International Journal of Computer Assisted Radiology and Surgery (2020) 15:341–349
https://doi.org/10.1007/s11548-019-02098-7

ORIGINAL ARTICLE

Robot-assisted flexible needle insertion using universal distributional deep reinforcement learning

Xiaoyu Tan[1] · Yonggu Lee[1] · Chin-Boon Chng[1] · Kah-Bin Lim[1] · Chee-Kong Chui[1]

Received: 4 September 2019 / Accepted: 17 November 2019 / Published online: 25 November 2019

Abstract

Purpose Flexible needle insertion is an important minimally invasive surgery approach for biopsy and radio-frequency ablation. This approach can minimize intraoperative trauma and improve postoperative recovery. We propose a new path planning framework using multi-goal deep reinforcement learning to overcome the difficulties in uncertain needle–tissue interactions and enhance the robustness of robot-assisted insertion process.

Methods This framework utilizes a new algorithm called universal distributional Q-learning (UDQL) to learn a stable steering policy and perform risk management by visualizing the learned Q-value distribution. To further improve the robustness, universal value function approximation is leveraged in the training process of UDQL to maximize generalization and connect to diagnosis by adapting fast re-planning and transfer learning.

Results Computer simulation and phantom experimental results show our proposed framework can securely steer flexible needles with high insertion accuracy and robustness. The framework also improves robustness by providing distribution information to clinicians for diagnosis and decision making during surgery.

Conclusions Compared with previous methods, the proposed framework can perform multi-target needle insertion through single insertion point qunder continuous state space model with higher accuracy and robustness.

Keywords Deep learning · Deep reinforcement learning · Needle steering · Tool–tissue interaction · Uncertainty

Introduction

Minimally invasive surgery (MIS) involves numerous surgical techniques which significantly limit the incision size and subsequently reduce the risk of infection, blood loss, and wound healing time. Therefore, MIS is applicable for a wide range of patients who are not suitable for open surgery [1–3]. Needle insertion is one of the most crucial clinical approach in MIS for biopsy and radiology treatment. To further reduce trauma, flexible needles are utilized and steered toward the treatment area to avoid puncturing critical organs and tissues. However, similar to other percutaneous approaches, the accuracy and robustness of flexible needle insertion are limited by multiple barriers including tissue deformation, diversity of tissue properties, and uncertainty in tissue–needle interaction [4]. The intraoperative accuracy and robustness directly contribute to treatment outcomes and postoperative recovery. Hence, there is a need to develop a robust path planning framework to further improve the robustness in obstacle avoidance and risk management under the complexity and uncertainty in needle–tissue interaction.

This paper introduced a new path planning framework to perform robot-assisted flexible needle insertion. In this framework, a new algorithm called universal distributional Q-learning (UDQL) is developed to learn the stable steering policy based on a CT reconstructed simulation model and display the learned Q-value distribution (i.e., distribution of expected return) for further risk management by surgeons. The universal value function approximation (UVFA) [5] is

✉ Xiaoyu Tan
xiaoyu_tan@u.nus.edu

Yonggu Lee
yonggu.lee@nus.edu.sg

Chin-Boon Chng
chinboon@nus.edu.sg

Kah-Bin Lim
limkahbin@nus.edu.sg

Chee-Kong Chui
mpecck@nus.edu.sg

[1] Department of Mechanical Engineering, National University of Singapore, Singapore 117575, Singapore

utilized in the training process of the UDQL agent to achieve multi-goal deep reinforcement learning (DRL) which enables the agent to perform multiple targets insertion through one single insertion point (SIP) with only one training process. Compared with our previous work [6], the proposed path planning framework can steer the flexible needle in 3-dimensional (3D) model under continuous state space. The surgeons decision could be involved in the model updates and transfer learning to generate a patient-specific treatment program without complicated reward engineering (e.g., compare the value distribution on different SIPs). Based on the computer simulation and experimental results, the proposed path planning framework can achieve high accuracy and robustness in steering flexible needle insertion on both simulation and robot platform [7].

Related works

Needle insertion techniques have been widely utilized to achieve biopsy, brachytherapy, anesthesia, and other MIS percutaneous therapies [4,8]. However, the insertion procedures were usually applied on nonhomogeneous soft tissue by physicians with only kinesthetic feedback from the tools. Therefore, the precision and treatment effect were highly related to the individual experience. Due to the steep learning curve of studying MIS and prolonged preoperative preparation, applying needle insertion related MIS in clinical treatment is challenging. To overcome these issues, several surgical robotic systems were developed to perform the robot-assisted needle insertion [3,7,9,10]. The insertion precision and robustness were reported to be improved due to the application of highly accurate and reliable robotic mechanisms.

To further reduce the intraoperative trauma, flexible needles were utilized in robot-assisted needle insertion steering with various path planning algorithms [6,11–13]. These methods were introduced to perform planning on preoperative off-line medical images (e.g., CT and MRI) due to its clear segmentation and accurate registration. The versatility of path planning algorithms is also enhanced due to the separation of medical imaging and planning. These methods utilized Markov decision processes (MDPs) and dynamic programming (DP) to consider the inherent uncertainty of flexible needle–tissue interaction as transition probabilities to produce steering policies [6,12,13].

However, these methods used discrete 2D models and described the uncertainty in numerical representation (e.g., single transition probability), making it difficult to perform multiple needle insertion tasks in 3D surgical scenarios and assisting the physician in understanding intuitively the uncertainty and the risk of the generated policy. In this paper, we introduce the UDQL path planning algorithm which utilizes multi-goal distributional DRL method to perform path planning on multi-target 3D work space through SIP. The distribution of expected return could be directly accessed for risk management. This work is a direct follow-up of traditional MDPs methods [12,13] and our previous work using robust MDPs method [6] in discrete 2D work space.

Methods

In this section, the main approaches used in our proposed path planning framework are presented, including the proposed UDQL algorithm, risk management, model updates, and transfer learning. First, a high-level overview of our framework is presented in section "Framework architecture".

Framework architecture

The flowchart shown in Fig. 1 demonstrates the major process of implementing the proposed path planning system with UDQL, including model construction, UDQL agent training, risk management, and fast re-planning with transfer learning. However, the proposed UDQL algorithm could adapt other work flow with different clinical procedures.

In this paper, we utilized segmented CT images to construct simulation environment. The CT images could be segmented by either physicians or other existing image processing methods [14,15]. Since the CT image segmentation process is not related to our main contribution, segmented images from our collaborating physicians are utilized in constructing the simulation. UDQL could achieve multiple-target needle insertion on SIP by one well-trained agent. The physicians and surgeons could observe the Q-value distribution at any position of the generated trajectory which indicate the expected return distribution. Normally, single

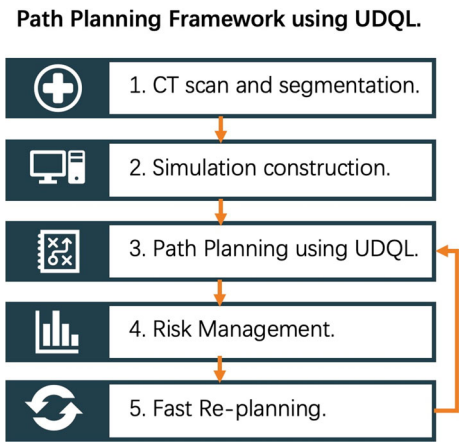


Fig. 1 The flowchart of our proposed path planning system using UDQL

modal distribution with high mean and low variance would be considered as a low-risk plan. Consequently, the SIP and target positions could be updated by surgeons to minimize the risk.

Universal distributional *Q*-learning

Reinforcement learning (RL) is utilized to find the optimal control policy (i.e., a sequence of actions) in MDPs and navigate the agent to achieve the best cumulative reward in an uncertain environment [16]. In combination with deep learning (DL) [17], DRL can adapt to the partially observed MDPs problems and find the optimal control policy in high-dimensional and complex environment [18,19].

Considering a normal RL setup constituted with MDPs, the RL agent will interact with Environment E, with action $a_t \in A$, on state $s_t \in S$, at time step t. At the next time step, the agent would approach to the new state s_{t+1} due to the transition probability $p(s_{t+1}|s_t, a_t)$. We represent the unpredictable needle–tissue interaction as transition probability, because it naturally indicate the stochasticity of next needle tip position. This definition is based on the previous works [6,13] and is supported by studies in needle–tissue interaction [20,21].

For each time step, the agent will perform an action from a parameterized deterministic policy $\mu_\theta(s_t)$ and obtain a reward through a reward function $r(s_t, a_t)$. The objective of training a RL agent is to find the θ that can maximize the expected total return $\mathbb{E}_\tau(R(\tau))$ along the trajectory $\tau \sim \mu_\theta$. Normally, the $R(\tau)$ is a discounted return with discount factor γ. The action-value function $Q^\mu(s, a)$ is the estimation of future return following the policy μ which takes the action a on state s, which could be achieved by solving the Bellman equation:

$$Q^\mu(s_t, a_t) = \mathbb{E}[r(s_t, a_t) + \gamma \sum_{s_{t+1}} p(s_{t+1}|s_t, a_t) \times Q^\mu(s_{t+1}, \mu(s_{t+1}))]. \quad (1)$$

To observe the Q-value distribution, we replace the study on expectation value $Q(s, a)$ by considering the value distribution $Z(s, a)$ which represents the stochastic nature of future return. The value distribution evaluation $Z^\mu(s, a)$ by policy μ could be evaluated by the distributional Bellman equation recursively:

$$Z^\mu(s_t, a_t) \overset{D}{=} R(s_t, a_t) + \gamma Z^\mu(s_{t+1}, \mu(s_{t+1})) \quad (2)$$

which R is a reward function characterized by value distribution Z and $\overset{D}{=}$ represents "equal in distribution" [22,23]. Normally, the value distribution Z is modeled by categorical distribution along the value interval $[V_{\min}, V_{\max}]$ with the set of atoms $\{z_i = V_{\min} + i\Delta z : 0 \leq i < N\}$ and $\Delta z := (V_{\max} - V_{\min})/(N - 1)$. Therefore, the distribution $Z_\theta(s, a) = z_i$ could be estimated by parametric approximator θ (e.g., deep neural networks) on given state s and action a with the probability:

$$p_i(s, a) := \frac{e^{\theta_i(s,a)}}{\sum_j e^{\theta_j(s,a)}}. \quad (3)$$

Compared with our previous work [6], the distributional Q-learning framework can consider the randomness caused by transition uncertainty and represent the Q-value following its stochastic nature [23]. The multimodality of value distribution could be also accurately represented in the approximation. Subsequently, we leverage UVFA and follow the definition in the hindsight experience replay (HER) [24] algorithm to solve the sparse reward problem and achieve the multiple target planning through one agent.

In our setup, we concatenate different goals on the state observation $Z(s||g, a)$ as a richer state representation which $||$ denotes the concatenation. The goal should be a mapping $m : S \to G$ s.t. $\forall_{s \in S} f_{m(s)} = 1$ with goal predicate $f_g : S \to \{0, 1\}$. HER implements a heuristic re-sampling method which replaces the goal g with other states s_t as the new goals for all transitions on state $s \in \{s_0, s_1, \ldots, s_{t-1}\}$. Following the idea of HER, we proposed UDQL algorithm which solves the multi-goal RL problem with distributional Q-learning. To further improve the sample efficiency, we also implemented the prioritized experience replay (PER) [25] in UDQL. This method preforms importance sampling which treats the TD error as prioritization. Compared with the original HER algorithm framework, the UDQL agent could initialize a larger target area which is beneficial for effectively gathering successful trajectories in exploration. The data efficiency and generalization could be also improved by utilizing PER and distributional Q-learning. The algorithm of UDQL is shown in Algorithm 1.

Risk management and transfer learning

After the training, the optimal action $a^\star$ could be obtained by the optimal policy $\pi^\star$. The value distribution along the trajectory could be observed through inference $Z_\theta(s, a^\star, g)$ and validated by medical professionals based on patient-specific conditions. Since the value distribution can represent the uncertain needle–tissue interaction, the risk of intraoperative implementation could be explicitly evaluated. By utilizing the value distribution, other risk-sensitive RL algorithms could be also implemented to assist the preoperative planning [22,26].

Based on the planning evaluation from surgeons and physicians, the model could then be manually updated and retrained on an updated simulation. Exploiting the UVFA,

Algorithm 1 Universal Distributional Q-learning with Prioritized Experience Replay

Require: Labeled targets set G
Initialize function approximator θ, PER replay buffer E_p, and local experience replay buffer E_l
for episode $= 1, M$ **do**
  for $g \in G$ **do**
    for $t = 0, T-1$ **do**
      $a_t \leftarrow \mu_\theta(s_t, g)$,$r_t := r(s_t, a_t, g)$
      Implement a_t and observe s_{t+1}
      Store (s_t, a_t, r_t, s_{t+1}) in E_l
    end for
    for All transitions in E_l **do**
      Set the priority as maximum p_{imax}
      Store $(s_t||g, a_t, r_t, s_{t+1}||g)$ in E_p
      if $f_g(s_T) = 1$ **then**
        Store $(s_t||s_T, a_t, r_t, s_{t+1}||s_T)$ in E_p
      end if
    end for
  end for
  for $t = 1, K$ **do**
    Sample a minibatch B from E_p
    for Transition in B **do**
      Calculate the Q-value distribution target $\boldsymbol{m}$ using (2) and (3)
    end for
    Optimize the cross-entropy loss with target $\boldsymbol{m}$
    Update the priority of B with δ_{ce} in E_p
  end for
end for

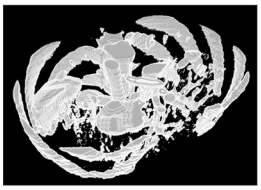
the value distribution could be continuously evaluated by the approximator $Z_\theta(s, a, g)$. Hence, the agent could perform fast transfer learning to adapt the updated SIP (initial position of simulation) and target set G.

Experiments

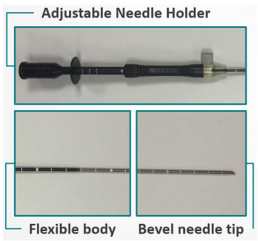
Experiment design and tasks setup

To fully test the proposed path planning framework using UDQL, we designed three experiments to evaluate its performance in simulation and phantom experiment. The first two experiments were performed on 2D and 3D models to compare the path planning performance with our previous method [6] and test both its accuracy and robustness in new 3D simulation scenarios. The simulation environment was constructed from real patient CT images and segmented by our medical collaborators. The third experiment was implemented on a prototype RFA surgical robot [7]. In this experiment, the UDQL agent needed to execute a flexible RFA needle insertion on a phantom model following all procedures described in Fig. 1.

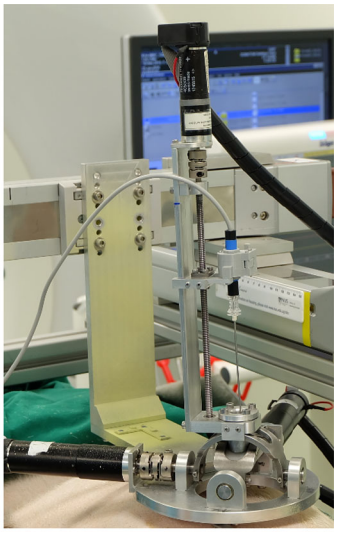
The prototype surgical robot shown in Fig. 2a, c is the end effector of our prototype image-guided RFA surgical system from our previous work [7,27–29]. Flexible RITA needle (RITA®, StarBurst™, USA) shown in Fig. 2b is used in our experiments which is also clinically utilized in our collaborating hospital. The 3D model acquired by CT data reconstruction is shown in Fig. 2a, c.



(a) 3D model for experiment



(b) RITA flexible needle



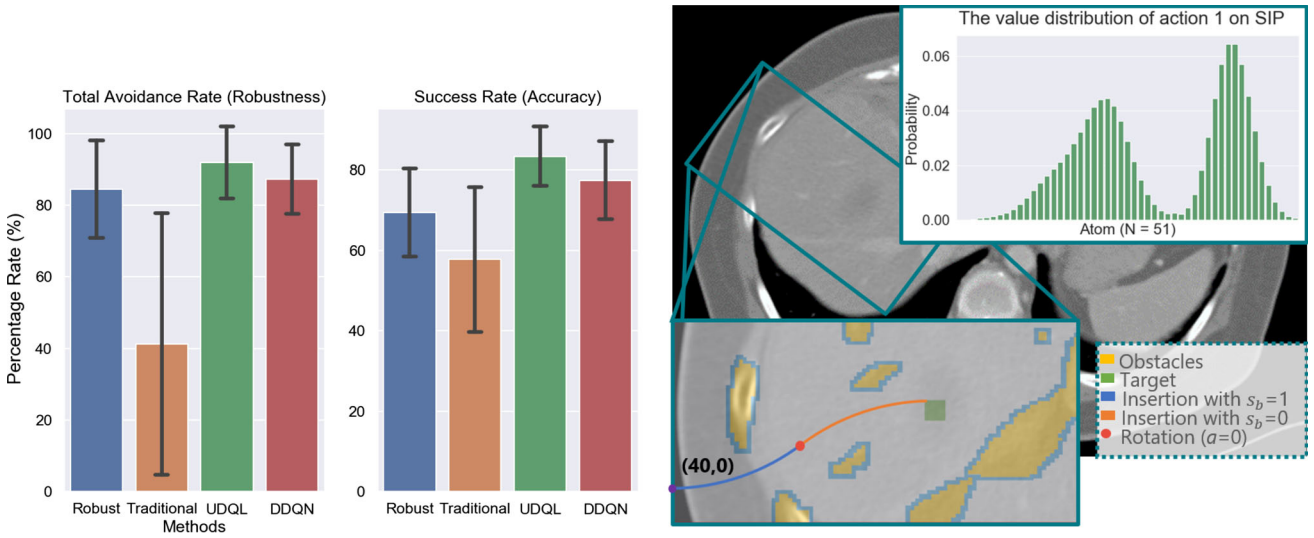
(c) Prototype needle insertion robot

Fig. 2 Flexible RFA needle path planning in liver tumor RFA surgery: simulation environment, flexible needle, and prototype surgical robot

2D RFA simulation with CT data

In this experiment, we compared the UDQL path planning algorithm with robust MDPs algorithm [6], traditional MPDs path planning algorithm [12], and deep double Q-learning network (DDQN) agent [30]. We follow the simulation experiment in the previous studies [6,12]. Following our previous study [6], the work space for robust MDPs algorithm and traditional MDPs algorithm was discretized on a 2D rectangle area with single target. However, for our proposed UDQL algorithm, we can directly perform planning on continuous state space.

The state definition follows our previous work [6] which could be fully characterized by needle tip position on X-axis and Y-axis, direction angle, and bevel direction. Hence, the state dimension is four. The reward function for traditional and robust MDPs was similar to the reward engineering in previous work [6] with five individual cost values on different conditions. Our proposed UDQL agent was trained on sparse reward function $r(s_t, a_t, g) = [f_g(s_{t+1}) = 0] + C[f_g(s_{t+1}) = 1]$ with $C = 50$ which avoids complicated reward engineering. The success rate P_s and the total avoidance ratio P_a are defined following our previous work [6]. The network architectures for UDQL agent and DDQN agent used in this experiment are three layers deep neural networks with 128 units on each layer. We tested the algorithm on all



(a) The barplot for P_s and P_a.

(b) Trajectory demonstration.

Fig. 3 **a** The barplot for insertion on all available SIPs with mean and standard deviation. "Robust" and "Traditional" indicate robust MDPs algorithm and traditional MDPs algorithm, respectively. **b** A demonstration of UDQL generated path on 2D CT-based simulation with value distribution on SIP (40, 0) and CT segmentation

available SIPs in the points set P_{SIP} from previous study [6]. A barplot is demonstrated in Fig. 3a to show the mean and standard deviation for each method across all available points in P_{SIP}. Our proposed UDQL algorithm achieves the highest mean avoidance rate and success rate with smallest variance which shows a superior robustness and accuracy in simulation.

To demonstrate the statistical significance of our work, we perform two sets with six groups statistical analysis (student's t test) on P_s and P_a which are similar to the previous study [6]. For the first test, we evaluate the UDQL and DDQN methods to show the significance of the proposed method compared with the DRL baseline. For the second test, we evaluate the UDQL algorithms and robust algorithm to demonstrate the improvement based on our previous work. In three conditions, we propose null hypothesis based on the mean of two populations: $H_0 = \mu_u - \mu_b = 0$ where μ_u denotes the mean of UDQL algorithm indicators and μ_b denotes the mean of baseline algorithms. Normally, the alpha level a of t test is set to 0.05 to indicate the significance of the rejection. In our test, we proposed a strict alpha level $a = 0.005$ to indicate the significance of the improvement.

Condition 1 *For all SIPs* $p_{\text{SIP}} \in P_{\text{SIP}}$.

Condition 2 $\{P_{au} > 80.0\%$ *and* $P_{su} > 70.0\%\}$ *or* $\{P_{ab} > 80.0\%$ *and* $P_{sb} > 70.0\%\}$.

Condition 3 $\{P_{au} > 80.0\%$ *and* $P_{su} > 70.0\%\}$ *and* $\{P_{ab} > 80.0\%$ *and* $P_{sb} > 70.0\%\}$.

In Condition 1, we analyze the results on all available SIPs. For the second condition, we filter the points which can perform an acceptable needle insertion steered by both algorithms. P_{au} and P_{su} denote the avoidance rate and success rate of UDQL algorithm and P_{ab} and P_{sb} indicate the performance of baseline methods, respectively. Condition 3 analyses on the points that achieved high P_a and P_s in baseline methods.

The two sets statistical analysis results are given in the Table 1. In both tests, the p value [31] is smaller than the default alpha value a for P_a and P_s in each condition. Hence, the null hypothesis H_0 is rejected and the results demonstrate the significance of UDQL algorithm on alpha level $a = 0.005$ and confidence level 99.5%. Our proposed path planning UDQL algorithm achieved higher accuracy and robustness compared with robust MDP and DDQN agent. The improvement in UDQL algorithm is statistical significant. The value distribution of SIP (40, 0) with insertion action $a = 1$ shown in Fig. 3b also demonstrates the capability of approximating the multimodality of the distribution which is important in risk management, because it represents the stochastic nature in transition uncertainty.

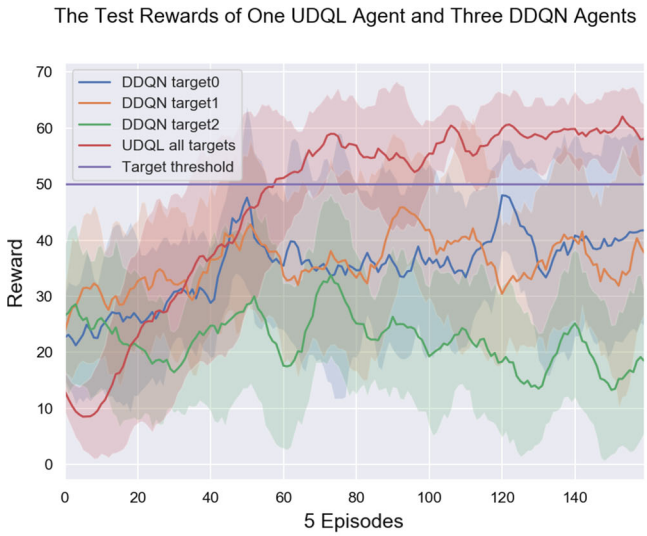
3D RFA simulation with CT data

Based on the results shown in section "2D RFA simulation with CT data", we have demonstrated that the improvement in both accuracy and robustness is statistical significant in 2D simulation. In this section, we test the feasibility of the UDQL

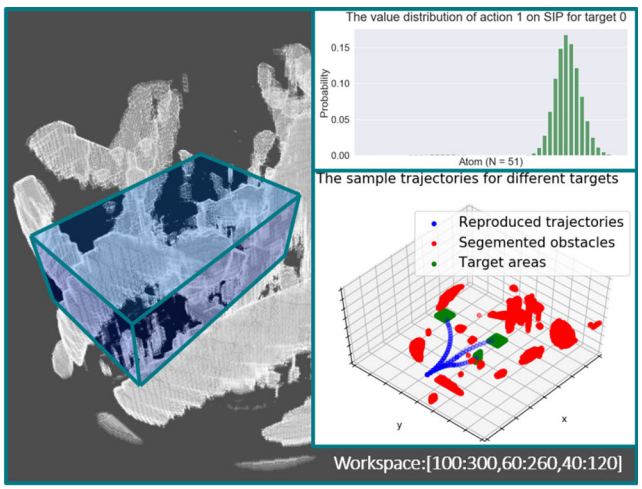
Table 1 Six groups of student's t test results

Conditions (with alpha level $a = 0.005$)	Groups	t value	p value	Results	Effect size
Condition 1 (UDQL & DDQN)	P_a	6.986	$p < 0.0005$	$p < a$	34
	P_s	6.554	$p < 0.0005$	$p < a$	
Condition 2 (UDQL & DDQN)	P_a	6.386	$p < 0.0005$	$p < a$	28
	P_s	6.046	$p < 0.0005$	$p < a$	
Condition 3 (UDQL & DDQN)	P_a	9.146	$p < 0.0005$	$p < a$	22
	P_s	7.844	$p < 0.0005$	$p < a$	
Condition 1 (UDQL & robust MDP)	P_a	6.840	$p < 0.0005$	$p < a$	34
	P_s	8.814	$p < 0.0005$	$p < a$	
Condition 2 (UDQL & robust MDP)	P_a	5.809	$p < 0.0005$	$p < a$	25
	P_s	7.613	$p < 0.0005$	$p < a$	
Condition 3 (UDQL & robust MDP)	P_a	3.585	$p < 0.005$	$p < a$	18
	P_s	7.064	$p < 0.0005$	$p < a$	

Each group contains two results for P_a and P_s, respectively



(a) Rewards Plot.



(b) Trajectories and value distribution.

Fig. 4 **a** The test rewards plot for UDQL agent on all targets and three DDQN agents on each target with standard deviation. **b** The sample trajectories for three targets and value distribution on SIP for first target. For the simplicity of demonstration, only the obstacles along the trajectories surface are plotted

algorithms in 3D work space. Because of the discretization, robust MDP method and traditional MDP method cannot be implemented in 3D work space because the number of states would increase exponentially during state dimension expansion. UDQL algorithm on the other hand, leverage DL to perform function approximation on continuous state space and consequently can adapt to high dimension state space.

In the simulation, the state is defined as a tuple $s = (s_x, s_y, s_z, s_\theta, s_\omega, s_\gamma)$, where (s_x, s_y, s_z) indicates the position of needle body-fix coordinate system C_t in global coordinate system C_{global} and $(s_\theta, s_\omega, s_\gamma)$ represents the rotation of C_t on Z-axis, Y-axis, and X-axis respectively. Hence, the transformation matrix from C_{global} to C_t could be calculated as $T_t = T_R(s_\theta)T_R(s_\omega)T_R(s_\gamma)$, where $T_R(\cdot)$ indicates the Euler angel rotation matrix. We only study the discrete action scenario due to the assumption in the studies [20,21,32]. The insertion unit is δ sampled from $\delta_\epsilon \sim \mathcal{N}(\delta, \sigma)$ for each time step to indicate the uncertainty. We follow all the other setup in section "Experiments" but augment the simulation to a 3D environment.

For comparison, DDQN agents are trained on each target individually. The test rewards of 10 repeated training process are shown in Fig. 4a. The P_a and P_s for both algorithms on each target are given in Table 2. The sample trajectory and value distribution are shown in Fig. 4b.

Based on the test rewards shown in Fig. 4a, the UDQL agent achieved a high average reward return with small standard deviation which demonstrates a stable learning pro-

Table 2 3D experiment results on designated SIP

Different segmented targets	Indicators of different methods			
	DDQN		UDQL	
	Accuracy (%)	Avoidance (%)	Accuracy (%)	Avoidance (%)
Target 0	84.9	90.5	85.2	98.0
Target 1	86.6	84.9	88.5	90.5
Target 2	73.4	83.7	86.4	86.7

cedure and a superior generalization on all targets. The baseline DDQN methods cannot achieve a stable learning procedure that the rewards have large deviations. The UDQL agent takes advantage of leveraging multi-goal setup with HER sampling to achieve a good generalization on all targets and utilizing value distribution representation to obtain a stable learning procedure.

Table 2 confirms the improvement in robustness and accuracy of utilizing UDQL in path planning. The value distribution of SIP with action $a = 1$ for target 0 shown in Fig. 4b indicates a high unimodality of future return which supports the high avoidance rate of inserting target 0 given in Table 2. Normally, for clinical usage, a unimodal value distribution with high mean and low variance indicates a low-risk high reward plan.

Phantom validation and error comparison

To validate the feasibility of implementing UDQL in robot-assisted needle insertion scenario, a phantom experiment is designed and performed on our prototype needle insertion surgical robot. In this experiment, we simulate the liver tumor RFA surgery on a liver phantom within a full-size human model to demonstrate the clinical work space. The liver model is constructed by a 3D printing base to fix the position in human model and a soft tissue layer which is made by mixing 5 g agarose powder (SeaKem® LE Agarose, USA) with 400 ml water. Three hemisphere targets with radius $r = 1$ cm are located in the bottom of soft layer. The objective of this phantom experiment is to manipulate the robot and steer the flexible RFA needle (Fig. 2b) to reach the three targets through SIP by utilizing the UDQL generated policy.

The experiment was repeated five times on five identical models. The phantom design, experiment instruments, and sample trajectories are shown in Fig. 5. The experimental results showed that the needle insertion manipulated by UDQL agent reached targets with error ±2.3 mm. Since the registration and kinematic modeling in robot manipulation were based on our previous work, we refer interested reader to our work in the image-guided robotic system for RFA [7] for more details.

In the simulation, the error is calculated by measuring the Euclidean distance between the needle tip position at the final time step and the center of target area on 1000 trails. For the phantom validation experiment, we manually measure the distance between the needle tip at final time step and the target centroids as error on each target insertion. The experimental errors in computer simulation and phantom experiment are given in Table 3. In 2D simulation, we leveraged UDQL algorithm, DDQN algorithm, robust MDP algorithm, and traditional MDP algorithm. The methods that were implemented in 2D discrete state space would induce a discretization error bounded by $e_s = \Delta\sqrt{2}(n+0.5)$, where n is the number of rotations [13]. However, the DRL methods were implemented on the continuous state space, there is no discretization error in the planning stage.

In the phantom experiment, our method induced ±2.3 mm error in inserting designated targets. Compared with other state-of-art needle insertion robots utilized in biopsy and surgery, the magnitude of the error on our simulation and validation may not be impressive. However, the target error highly depends on the instruments usage and dedicated experiment environment. In our work, we implement flex-

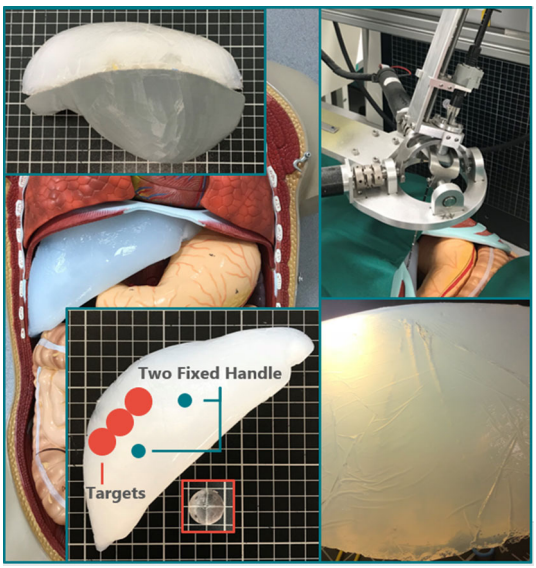


Fig. 5 The liver phantom, experiment instruments, and sample trajectory. The liver model was constructed by half soft tissue on hard base (upper-left) to fix the position in the human model (left). The targets in the soft part are shown in the figure at bottom-left. The bottom-right figure demonstrates the sample trajectory inserted by flexible needle

Table 3 The error measured in 2D simulation, 3D simulation, and phantom validation

Experiments	Methods	Experimental error (mm)	Discretization error (mm)
2D simulation	UDQL	± 0.8	–
	DDQN	± 1.3	–
	Robust MDP	±1.6	± 0.7
	Traditional MDP	± 1.9	± 0.7
3D simulation	UDQL	± 1.8	–
	DDQN	± 3.0	–
Phantom validation	UDQL	± 2.3	–

ible RFA needle which could induce more uncertainty in needle–tissue interaction compared with rigid needle used in other works. Also, the flexible RFA needle has a larger diameter (approximate 11-gauge) compared with the biopsy needle (normally 17–19 gauge) used in other state-of-art works [3,33–35].

Conclusion

In this paper, we proposed a new path planning framework called UDQL which can perform multi-goal DRL and provide value distribution for risk management. Compared with previous works [6,12], UDQL agents could perform planning on 3D continuous state space without introducing error caused by state discretization and counter the uncertainty in a more realistic representation. The value distribution indicates the distribution of expected return based on specific reward function design. In the flexible needle insertion scenario, the value distribution on SIP indicates the distribution of expected insertion outcome. By analyzing the value distribution, the surgeons could perform risk management manually based on patients specific conditions and the distribution multimodality.

Our algorithm is performed on the off-line simulation without considering the closed-loop control on the current state. The uncertainty of the needle–tissue interaction is implicitly considered in the planning stage as transition probability. Based on the experimental results, UDQL algorithm is capable of performing path planning at SIP with multiple targets and achieving a stabilized learning procedure without complicated reward engineering. The phantom experimental results and error comparison shown in section "Phantom validation and error comparison" demonstrate the feasibility of UDQL agent in steering a flexible RFA needle with a prototype needle insertion surgical robot under an acceptable error level [7].

For future works, we would like to fully test the proposed path planning framework in both complex ex vivo and in vivo experiments with various CT segmentation methods. The uncertainty in multi-layer needle–tissue interaction will be evaluated and modeled in the simulation to achieve a more robust and accurate agent.

Acknowledgements The last author would like to acknowledge the contribution of A/Prof Stephen Chang of Mount Elizabeth Hospital, Singapore for his input on surgeries and medical education.

Funding The research and development of the prototype Image-guide Radio-frequency Ablation Surgical System was supported in parts by Research Grants from Singapore Agency of Science and Technology (A*Star) and Ministry of Education, Singapore respectively.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Informed consent Informed consent was obtained from all individual participants included in the study.

Human and animal rights All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

References

1. Hamad GG, Curet M (2010) Minimally invasive surgery. Am J Surg 199(2):263–265
2. Tan X, Chng C-B, Ye S, Lim K-B, Chui C-K (2019) Robot-assisted training in laparoscopy using deep reinforcement learning. IEEE Robot Autom Lett 4(2):485–492
3. Hiraki T, Kamegawa T, Matsuno T, Sakurai J, Kirita Y, Matsuura R, Yamaguchi T, Sasaki T, Mitsuhashi T, Komaki T (2017) Robotically driven ct-guided needle insertion: preliminary results in phantom and animal experiments. Radiology 285(2):454–461
4. Abolhassani N, Patel R, Moallem M (2007) Needle insertion into soft tissue: a survey. Med Eng Phys 29(4):413–431
5. Schaul T, Horgan D, Gregor K, Silver D (2015a) Universal value function approximators. In: International conference on machine learning. pp 1312–1320
6. Tan X, Yu P, Lim K-B, Chui C-K (2018) Robust path planning for flexible needle insertion using Markov decision processes. Int J Comput Assist Radiol Surg 13(9):1439–1451
7. Duan B, Wen R, Chng C-B, Wang W, Liu P, Qin J, Peneyra JL, Chang SK-Y, Heng P-A, Chui C-K (2015) Image-guided robotic system for radiofrequency ablation of large liver tumor with sin-

gle incision. In: 2015 12th International conference on ubiquitous robots and ambient intelligence (URAI). IEEE, pp 284–289
8. DiMaio SP, Salcudean SE (2005) Needle steering and motion planning in soft tissues. IEEE Trans Biomed Eng 52(6):965–974
9. Taylor RH, Menciassi A, Fichtinger G, Fiorini P, Dario P (2016) Medical robotics and computer-integrated surgery. In: Siciliano B, Khatib O (eds) Springer handbook of robotics. Springer, Cham, pp 1657–1684
10. Liu P, Qin J, Duan B, Wang Q, Tan X, Zhao B, Jonnathan PL, Chui C-K, Heng P-A (2018) Overlapping radiofrequency ablation planning and robot-assisted needle insertion for large liver tumors. Int J Med Robot Comput Assist Surg 15:e1952
11. Chatelain P, Krupa A, Navab N (2015) 3d ultrasound-guided robotic steering of a flexible needle via visual servoing. In: IEEE international conference on robotics and automation, ICRA'15
12. Alterovitz R, Siméon T, Goldberg KY (2007) The stochastic motion roadmap: a sampling framework for planning with Markov motion uncertainty. In: Robotics: science and systems, vol 3, pp 233–241
13. Alterovitz R, Branicky M, Goldberg K (2008) Motion planning under uncertainty for image-guided medical needle steering. Int J Robot Res 27(11–12):1361–1374
14. Morar A, Moldoveanu F, Gröller E (2012) Image segmentation based on active contours without edges. In: 2012 IEEE 8th international conference on intelligent computer communication and processing. IEEE, pp 213–220
15. Chen X, Nguyen BP, Chui C-K, Ong S-H (2016) Automated brain tumor segmentation using kernel dictionary learning and superpixel-level features. In: 2016 IEEE international conference on systems, man, and cybernetics (SMC). IEEE, pp 002547–002552
16. Sutton RS, Barto AG (1998) Introduction to reinforcement learning, vol 135. MIT Press, Cambridge
17. Deng L, Yu D (2014) Deep learning: methods and applications. Found Trends® Signal Process 7(3–4):197–387
18. Silver D, Schrittwieser J, Simonyan K, Antonoglou I, Huang A, Guez A, Hubert T, Baker L, Lai M, Bolton A (2017) Mastering the game of go without human knowledge. Nature 550(7676):354
19. Schulman J, Wolski F, Dhariwal P, Radford A, Klimov O (2017) Proximal policy optimization algorithms. arXiv:1707.06347
20. Fu YB, Chui CK, Teo CL (2013) Liver tissue characterization from uniaxial stress–strain data using probabilistic and inverse finite element methods. J Mech Behav Biomed Mater 20:105–112
21. Fu YB, Chui CK (2014) Modelling and simulation of porcine liver tissue indentation using finite element method and uniaxial stress–strain data. J Biomech 47(10):2430–2435
22. Qu C, Mannor S, Xu H (2018) Nonlinear distributional gradient temporal-difference learning. arXiv:1805.07732
23. Bellemare MG, Dabney W, Munos R (2017) A distributional perspective on reinforcement learning. arXiv:1707.06887
24. Andrychowicz M, Wolski F, Ray A, Schneider J, Fong R, Welinder P, McGrew B, Tobin J, Abbeel OP, Zaremba W (2017) Hindsight experience replay. In: Advances in neural information processing systems, pp 5048–5058
25. Schaul T, Quan J, Antonoglou I, Silver D (2015b) Prioritized experience replay. arXiv:1511.05952
26. Tamar A, Di Castro D, Mannor S (2016) Learning the variance of the reward-to-go. J Mach Learn Res 17(1):361–396
27. Yang L, Wen R, Qin J, Chui C-K, Lim K-B, Chang SK-Y (2010) A robotic system for overlapping radiofrequency ablation in large tumor treatment. IEEE/ASME Trans Mechatron 15(6):887–897
28. Tan X, Chng C-B, Duan B, Ho Y, Wen R, Chen X, Lim K-B, Chui C-K (2017) Cognitive engine for robot-assisted radio-frequency ablation system. Acta Polytech Hung 14(1):129–145
29. Tan X, Chng C-B, Duan B, Ho Y, Wen R, Chen X, Lim K-B, Chui C-K (2016) Design and implementation of a patient-specific cognitive engine for robotic needle insertion. In: 2016 IEEE international conference on systems, man, and cybernetics (SMC). IEEE, pp 000560–000565
30. Van Hasselt H, Guez A, Silver D (2016) Deep reinforcement learning with double Q-learning. In: AAAI, vol 16, pp 2094–2100
31. Wasserstein RL, Lazar NA (2016) The asa's statement on *p* values: context, process, and purpose. Am Stat 70(2):129–133
32. Leong F, Huang W-H, Chui C-K (2013) Modeling and analysis of coagulated liver tissue and its interaction with a scalpel blade. Med Biol Eng Comput 51(6):687–695
33. Tokuda J, Song S-E, Fischer GS, Iordachita II, Seifabadi R, Cho NB, Tuncali K, Fichtinger G, Tempany CM, Hata N (2012) Preclinical evaluation of an MRI-compatible pneumatic robot for angulated needle placement in transperineal prostate interventions. Int J Comput Assist Radiol Surg 7(6):949–957
34. Krieger A, Susil RC, Fichtinger G, Atalar E, Whitcomb LL (2004) Design of a novel MRI compatible manipulator for image guided prostate intervention. In: IEEE international conference on robotics and automation, proceedings on ICRA'04, vol 1. IEEE, pp 377–382
35. Schouten MG, Ansems J, Renema WKJ, Bosboom D, Scheenen TWJ, Fütterer JJ (2010) The accuracy and safety aspects of a novel robotic needle guide manipulator to perform transrectal prostate biopsies. Med Phys 37(9):4744–4750