SCIENCE CHINA
Information Sciences

• MOOP •

August 2022, Vol. 65 184201:1–184201:2
https://doi.org/10.1007/s11432-021-3363-0

Learning ultrasound scanning skills from human demonstrations

Xutian DENG[1], Ziwei LEI[1], Yi WANG[1], Wen CHENG[1], Zhao GUO[1], Chenguang YANG[3] & Miao LI[1,2*]

[1]*School of Power and Mechanical Engineering, Wuhan University, Wuhan* 430072*, China;*
[2]*The Institute of Technological Sciences, Wuhan University, Wuhan* 430072*, China;*
[2]*School of Automation Science and Engineering, South China University of Technology, Guangzhou* 510640*, China*

Received 17 April 2021/Revised 30 July 2021/Accepted 8 October 2021/Published online 27 June 2022

Citation Deng X T, Lei Z W, Wang Y, et al. Learning ultrasound scanning skills from human demonstrations. Sci China Inf Sci, 2022, 65(8): 184201, https://doi.org/10.1007/s11432-021-3363-0

Recently, the robotic ultrasound system has become an emerging topic owing to the widespread use of medical ultrasound. According to the level of the system autonomy, robotic ultrasound systems can be categorized into three levels: teleoperated, semiautonomous and full-autonomous. Related work to fulfill ultrasound scanning guidance could be divided into three stages in autonomous ultrasound systems according to the dimensions of the input information: (1) manually planning trajectory (human guidance), (2) ultrasound image evaluation model with proper servo controller (ultrasound image), (3) learning part of ultrasound scanning skills (ultrasound image, probe pose). Although studies [1–4] have achieved certain results, the modeling method of ultrasound scanning skills needs to be improved. In this study, focusing on extracorporeal ultrasound, we not only consider ultrasound images and probe pose, but also encode the contact force between the probe and humans into ultrasound scanning guidance, which is regarded as a critical factor during ultrasound scanning. The main contribution of this study is twofold. (1) A multimodal model of ultrasound scanning skills is proposed and trained from human demonstrations, which considers the ultrasound images, probe pose and contact force. (2) A sampling-based strategy is proposed with the learned model to adjust the extracorporeal ultrasound scanning process to obtain a high-quality ultrasound image. Note that the primary goal of this article is to offer a learning-based framework to understand and acquire extracorporeal ultrasound scanning skills from human demonstrations. However, the learned model can apparently be applied to a robot system, which is our future work.

Research target. Our goal is to learn freehand ultrasound scanning skills from human demonstrations (see Appendix A). Furthermore, we aim to evaluate the multimodal task quality of combining multiple sensory information, including ultrasound images, contact force and probe pose (as shown in the left part of Figure 1(a)), to extract the skill from the task representation and even to transfer the skill across tasks. The details of the method are described in the following.

Learning of ultrasound task representation. For a freehand ultrasound scanning task, we propose a domain-specific encoder, as shown in Figure 1(a). Three types of sensory feedback are available, namely, ultrasound images, probe pose, and contact force. We use a deep neural network to encapsulate the heterogeneous nature of these sensory data. Convolutional layers are used to extract the feature vector of an ultrasound image, and fully connected layers are used to extract the feature vector of pose and force. The resulting two feature vectors are concatenated, and further, yield a task feature vector. The model for multimodal task representation is a neural network whose parameters are denoted by Ω_θ. The training process is described in the following.

Data collection via human demonstrations. The multimodal model, as shown in Figure 1(a), has several learnable parameters. Therefore, we design a procedure to collect the ultrasound scanning data from human demonstrations for the training data. A novel probe holder is designed with intrinsically mounted IMU and F/T sensors (see Appendix B). The collected data are described as follows:

- $D = \{(S_t, P_t, F_t)\}_{t=1,\ldots,N}$ denotes a dataset with N observations.
- $S_t \in \mathbb{R}^{224\times224\times3}$ denotes the collected ultrasound image with cropped size at time t.
- $P_t \in \mathbb{R}^4$ denotes the probe pose in terms of quaternion at time t.
- $F_t \in \mathbb{R}^6$ denotes the contact force/torque between the probe and the human skin at time t.

Three sonographers evaluated the quality of the obtained ultrasound image for each piece of recorded data in the dataset D and labeled with 1/0 ('good'/'bad'). The neural network model Ω_θ is trained with a crossentropy loss function. During training, we minimize the loss function with

* Corresponding author (email: miao.li@whu.edu.cn)

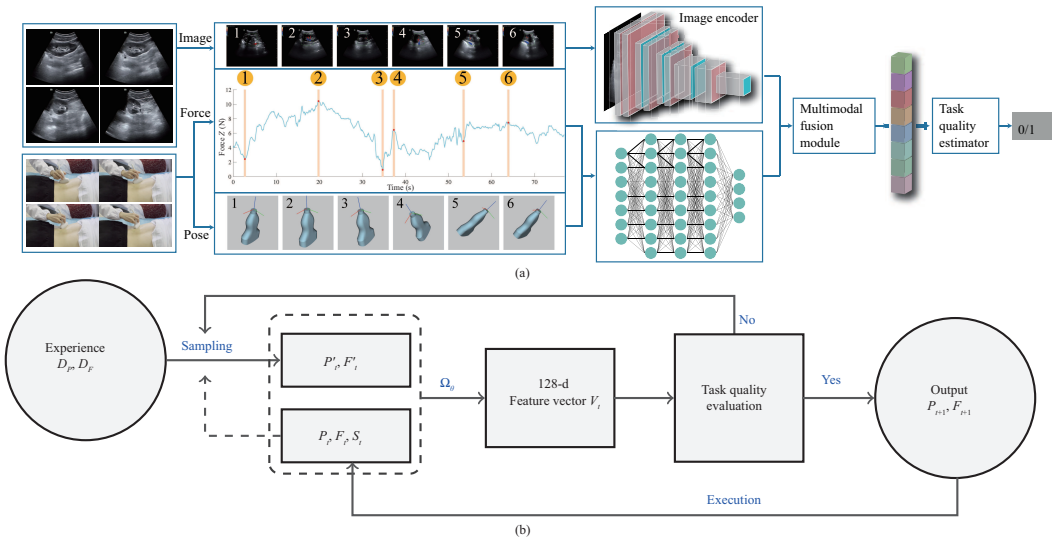

Figure 1 (Color online) (a) The multimodal task learning architecture with human annotation; (b) the sampling-based online adaption strategy.

stochastic gradient descent. As a result, the trained network could evaluate the quality of tasks. Given the task representation model Ω_θ, an online adaptation strategy is proposed to improve the task quality by leveraging the multimodal sensory feedback, as discussed in the following.

Ultrasound skill learning. We devise the online adaptation policy as follows:

$$P_{t+1}, F_{t+1} = \arg\max_{P_t^{'}, F_t^{'}} (Q_1 + Q_2), \tag{1}$$

$$Q_1 = f_{\Omega_\theta}\left(S_t, P_t^{'}, F_t^{'} | P_t^{'} \in D_P, F_t^{'} \in D_F\right), \tag{2}$$

$$Q_2 = 1 - f_{\mathrm{RMSE}}\left(\langle P_t, F_t\rangle, \langle P_t^{'}, F_t^{'}\rangle | P_t^{'} \in D_P, F_t^{'} \in D_F\right), \tag{3}$$

where f_{Ω_θ} denotes the predicted quality, which is evaluated using the learned neural network model Ω_θ, f_{RMSE} denotes the root mean squared error of two vectors, Q_1 and Q_2 denote two weights of the task quality, $F_t^{'}$ and $P_t^{'}$ denote the randomly sampled results from collected dataset at time t, and D_P and D_F denote two feasible sets of pose and force, respectively. Here, these two feasible sets are determined via human demonstrations. However, notably, other task-specific constraints for the pose and contact force can also be adopted here.

This model-free policy does not require prior knowledge of the ultrasound scanning process dynamics, namely the transition probabilities from one state to another (from current to next ultrasound image). We choose the Monte Carlo policy for optimization [5], where the potential actions are sampled and selected directly from previous demonstrated experience, as shown in Figure 1(b). The predicted quality from the learned neural network model Ω_θ is the most intuitive evaluation result. Considering multiple results of (2), we impose a bound between $P_t^{'}$, $F_t^{'}$ and P_t, F_t, as shown in (3), which prevents the next state from moving too far away from the current state. If the new state $\langle S_t, P_t^{'}, F_t^{'}\rangle$ is evaluated as suitable, the desired pose $P_t^{'}$ and contact force $F_t^{'}$ are used as the goal for the human ultrasound scanning guidance. Otherwise, new $P_t^{'}$ and $F_t^{'}$ are sampled from previous demonstrated experience. This process repeats $\mathcal{N}$ times, and $P_t^{'}$ and $F_t^{'}$ with the best task quality are chosen as the final goal for the human scanning guidance.

Conclusion. This article presents a framework for learning ultrasound scanning skills from human demonstrations. We summarize ultrasound images, probe pose and contact force into a learnable model. Further, a sampling-based strategy is proposed to guide the extracorporeal ultrasound scanning process, based on the learned model. Finally, we have designed some experiments for verification (see Appendixes A–D). Experimental results show that this framework for ultrasound scanning guidance is robust. This work will be applied to an existing robot system in the future.

Acknowledgements This work was supported by Suzhou Key Industrial Technology Innovation Project (Grant No. SYG202121) and Natural Science Foundation of Jiangsu Province (Grant No. BK20180235).

Supporting information Videos and Appendixes A–D. The supporting information is available online at info.scichina.com and link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

References

1 Karamalis A, Wein W, Klein T, et al. Ultrasound confidence maps using random walks. Med Image Anal, 2012, 16: 1101–1112
2 Chatelain P, Krupa A, Navab N. Confidence-driven control of an ultrasound probe: target-specific acoustic window optimization. In: Proceedings of the 2016 IEEE International Conference on Robotics and Automation, Stockholm, 2016. 3441–3446
3 Virga S, Zettinig O, Esposito M, et al. Automatic force-compliant robotic ultrasound screening of abdominal aortic aneurysms. In: Proceedings of the 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems, Daejeon, 2016. 508–513
4 Droste R, Drukker L, Papageorghiou A T, et al. Automatic probe movement guidance for freehand obstetric ultrasound. In: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, 2020. 583–592
5 Sutton R S, Barto A G. Reinforcement Learning: An Introduction. Cambridge: MIT Press, 2018