International Journal of Computer Science, Communication & Information Technology (CSCIT)
Vol.6 pp.49-53

Simple and Efficient Recurrent Neural Network to Evaluate Classified Surgery Tasks

Malik Benmansour[#1], Abed Malti[#2]

[#]*GEE-GEE Department, Tlemcen-Tlemcen University*
13000,Tlemcen, Algeria

[1]malik.benmansour@gmail.com

[2]abed.malti@gmail.com

***Abstract*— In this work, we propose to use recurrent neural network (RNN) architecture to provide a dynamic evaluation of performing surgery tasks. The task is considered to be known and is represented by a sequence of kinematic data recorded from DaVinci Robot. The sequence of output represents a dynamic evaluation which gets updated while the sequence of input is feeded to the RNN. To train the RNN we use three levels of skills: expert, intermediate and novice to which we associate three scores: 1, 0.7 and 0.4 respectively. We test the performance of the proposed method on three different surgical gestures: knot tying, needle passing and suturing. We use one RNN per-gesture which we train with the corresponding data from JIGSAWS; a publicly available dataset of surgery tasks. We compare this approach with a static method that uses Deep Neural Network which provides a global score of the whole surgery task.**

Keywords—surgery task, skill assessment, recurrent neural network.

I. Introduction

During the past few years, many classification approaches have been used to sort out surgical skills in both coarse or fine grained details [1, 2, 3]. Most of this approaches rely on Recurrent Neural Network (RNN) architectures because of the sequential nature of the data. In general, these data are video sequences containing the surgery intervention within organs and surgery tools. When the procedure is executed with a surgical robot, these data can be also kinematic trajectories containing positions, velocities and accelerations of articulated joints: joystick and robotic arm. Despite the importance of surgery task classification, few efforts have been made to automate the evaluation by providing numerical scores on how good or bad a surgery task is achieved.

An automatic and thus objective numerical evaluation of surgery tasks is crucial to help surgereon trainees improving their performance. Moreover it is scalable for continuous evaluation of multiple trainees. It provides the trainees with reliable feedback to improve their surgical skills.

In this paper, we propose to use a simple RNN to provide numerical scores to evaluate the performance of a given surgery task. The input are kinematic trajectories of positions, velocities and accelerations. They correspond to joystick and robotic arm motions from the DaVinci Robot. The provided scores are shaped as a time series of equal length as the input kinematic sequence. With this approach we are able to provide sequential scores with which we compute an average score for the whole task. We are also able to address different input sequence lengths without being obliged to downsample or upsample the sequence to fit a constant input length as it is the case when we used Deep Neural Networks [14, 15]. Our experiment runs on three different surgery tasks: suturing, knot tying and needle passing using the kinematic data and three levels of evaluation: expert, medium and novice.

This paper is organized as follows. First, we present an overview of the state-of-the-art of surgicall skills assessment. Second, we present the proposed setup with the RNN architecture and the training approach. Third, we show our experimental results with discussions and comments. Finally, we draw conclusions and future works on the basis of the obtained results.

II. Related work

The state-of-the-art of surgical skills assessment relies mostly on publicly available dataset. The widely used one is JIGSAWS [1, 5] which contains both kinematic and video data of surgery gestures. During the past five years, artificial neural networks architectures have proven to be more efficient above other machine learning approaches [9]. Most of the proposed approaches addressed mainly classification approaches either for coarse tasks or fine grained sub-tasks. They also differ by relying on surgery videos or on kinematic data. [2] proposed a spatiotemporal RNN-CNN architecture for Fine-grained surgical skills segmentation based on image sequences. [3] utilized a labeled data from real surgery videos to train an accurate and efficient robotic instrument tracker based on the state-of-the- art Hourglass Networks. They evaluated the movement of the robotic instruments and automatically classified the technical level of a surgeon with a linear classifier, using peer evaluations of skill as the reference standard. [4] proposed both tool detection and operative skill assessment using region-based Convolutional Neural Networks. [7] proposed to use both videos and kinematic data

ISSN 1737-930X

International Journal of Computer Science, Communication & Information Technology (CSCIT)
Vol.6 pp.49-53

to classify surgical gestures. [8] suggested to recognize both gestures and longer, higher-level maneuvers with a model of the mapping from kinematics to gestures/maneuvers with recurrent neural networks.

In real surgery applications, several deep network framerworks have been proposed. A. Twinanda et al. [10] used CNN for surgery phase recognition from cholecystectomy videos. The approach carried out both phase recognition and tool presence detection. A similar method was recently used for cataract surgery [11]. Rafii-Tari et al. [12] used catheter-tissue interaction and motion patterns across different skill levels to deliver an automated and objective assessment of performance. [6] proposed to assess skill and provide effective feedback to trainees for unstructured surgical tasks in the operating room, such as tissue dissection in septoplasty.

III. METHODOLOGY

RNNs have been widely used into several sequence modeling of dynamic systems. Speech, text, handwriting are the main contexts where RNNs were originally applied.

Structure: an RNN is composed of a set of inputs $\{x_t\}$, hidden states $\{h_t\}$, outputs $\{m_t\}$ and a nonlinear block. A non-linear block in the RNN takes an input x_t and a hidden state $h_{(t-1)}$ to produce an output m_t and a hidden state h_t which feeds the next nonlinear block with an input $x_{(t+1)}$ and so forth (see figure 1). In the case of RNN with feedback, which our case, the output m_t is back connected to the nonlinear block.

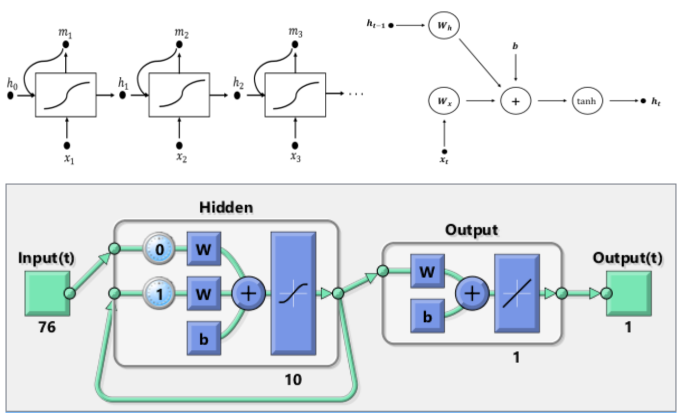


Fig. 1. A single Recurrent Neural Network (RNN) per given surgery task.

The relation involved by the nonlinear block can be written as

$$h_t = tanh(W_x . x_t + W_h . h_{\{t-1\}} + W_m . m_t + b) \quad (1)$$

W_x, W_h and W_m are weights to be learnt. b is the bias. To obtain the output score y_t, we use a linear layer

$$y_t = W_y . m_t + b_y \quad (2)$$

W_y and b_y being also scalars to be learnt.

Inputs: the inputs are kinematic sequences from the JIGSAWS dataset [11]. Each sequence length represents the duration of the task completed by the surgeon. The input is a 76 vector of kinematic data containing positions, velocities and accelerations from the joytsticks and robotic arms of the DaVinci robot [7].

Outputs: The RNN is a single output scalar rating the task between 0 and 1. For every 76 kinematic input, the RNN delivers a score given that the input is part of a sequence of known surgery task.

One RNN per-surgery-task:

In our work we train three different RNNs with the same structure. They are described as follows :

KT-RNN: is dedicated to provide sequences of scores related to the Knot-Tying gesture. It has 76 input vector of kinematic data of Knot-Tying task from any level of knowledge: expert, intermediate and novice subjects. In the learning step three trials from each subject are taken. Two trials of the novice and intermediate subjects are left to for the test step. We refer to this model as "KT-RNN". Figure 2 displays a sample frame from the Knot-Tying.

Fig. 2. A sample frame of the Knot-Tying task.

NP-RNN: is dedicated to provide sequences of scores related to the Needle-Passing gesture. It has 76 input vector of kinematic data of Needle-Passing task from any level of knowledge: expert, intermediate and novice subjects. In the learning step three trials from each subject are taken. Two trials of the novice and intermediate subjects are left to for the test step. We refer to this model as "NP-RNN". Figure 3 displays a sample frame from the Needle-Passing.

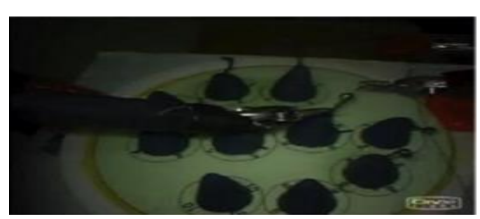


Fig. 3. A sample frame of the Needle-Passing task.

ISSN 1737-930X

International Journal of Computer Science, Communication & Information Technology (CSCIT)
Vol.6 pp.49-53

ST-RNN: this third follows the same principle as the two previous onesand is related to Suturing. We refere to this model as ST-RNN. Figure 4 displays a sample frame from the Suturing gesture.

Fig. 4. A sample frame of the Suturing task.

IV. EXPERIMENTAL RESULTS

In the following section, we will present the setup of this paper but also, we will present shortly two others setups we presented in papers that we established for previous conferences [14, 15]. In these two setups, we used deep neural networks instead of recurrent neural networks to assess performance of the JIGSAWS subjects on the three tasks. Details are shown in the following subsection.

A. Experimental setup and compared methods

RNNs setup: We used three independent recurrent neural networks. Each one is intended to assess all subjects of the JIGSAWS dataset on one task only. Therefore, each network takes as input, learning kinematic data of one of the three tasks. Each RNN contains 10 hidden layers. The training finishes after 50 epochs. The Levenberg-Marqaurdt algorithm is employed for the training process and the performance is calculated through the Mean Squared Error. The following figures depict the architecture of our three RNNs.

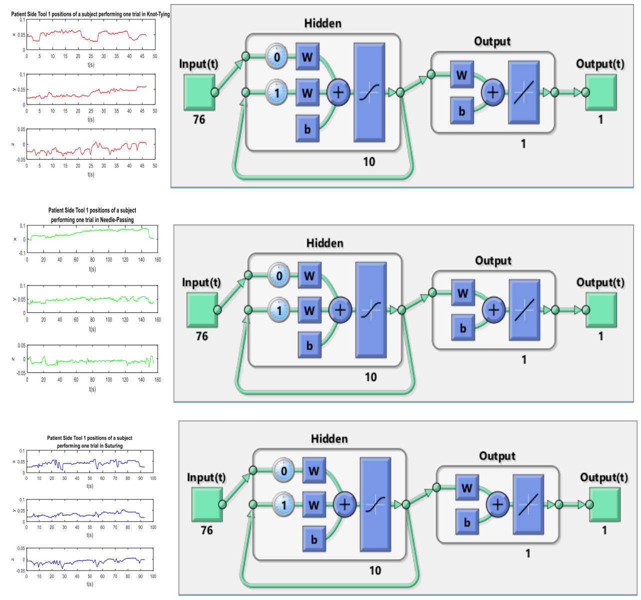


Fig. 5. Architechture of the proposed RNNs for specific task evaluation.

Remark : On the previous figure, each row represents the architecture of a RNN intended to assess one of the three tasks. On the left of each row is shown the evolution of the spatial positions of one Patient Side Tool (which is a component of the DaVinci surgery robot) while the task is performed by a surgeon.

Compared methods: The setups established in [14, 15] use deep neural networks. In [14], the goal was to design a deep neural network containing three hidden layers with 20 nodes per hidden layer. This network was trained with kinematic data of expert subjects of the JIGSAWS dataset performing the three tasks and next, kinematic data of intermediate and novice subjects were chosen to test the network. This network was intended to assess objectively the subjects on the three tasks at once. On the other hand, in [15], we established three independent neural networks and each one is responsible to assess subjects on one task only. Thus, each network is trained with kinematic data of some trials of all subjects of the JIGSAWS dataset, which means novice, intermediate and exepert subject and the test is completed using kinematic data of the remaining trials. Please refer to [1] for more details about the JIGSAWS content. Each of the three networks is composed of three hidden layers with 20 nodes per hidden layer. As it can be noticed through its description, the model of this setup is globally similar to the RNNs setup but we have used deep neural networks instead of recurrent networks.

B. Surgical skill assessment and disscussion

We run the assessment procedure on the test data. It contains two intermediates and four novices. The intermediate subjects are labeled with the letters C and F and the novice subjects are labeled with the letters B, G, H and I. For the setup of [15] and for the RNNs, we give for each network,as

ISSN 1737-930X

test data, one trial of each subject performing the task the network is responsible to assess. For the model of the setup in [14], the test data contains one trial of each intermediate and novice subject performing each task.

The following tables show the scores of each novice and intermediate subject after performing the three tasks. For more convenience, we abbreviate the names of the tasks Knot-Tying, Needle-Passing and Suturing respectively by KT, NP and ST. We labeled any task performed by any subject by the abbreviation of the task concatenated with the label representing the subject followed by the number of the trial. For example, KTB4 stands for "novice subject B performing Knot-Tying task, trial 4".For the 1st table, in the "Rate" rows, the green colored cells show the test results of the model of the setup used in [14] while the yellow colored cells show the test results of the three independent models of the setup used in [15]. In the 2nd table are shown the results of RNNs setup.

The results prove that the RNNs yielded the best performance evaluation. Indeed, we can see that particularly, the two intermediate subjects C and F got better ratings than all the novice subjects B,G,H and I with the recurrent neural networks than with the deep neural networks. These results are plausible since intermediate subjects have more time experience in the domain than the novice subjects.

TABLE I
SCORES OF NOVICE AND INTERMEDIATE SUBJECTS AFTER ASSESSING THEIR PERFORMANCE THROUGH THE SETUPS USING DEEP NEURAL NETWORKS [14, 15]

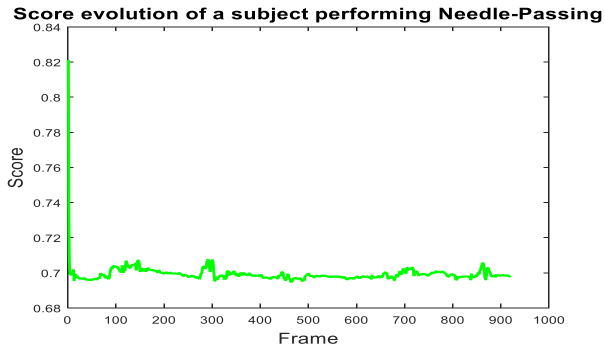
Subject and task	**KTB4**		**KTG4**		**KTH4**		**KTI5**		**KTC4**		**KTF4**	
Rate	60.8%	39.00%	98.6%	36.5%	58.2%	39.9%	0.8%	50.8%	98.9%	87.7%	85.2%	50.9%
Subject and task	**NPB4**		**NPG4**		**NPH4**		**NPI4**		**NPC4**		**NPF4**	
Rate	99.7%	43.58%	32.9%	50.0%	99.9%	31.1%	99.8%	33.7%	32.5%	100%	60.3%	64.4%
Subject and task	**STB4**		**STG4**		**STH4**		**STI4**		**STC4**		**STF4**	
Rate	0.05%	26.93%	0.01%	46.8%	0.00%	27.3%	3.8%	34.5%	0.02%	78.2%	27.3%	45.2%

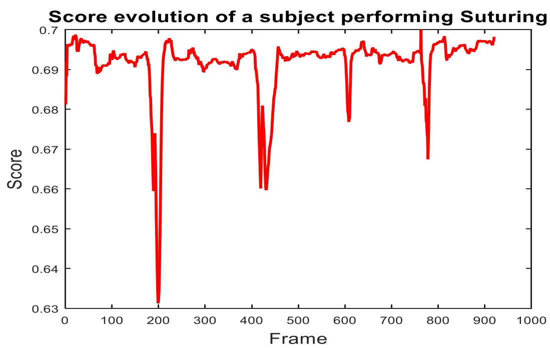
TABLE III
SCORES OF NOVICE AND INTERMEDIATE SUBJECTS AFTER ASSESSING THEIR PERFORMANCE THROUGH THE RNNS.

Subject and task	**KTB4**	**KTG4**	**KTH4**	**KTI5**	**KTC4**	**KTF4**
Rate	40.1%	39.9%	40.2%	39.7%	99.2%	69.7%
Subject and task	**NPB4**	**NPG4**	**NPH4**	**NPI4**	**NPC4**	**NPF4**
Rate	40.1%	40.0%	39.9%	40.5%	70.3%	66.5%
Subject and task	**STB4**	**STG4**	**STH4**	**STI4**	**STC4**	**STF4**
Rate	39.9%	40.0%	40.7%	40.0%	69.3%	59.6%

As additional results, here are some figures representing the score evolution of the intermediate subject F performing each task. The "x" axis represents the duration (in frames) of the trial while the "y" axis represents the score at each frame.

ISSN 1737-930X





V. CONCLUSION

In this paper, we presented an approach with the purpose of surgical skills evaluation. This method consists on assigning three independent recurrent neural networks that were trained on all subjects performing three corresponding gestures. The training scores for assessment were assigned according to the subject experience. Thus, each network is responsible on rating the performance of one task. We compared this approach with other approches that use deep neural netwroks We have experimentally shown that the proposed approach yields better and more logical assessment scores based in the fact that intermediate subjects have more experience than the novice subjects in the surgery domain. As future work, we aim to design Long Short Term Memory networks that will be able to learn from sequences of images that show the gestures composing each task and see if we will get any better results.

REFERENCES

[1] Yixin Gao, S. Swaroop Vedula, Carol E. Reiley, and Narges Ahmidi, “JIGSAWS: The JHU-ISI Gesture and Skill Assessment Working Set,” *CIRL*, 04-Sep-2014. [Online]. Available: https://cirl.lcsr.jhu.edu/research/hmm/datasets/jigsaws_release/.

[2] C. Lea, A. Reiter, R. Vidal, and G. D. Hager, “Segmental Spatiotemporal CNNs for Fine-grained Action Segmentation,” *ArXiv160202995 Cs*, Feb. 2016.

[3] H. Law, K. Ghani, and J. Deng, “Surgeon Technical Skill Assessment using Computer Vision based Analysis,” in *Machine Learning for Healthcare Conference*, 2017, pp. 88–99.

[4] A. Jin *et al.*, “Tool Detection and Operative Skill Assessment in Surgical Videos Using Region-Based Convolutional Neural Networks,” *ArXiv180208774 Cs*, Feb. 2018.

[5] N. Ahmidi *et al.*, “A Dataset and Benchmarks for Segmentation and Recognition of Gestures in Robotic Surgery,” *IEEE Trans. Biomed. Eng.*, vol. 64, no. 9, pp. 2025–2041, 2017.

[6] H. Doughty, D. Damen, and W. Mayol-Cuevas, “Who’s Better? Who’s Best? Pairwise Deep Ranking for Skill Determination,” *ArXiv170309913 Cs*, Mar. 2017.

[7] N. Ahmidi *et al.*, “Automated objective surgical skill assessment in the operating room from unstructured tool motion in septoplasty,” *Int. J. Comput. Assist. Radiol. Surg.*, vol. 10, no. 6, pp. 981–991, Jun. 2015.

[8] L. Zappella, B. Béjar, G. Hager, and R. Vidal, “Surgical gesture classification from video and kinematic data,” *Med. Image Anal.*, vol. 17, no. 7, pp. 732–745, Oct. 2013.

[9] R. DiPietro *et al.*, “Recognizing Surgical Activities with Recurrent Neural Networks,” in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*, 2016, pp. 551–558.

[10] L. Sbernini, L. R. Quitadamo, F. Riillo, N. D. Lorenzo, A. L. Gaspari, and G. Saggio, “Sensory-Glove-Based Open Surgery Skill Evaluation,” *IEEE Trans. Hum.-Mach. Syst.*, vol. 48, no. 2, pp. 213–218, Apr. 2018.

[11] A. P. Twinanda, S. Shehata, D. Mutter, J. Marescaux, M. Mathelin and N. Padoy, “EndoNet: A Deep Architecture for Recognition Tasks on Laparoscopic Videos ” *IEEE Trans. on medical imaging*, vol. 36, no. 1, pp. 86-97.

[12] K. Charrière, G. Quellec, M. Lamard, D. Martiano, G. Cazuguel, G. Coatrix and B. Cochener, “Real-time analysis of cataract surgery videos using statistical models ” *Multimedia tools and application*”, vol. 76, no. 21, pp. 22473-22491.

[13] H. Rafii-Tari, C. J. Payne, J. Liu, C. Riga, C. Bicknell and G-Z. Yang, ”Towards automated surgical skill evaluation of endovascular catheterization tasks based on force and motion signatures”, IEEE International Conference of Robotics and Automation – *ICRA 2015*, 2015, pp. 1789-1794.

[14] Malik Benmansour, Wahida Handouzi, Abed Malti, ”A neural network architecture for automatic and objective surgical skill assessment”, IEEE, The International Conference on Electrical Sciences and Technologies in Maghreb – *CISTEM 2018.*

[15] Malik Benmansour, Wahida Handouzi, Abed Malti, ”Task-Specific Surgical Skill Assessment With Neural Networks”, Advanced Intelligent Systems for Sustainable Development – *AI2SD’2018.*

ISSN 1737-930X