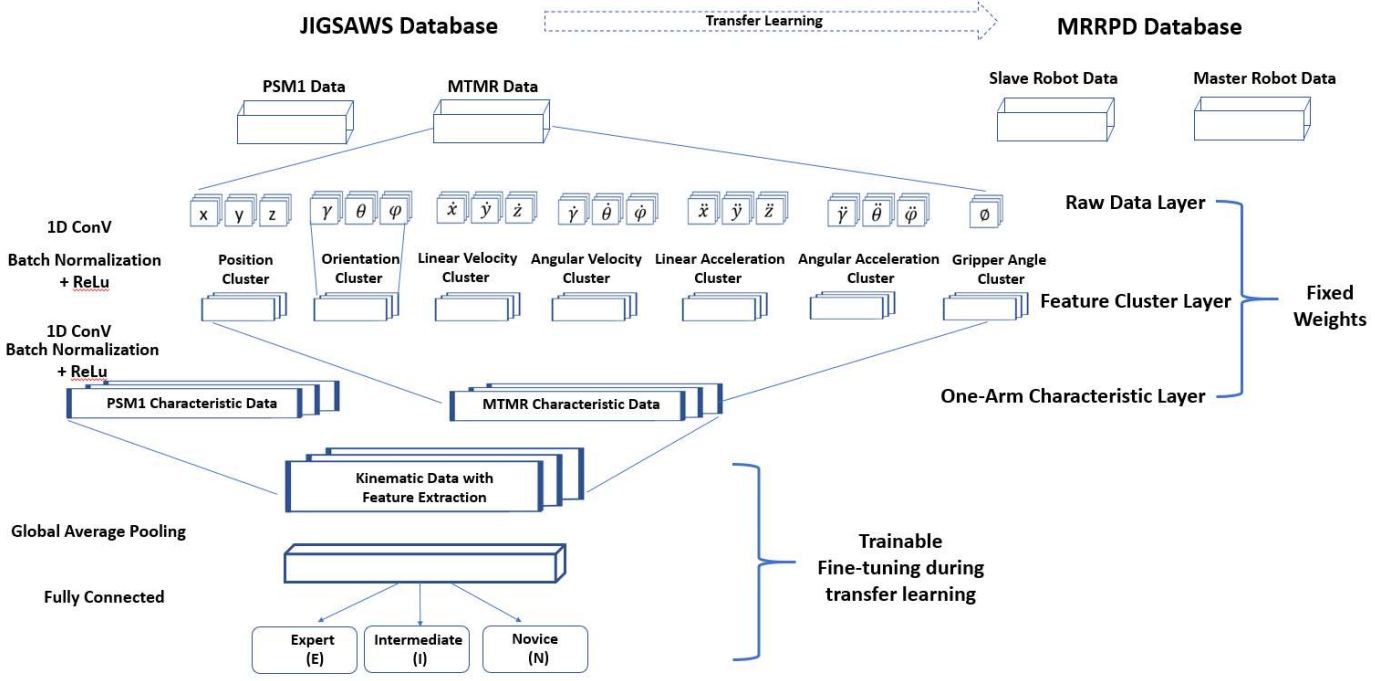


Fig. 4: The architecture for the Deep Neural Network Model and the implementation of transfer learning.

C. 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 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 information, velocity information, acceleration information and gripper information.

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

The multinomial cross-entropy is used as our objective cost function when training the network. As for parameter optimization, Adam is utilized [30]. The architecture for the Deep Neural Network Model training is shown in Fig. 4.

The network can be re-organized based on different configuration of data structure collected during experiments. For example, if no master manipulators are used during experiments, the one-arm characteristic layer merging process can only include 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 acceleration 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 first and second layers 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 preliminary test using the 1D CNN 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, we use 1D CNN in this paper.

D. 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 increase generalization [31].

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 [32], 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.

E. 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 for the Deep Neural Network Model

Macro Average				
Tasks	1 _{st} Test	2 _{nd} Test	3 _{rd} Test	Mean
Suturing	100%	100%	100%	100%
Needle Passing	88.9%	100%	100%	96.3%
Knot Tying	88.9%	91.7%	93.3%	91.3%
RAMIS	100%	97.0%	86.9%	94.6%
Micro Average				
Tasks	1 _{st} Test	2 _{nd} Test	3 _{rd} Test	Mean
Suturing	100%	100%	100%	100%
Needle Passing	83.3%	100%	100%	94.4%
Knot Tying	85.7%	85.7%	87.5%	86.3%
RAMIS	100%	95.5%	85.0%	93.5%

TABLE II: Results for the Transfer Learning

Macro Average				
Transfer Type	1 _{st} Test	2 _{nd} Test	3 _{rd} Test	Mean
Suturing → Path Following	91.7%	100%	83.3%	91.7%
Suturing → Positioning	100%	100%	100%	100%
JIGSAWS → RAMS	100%	95.2%	91.7%	95.6%
Micro Average				
Transfer Type	1 _{st} Test	2 _{nd} Test	3 _{rd} Test	Mean
Suturing → Path Following	87.5%	100%	75.0%	87.5%
Suturing → Positioning	100%	100%	100%	100%
JIGSAWS → MRRPD	100%	93.8%	87.5%	93.8%

We employ the cross-validation for evaluation. In order to remove the outlier, the best and the worst results were abandoned while the mean value is calculated. The learning rate we used here is 0.0001, while 300 epoches are used to train the model.

For each cross-validation setting, the performance of each technique is evaluated by the Micro average accuracy and the Macro average recall 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.

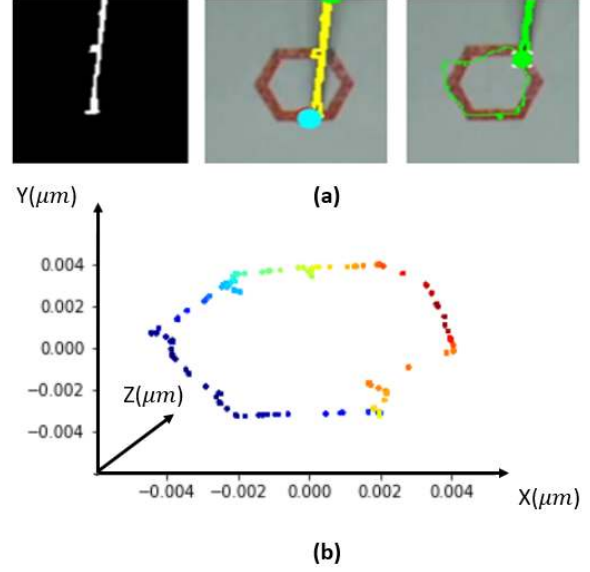


Fig. 5: (a) Visual tracking feedback. (b) Class Activation Map.

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

Table I shows the training results based on the JIGSAWS database. In order to let the neural model learn more features from different surgical tasks, the trials from different surgical tasks are mixed together. 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 acceptable.

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 and the positioning task, the average testing accuracy is 87.5% and 100% respectively. The mean testing accuracy for the transfer learning result from JIGSAWS database to MRRPD is 93.8%. The transfer learning is effective for microsurgical skill assessment.

F. Evaluation with Interpretability

The interpretability is important for the trainee to gain valuable feedback, which guides them to achieve higher level of skills. Class activation map can be implemented through a GAP 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 [33]. By constructing a heatmap from the class activation map, the main reason behind the subject's classification results can be identified.

Fig. 5 (b) 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.

G. Task-Specific Metrics Visualization

Some task-specific features are required to capture adequate information from a specific surgical procedure, while genetic metrics are not sufficient enough for evaluation of the surgical skill for different surgery types [34], [35]. 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 while Root Mean Square Error ($RMSE$) can be used to evaluate the path following task. The definitions of the two evaluation metrics are listed as follows.

- **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 for operation 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 [36].

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

$$\begin{cases} P = \frac{1}{m} \sum_{i=1}^m \min(\|\mathbf{P}_t(i) - \mathbf{P}g(t)\|, (t \in 1 : T)) \\ RMSE = \frac{1}{T} \sum_{t=1}^T \|\mathbf{P}r(t) - \mathbf{P}g(t)\| \end{cases} \quad (2)$$

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 who were involved in the user studies.

Fig. 5 (a) shows the process of visual data processing. The video data were collected through the microscopes. The RGB data of microscopic 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 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 three subjects after the video data processing.

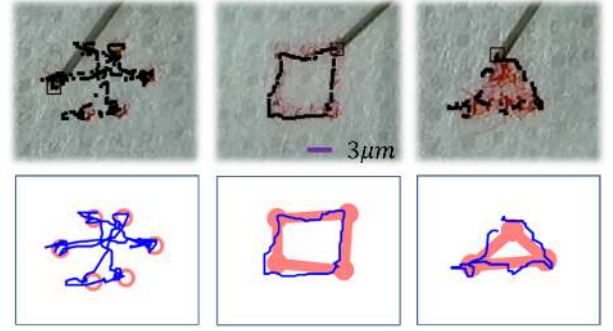


Fig. 6: Vision based real-time tracking results and analysis for a) Positioning of the vertices of a hexagon; b) Path following of a square; c) Path following of a triangle.

Except for task-specific features, other general evaluation metrics can be automatically calculated as a reference. For example, the movement time (T) [37], the operative speed (A_v) [13], instrument traveled length (P_s) [17], the times of the user using the clutching mechanism to reposition the master manipulator (C_f). Table III summarizes the task-specific evaluation metrics for three subjects 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.

TABLE III: User Studies Results

	Positioning Task	Path Following
P_i	674.8	218.5
T_s	84.4	24.2
A_s	8.0 s^{-1}	9.0 s^{-1}
C_f	2	3
P	0.13	/
$RMSE$	/	2.72

H. 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-forceps may need to follow a pre-defined trajectory to reach the targeted operation area, which requires proficient path following skill. Needle insertion and cannulation is one of the important surgical task related to microvascular anastomosis, which requires robotic assistance for targeting. The precision for needle placement is important, which is a kind of positioning task. In the future, more clinical data can be collected in the future to evaluate the effectiveness of the proposed method.

V. CONCLUSIONS

Due to the limitation of the data for RAMS, cross-domain transfer learning is explored and verified in this paper for automatic microsurgical skill level classification. 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 the surgical tasks for laparoscopic surgery. The pre-trained model is tuned and transferred for RAMS application.

The proposed method is validated on the JIGSAWS dataset and achieved very competitive results with 100% accuracy on the suturing and needle passing tasks. The transfer learning method is effective, since the model is trained based on the data of RAMS while the skill level classification layer is fine-tuned after feeding the data obtained from the MRRP. The macro accuracy and micro accuracy of microsurgical skill level classification based on transfer learning can reach 95.6% and 93.8% respectively.

The visualization of the interpretable features are employed as the feedback to trainees by explaining which period of the operation can be used to define the operator as a novice, intermediate or expert based on class activation map. This explains the underlying mechanism of the classification. 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 expert.

VI. ACKNOWLEDGEMENTS

The authors would like to acknowledge all the subjects who contributed the valuable data.

REFERENCES

- [1] B. Ben, A. D. Grayson, J. Mark, B. Nicholas, G. J. Grotte, D. J. M. Keenan, M. Russell, B. M. Fabri, and J. Mark, "Surgeon specific mortality in adult cardiac surgery: comparison between crude and risk stratified data," *Bmj*, vol. 327, no. 7405, pp. 13–17, 2003.
- [2] A. C. Goh, D. W. Goldfarb, J. C. Sander, B. J. Miles, and B. J. Dunkin, "Global evaluative assessment of robotic skills: validation of a clinical assessment tool to measure robotic surgical skills," *J Urol*, vol. 187, no. 1, pp. 247–252, 2012.
- [3] M. A. Aghazadeh, I. S. Jayaratna, A. J. Hung, M. M. Pan, M. M. Desai, I. S. Gill, and A. C. Goh, "External validation of global evaluative assessment of robotic skills (gears)," *Surgical Endoscopy*, vol. 29, no. 11, pp. 3261–6, 2015.
- [4] G. P. Moustris, S. C. Hiridis, K. M. Deliparaschos, and K. M. Konstantinidis, "Evolution of autonomous and semi-autonomous robotic surgical systems: a review of the literature," *International Journal of Medical Robotics + Computer Assisted Surgery Mrcas*, vol. 7, no. 4, pp. 375–392, 2011.
- [5] Y. Gao, S. S. Vedula, C. E. Reiley, N. Ahmidi, B. Varadarajan, H. C. Lin, L. Tao, L. Zappella, B. Béjar, D. D. Yuh *et al.*, "Jhu-isi gesture and skill assessment working set (jigsaws): A surgical activity dataset for human motion modeling," in *MICCAI Workshop: M2CAI*, vol. 3, 2014, p. 3.
- [6] J. Martin, G. Regehr, R. Reznick, H. Macrae, J. Murnaghan, C. Hutchison, and M. Brown, "Objective structured assessment of technical skill (osats) for surgical residents," *British journal of surgery*, vol. 84, no. 2, pp. 273–278, 1997.
- [7] P. A. Liverneaux, S. Hendriks, J. C. Selber, and S. J. Parekattil, "Robotically assisted microsurgery: development of basic skills course," *Archives of plastic surgery*, vol. 40, no. 4, p. 320, 2013.
- [8] P. U. Kalu, J. Atkins, D. Baker, C. J. Green, and P. E. Butler, "How do we assess microsurgical skill?" *Microsurgery: Official Journal of the International Microsurgical Society and the European Federation of Societies for Microsurgery*, vol. 25, no. 1, pp. 25–29, 2005.
- [9] W. Chan, N. Niranjana, and V. Ramakrishnan, "Structured assessment of microsurgery skills in the clinical setting," *Journal of Plastic, Reconstructive & Aesthetic Surgery*, vol. 63, no. 8, pp. 1329–1334, 2010.
- [10] J. C. Selber, E. I. Chang, J. Liu, H. Suami, D. M. Adelman, P. Garvey, M. M. Hanasono, and C. E. Butler, "Tracking the learning curve in microsurgical skill acquisition," *Plastic and reconstructive surgery*, vol. 130, no. 4, p. 550e, 2012.
- [11] D. G. Ezra, R. Aggarwal, M. Michaelides, N. Okhravi, S. Verma, L. Benjamin, P. Bloom, A. Darzi, and P. Sullivan, "Skills acquisition and assessment after a microsurgical skills course for ophthalmology residents," *Ophthalmology*, vol. 116, no. 2, pp. 257–262, 2009.
- [12] A. Steinfeld, T. Fong, D. Kaber, M. Lewis, J. Scholtz, A. Schultz, and M. Goodrich, "Common metrics for human-robot interaction," in *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*. ACM, 2006, pp. 33–40.
- [13] Y. Kassahun, B. Yu, A. T. Tibebe, D. Stoyanov, S. Giannarou, J. H. Metzen, and E. Vander Poorten, "Surgical robotics beyond enhanced dexterity instrumentation: a survey of machine learning techniques and their role in intelligent and autonomous surgical actions," *International journal of computer assisted radiology and surgery*, vol. 11, no. 4, pp. 553–568, 2016.
- [14] P. B. McBeth, D. F. Louw, F. Yang, and G. R. Sutherland, "Quantitative measures of performance in microvascular anastomoses," *Computer Aided Surgery*, vol. 10, no. 3, pp. 173–180, 2005.
- [15] Z. Lin, M. Zecca, S. Sessa, T. Sasaki, K. Suzuki, K. Itoh, H. Iseki, and A. Takanishi, "Objective skill analysis and assessment in neurosurgery by using an ultra-miniaturized inertial measurement unit wb-3—pilot tests—," in *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2009, pp. 2320–2323.
- [16] G. Forestier, F. Petitjean, P. Senin, F. Despinoy, A. Hualmé, H. I. Fawaz, J. Weber, L. Idoumghar, P.-A. Muller, and P. Jannin, "Surgical motion analysis using discriminative interpretable patterns," *Artificial intelligence in medicine*, vol. 91, pp. 3–11, 2018.
- [17] 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, p. 590, 2009.
- [18] 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," *International Journal of Medical Robotics + Computer Assisted Surgery Mrcas*, vol. 14, no. 1, p. e1845, 2017.
- [19] B. Poursartip, M. E. Lebel, R. Patel, M. Naish, and A. L. Trejos, "Analysis of energy-based metrics for laparoscopic skills assessment," *IEEE Transactions on Biomedical Engineering*, vol. PP, no. 99, pp. 1–1, 2017.
- [20] B. Poursartip, M.-E. LeBel, L. McCracken, A. Escoto, R. Patel, M. Naish, and A. Trejos, "Energy-based metrics for arthroscopic skills assessment," *Sensors*, vol. 17, no. 8, p. 1808, 2017.
- [21] M. Ershad, Z. Koesters, R. Rege, and A. Majewicz, *Meaningful Assessment of Surgical Expertise: Semantic Labeling with Data and Crowds*, 2016.
- [22] A. Zia and I. Essa, "Automated surgical skill assessment in rmis training," *International journal of computer assisted radiology and surgery*, vol. 13, no. 5, pp. 731–739, 2018.
- [23] S. S. Vedula, A. Malpani, N. Ahmidi, S. Khudanpur, G. Hager, and C. C. G. Chen, "Task-level vs. segment-level quantitative metrics for surgical skill assessment," *Journal of surgical education*, vol. 73, no. 3, pp. 482–489, 2016.
- [24] L. Tao, E. Elhamifar, S. Khudanpur, G. D. Hager, and R. Vidal, "Sparse hidden markov models for surgical gesture classification and skill evaluation," in *International conference on information processing in computer-assisted interventions*. Springer, 2012, pp. 167–177.
- [25] Z. Wang and A. M. Fey, "Deep learning with convolutional neural network for objective skill evaluation in robot-assisted surgery," 2018.
- [26] 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.
- [27] M. Turan, Y. Almalioglu, E. Konukoglu, and M. Sitti, "A deep learning based 6 degree-of-freedom localization method for endoscopic capsule robots," *arXiv preprint arXiv:1705.05435*, 2017.
- [28] M. Turan, E. P. Ornek, N. Ibrahimli, C. Giracoglu, Y. Almalioglu, M. F. Yanik, and M. Sitti, "Unsupervised odometry and depth learning for

- endoscopic capsule robots,” in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 1801–1807.
- [29] 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,” *International journal of computer assisted radiology and surgery*, pp. 1–11, 2019.
 - [30] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.
 - [31] 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 *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2017, pp. 23–30.
 - [32] S. J. Pan and Q. Yang, “A survey on transfer learning,” *IEEE Transactions on knowledge and data engineering*, vol. 22, no. 10, pp. 1345–1359, 2009.
 - [33] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, “Learning deep features for discriminative localization,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2921–2929.
 - [34] 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,” *International Journal of Medical Robotics + Computer Assisted Surgery Mrcas*, vol. 14, no. 1, 2017.
 - [35] 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,” *Simulation in Healthcare Journal of the Society for Simulation in Healthcare*, vol. 4, no. 4, pp. 104–108, 2009.
 - [36] K. Harada, Y. Minakawa, Y. Baek, Y. Kozuka, S. Sora, A. Morita, N. Sugita, and M. Mitsuishi, “Microsurgical skill assessment: toward skill-based surgical robotic control,” in *International Conference of the IEEE Engineering in Medicine Biology Society*, 2011.
 - [37] 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. 2106–19, 2014.