A Novel Method for Knot-Tying in Autonomous Robotic assisted Surgery Using Deep Learning

Zhenning Zhou\textsuperscript{1,a}, Xueying Zu\textsuperscript{1,b}, Qixin Cao\textsuperscript{1,c} and Xiaoxiao Zhu\textsuperscript{1,d*}

\textsuperscript{1}School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai, China
\textsuperscript{a}email: Zhenning_zhou@sjtu.edu.cn
\textsuperscript{d*}email: ttl@sjtu.edu.cn

Abstract. With the advent of robot-assisted surgery, surgical task automation, which means that robotic assistants could autonomously execute certain commonly occurring tasks, is more and more appealing and has been studied over the last several years because of the booming of deep learning. Mainly, such partial automation can help reduce the surgeon's workload and allow surgeons to focus more on critical elements of the surgical workflow. In this paper, we propose a novel method using deep learning based on Variational Autoencoders for robotic assistants to learn knot-tying with the Data Set of manual operation, instead of learning from video demonstrations. Taking the circle action of knot-tying as an example, we make use of the VAE network to conduct feature learning and autonomous generation of knot-tying trajectories. During this time, the appropriate VAE network is built and implemented training, and after 100 rounds of training, experimental results show that we successfully acquire trajectories as expected using smaller Data Set with VAE network.

1. Introduction

The rapidly increased use of surgery-assisted robots provides more and more methods for surgeons to cooperate, access and accomplish surgeries. Apparently, in the future of surgery, robotic assistants will be supposed to have the capacity to autonomously perform certain commonly occurring tasks because such partial automation can reduce the surgeon's workload and allow surgeons to focus more on critical elements of the surgical workflow. In addition, after collecting Data Set of target performance [1], intelligent robotic surgical assistants, as van den Berg states, which could “improve patient health by enhancing surgeon performance, reduce medical errors, and reduce costs by reducing operation time” [2]. As a consequence, enabling robotic assistants to learn from a large body of human demonstration and perform certain common tasks, such as knot-tying, is of great value.

Actually, from the last century, modeling human movement and learning human skills have been studied in a large amount of literature using hidden Markov models [3], fuzzy sets [4], as well as neural networks [5]. In recent years, there is more expectation on the automation of specific surgical subtasks performed by robots, and in the meantime, some research results have demonstrated the potential of autonomous robots to improve the efficacy, consistency and accessibility of surgical techniques [6-8]. Mayer et al. implemented automation of manual tasks for minimally invasive surgery in terms of robot learning to tie knots [9]. An automatic vision guided knot-tying method for robotic-assisted surgery is proposed in this paper [10], whose experimental results show that autonomous knotting is likely to perform faster than humans. Hermann Mayer et al. exploited more recurrent and powerful neural networks (RNNs) with adaptive internal states [11]. Ziheng Wang et al. provided an analytical deep
learning framework in surgical for training skill assessment [12], which means a lot. Except for automation of tasks, researches on surgical gesture recognition, action segmentation and imitation are also appealing and at the same time, have acquired some remarkable results [13-15].

Here, in order to make the operation easier and improve the efficiency of surgeons, we propose a novel method, which can help robotic assistants to learn tying knots with trajectory Data Set using deep learning based on Variational Autoencoders. Moreover, we take the circle knot as an example, as shown in Fig 1, demonstrate the feasibility and validity of this method with training results.

Fig. 1 The knot-tying operation

2. Methods
In order to obtain the Cartesian trajectory of surgical instrument, the knot-tying operation can be carried out by manual teleoperation, and the track coordinates of instrument’s end can be recorded at the same time. However, there is a problem with this method that the trajectory of each manual telecontrol is different, as illustrated in Fig. 2. Three knot-tying trajectories generated by manual teleoperation are different, but all of them can complete the corresponding tasks, which means that it is impossible to establish an index to measure which trajectory is the best. Therefore, it is inappropriate to choose any trajectory executed manually as the trajectory of robot motion planning. However, by observing the trajectories in Fig. 2, it could be easily found that manually operated trajectories of the winding action in the knot-tying task are different each time, but there are common characteristics behind them.

Fig. 2 The trajectory of teleoperation

2.1. Selection of Generative Networks
In this paper, we adopt a method based on deep learning to learn the manually executed trajectory and obtain the low-dimensional features. On this basis, the automatic generation of trajectory is achieved through the generation network. Among those deep generative models, two major families stand out and deserve a special attention: Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs). However, compared to GAN, the amount of data required by VAE is smaller, the training takes less time, and the networks is more stable. In conclusion, we adopt VAE to automatically generate trajectories.

2.2. Principles of VAE network
The principle block diagram of the Variational Autoencoders is shown in Fig. 3. The input of VAE is the original real sample $X$, which is encoded by the encoder, to achieve low-dimensional feature extraction. In order to make the variational autoencoder to be essentially a generative model rather than a discriminant model, the extracted low-dimensional feature implicit vector $z$ must conform to some prior distribution $P(z)$, which can be artificially specified, such as the simplest uniform distribution or
normal distribution. In order to obtain the final generated samples, the implicit vector $z$ conforming to the prior distribution is input into the decoder network, which represents the conditional probability distribution $P(x|z)$. The final sample $\tilde{x}$ is generated by sampling the joint probability distribution $P(x, z) = P(z)P(x|z)$ and the purpose of the variational autoencoder is to make the samples generated by the decoder as realistic as possible. Although the encoder and decoder are involved in the training process, only the decoder is finally applied after the network training is completed.

![Fig. 3 Schematic diagram of VAE](image)

From the view of probability, we can consider that the distribution of the Data Set is sampled from $P(x|z)$, where $z$ is a hidden low-dimensional feature variable, represents the data internally and is in accord with the known prior distribution $P(z)$. When the hidden variable $z$ is known in problem, and the network has learned the law of Data Set that means $P(x|z)$, we can generate qualified samples, which have common features of $z$.

2.3. The structure of VAE network
In this study, the prior distribution $P(z)$ is set as the normal distribution $\mathcal{N}(0,1)$. Under this circumstance, $P(z)$ is known, and the problem is transformed into a probabilistic model $P(z)$ that is expected to be generated by network learning. In the meantime, we use maximum likelihood estimation method here: the more excellent network, the probability of real samples generated should be greater. So if we parameterize the decoder model $P(x|z)$ with parameter $\Theta$, the optimization target of the network will be:

$$\max_{\Theta} p(x) = \int P(z)P(x|z)dz$$

The idea of variation is used here that we can approximate distribution $q_\Phi(z|x)$ with distribution $P(z|x)$ to minimize the distance between $q_\Phi(z|x)$ and $P(z|x)$

$$\min_{\Phi} D_{KL}(q_\Phi(z|x)||P(z|x))$$

where $D_{KL}$ represents KL divergence, is used to calculate the distance between two probability distributions, can be defined as

$$D_{KL}(q_\Phi(z|x)||P(z|x)) = \int q_\Phi(z|x) \log \frac{q_\Phi(z|x)}{P(z|x)} dz$$

and if we define that

$$-\int q_\Phi(z|x) \log \frac{q_\Phi(z|x)}{P(x,z)} dz = L(\Phi, \Theta)$$

the above equation will become
\[ D_{KL}(q(z|x)||P(x|z)) = -L(\Phi, \theta) + \log P(x) \]

This is a fact that KL divergence is equal or greater than zero, and the maximum likelihood probability of network target \( P(x) \) could be converted to maximize the \( L(\Phi, \theta) \)

\[
L(\Phi, \theta) = \int q_\phi(z|x) \log \frac{P_\theta(x, z)}{q_\phi(z|x)} \\
= \int q_\phi(z|x) \log \frac{P(z)}{q_\phi(z|x)} + \int q_\phi(z|x) \log P_\theta(x|z) \\
= D_{KL}(q_\phi(z|x)||P(z)) + \mathbb{E}_{z \sim q_\phi(z|x)} [\log P_\theta(x|z)]
\]

As a result, the encoder and decoder network can be parameterized as \( q_\phi(z|x) \) and \( \log P_\theta(x|z) \), respectively. Calculate the KL divergence, then the loss function constructed by likelihood probability \( \log P_\theta(x|z) \) of decoder and the network will be optimized. However, the input of decoder is sampled from \( \mathbb{N}(\mu, \sigma) \) and the gradient propagation is discontinuous, so this gradient descent algorithm cannot be used to optimize the network. For this problem, we adopt Reparameterization Trick method that we implement sampling of implicit variables by \( z = \mu + \sigma \) to achieve the target of network optimization.

According to the principle of VAE, we build our own VAE network as shown in the Fig. 4, where the input of network is the manual trajectory collected in advance. The trajectory data from different dimensions are trained after segmented, and finally the data generated by each dimension channel is merged and restored into a complete trajectory. The input data is extracted with low-dimensional features through two-layer RNN network. Afterwards, these extracted low-dimensional features pass through two separate fully connected networks FC1 and FC2 to obtain the hidden variable \( z \), then the hidden variable \( z \) is generated through the two-layer RNN network and fully connected layer to generate the required samples.

**3. Results & Discussion**

In order to verify the effect of the Variational Autoencoders proposed in this paper for trajectory generation, we build the network for training based on TensorFlow2.0.

The Data Set still adopts the winding action in the knot-tying track mentioned in the last section, as shown in Fig. 5, which shows the process of network training. It can be seen that the generated trajectories are chaotic at the beginning of the training, then the outline can be basically found after 20-30 rounds of training. However, after 100 rounds of training, the generated trajectories have a high similarity with the manual trajectory, which can completely replace the latter, as shown in Fig. 6.
Meanwhile, the change trend of loss during training is demonstrated in Fig. 7. It may be easily found that the convergence rate of VAE training is awfully such fast and task demand is basically satisfied after 100 rounds of training. The advantages of VAE network over GAN network are highlighted that smaller Data Set is required and training process is more stable.

4. Conclusions
Surgical task automation is more and more appealing with the booming of deep learning. Unlike other research on the surgical automation learning from video demonstrations, we propose a novel method which can learn to tie knots with trajectory Data Set using deep learning based on Variational Autoencoders, in order to make the operation easier and improve the efficiency of surgeons.

In this paper, we take the circle action as an example, and the VAE network is firstly used to conduct feature learning and autonomous generation of circle trajectories. We build appropriate VAE network and implement training, after 100 rounds of training, the experimental results show that we successfully obtain trajectories as expected with VAE network. In addition, advantages of VAE are fully demonstrated because of smaller Data Set required in training process and more stable performance.

Acknowledgments
The work is funded by the foundation from the National Key R&D Program of China (2019YFC1711800).
References

[1] D. Nagy T, Haidegger T. (2019) A DVRK-based framework for surgical subtask automation. Acta Polytechnica Hungarica, 16: 61–78.

[2] J. van den Berg, S. Miller, D. Duckworth, H. Hu, A. Wan, X.-Y. Fu, K. Goldberg, and P. Abbeel. 2010. Superhuman performance of surgical tasks by robots using iterative learning from human-guided demonstrations. In: ICRA. Anchorage. 2074–2081

[3] K. Itabashi, K. Hayakawa, T. Suzuki, S. Okuma, F. Fujiwara. 1997. Modeling of the peg-in-hole task based on impedance parameters and HMM. In: Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems. 451-457.

[4] M. Skubic, R. Volz. (2000) Acquiring robust, force based assembly skills from human demonstration. IEEE Trans. on Robotics and Automation, 16(6):772–781.

[5] M. Kaiser, R. Dillmann. 1996. Building elementary robot skills from human demonstration. In: Proc. IEEE Int. Conf. on Robotics and Automation. 2700-2705.

[6] Azad Shademan, Ryan S. Decker, Justin D. Opfermann (2020) Supervised autonomous robotic soft tissue surgery. Science Translational Medicine, 8: 337ra64.

[7] Chow D.L, Newman W. 2013. Improved Knot-Tying Methods for Autonomous Robot Surgery. In: IEEE International Conference on Automation Science and Engineering. Wisconsin. 461-465.

[8] Moustris G.P, Hiridis S.C, Deliparaschos K.M, Konstantinidis K.M. (2011) Evolution of autonomous and semi-autonomous robotic surgical systems: a review of the literature. Int J Med Robotics Comput Assist Surg, 7: 375–392.

[9] H. Mayer, I. Nagy, D. Burschka, A. Knoll, E. Braun, R. Lange, R. Bauernschmitt. 2008. Automation of manual tasks for minimally invasive surgery. In: Int. Conf. on Autonomic and Autonomous Systems. 260-265.

[10] Chow D.L, Russell C. Jackson, Wyatt Newman. 2014. A Novel Vision Guided Knot-Tying Method for Autonomous Robotic Surgery. In: IEEE International Conference on Automation Science and Engineering, Taipei. 504-508.

[11] Mayer M, Gomez F, Wierstra D, Nagy I. 2006. A System for Robotic Heart Surgery that Learns to Tie Knots Using Recurrent Neural Networks. In: International Conference on Intelligent Robots and Systems. Beijing. 543-548.

[12] Wang Z.H, Fey A.M (2018) Deep learning with convolutional neural network for objective skill evaluation in robot-assisted surgery. International Journal of Computer Assisted Radiology and Surgery, 13:1959–1970.

[13] Tanwani A.K, Sermanet P, Yan A. 2020. Motion2Vec: Semi-Supervised Representation Learning from Surgical Videos. In: ICRA.

[14] Amsterdam B.V, Clarkson M.J. 2020. Multi-Task Recurrent Neural Network for Surgical Gesture Recognition and Progress Prediction. In: ICRA. 1380-1386.

[15] K. Fang, T. Wu, D. Yang, S. Savarese, and J. J. Lim. 2018. Demo2vec: Reasoning object affordances from online videos.In: IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2139–2147.