

Bimanual Regrasping for Suture Needles using Reinforcement Learning for Rapid Motion Planning

Zih-Yun Chiu¹, Florian Richter¹ *Student Member, IEEE*,
 Emily K. Funk², Ryan K. Orosco² *Member, IEEE*, and Michael C. Yip¹ *Member, IEEE*

Abstract—Regrasping a suture needle is an important yet time-consuming process in suturing. To bring efficiency into regrasping, prior work either designs a task-specific mechanism or guides the gripper toward some specific pick-up point for proper grasping of a needle. Yet, these methods are usually not deployable when the working space is changed. Therefore, in this work, we present rapid trajectory generation for bimanual needle regrasping via reinforcement learning (RL). Demonstrations from a sampling-based motion planning algorithm is incorporated to speed up the learning. In addition, we propose the ego-centric state and action spaces for this bimanual planning problem, where the reference frames are on the end-effectors instead of some fixed frame. Thus, the learned policy can be directly applied to any feasible robot configuration. Our experiments in simulation show that the success rate of a single pass is 97%, and the planning time is 0.0212s on average, which outperforms other widely used motion planning algorithms. For the real-world experiments, the success rate is 73.3% if the needle pose is reconstructed from an RGB image, with a planning time of 0.0846s and a run time of 5.1454s. If the needle pose is known beforehand, the success rate becomes 90.5%, with a planning time of 0.0807s and a run time of 2.8801s.

I. INTRODUCTION

On the journey towards complete autonomy in surgical robotics [1], automation of suturing has engrossed the community. There is a good reason for this, as manual suturing has been cited as being time consuming [2], and suturing on Robot Assisted Minimally Invasive Surgeries (RAMIS) can be tedious and challenging [3]. Hence, automation of suturing becomes an attractive goal for improving a surgeon's quality of life during RAMIS. Emergence of surgical robotics simulators [4] and open-sourced robotic platforms such as the da Vinci Research Kit (dVRK) [5] provide the ability for rapid development towards realizing this goal.

Suturing requires a wide range of considerations for effective automation including needle manipulation [6], knot tying [7], identification of entry and exit points for the suture throw [8], and interfacing the automation with the surgeon for effective deployment [9]. In this work, we investigate the specific sub-challenge of needle manipulation from the perspective of bimanual regrasping, which is to regrasp the

This research was supported by the Telemedicine and Advanced Technology Research Center (TATRC) T.R.O.N. program.

¹Zih-Yun Chiu, Florian Richter, and Michael C. Yip are with the Department of Electrical and Computer Engineering, University of California San Diego, La Jolla, CA 92093 USA. {zchiu, frichter, yip}@ucsd.edu

²Emily K. Funk and Ryan K. Orosco are with the Department of Surgery - Division of Head and Neck Surgery, University of California San Diego, La Jolla, CA 92093 USA. {ekfunk, rorosco}@ucsd.edu

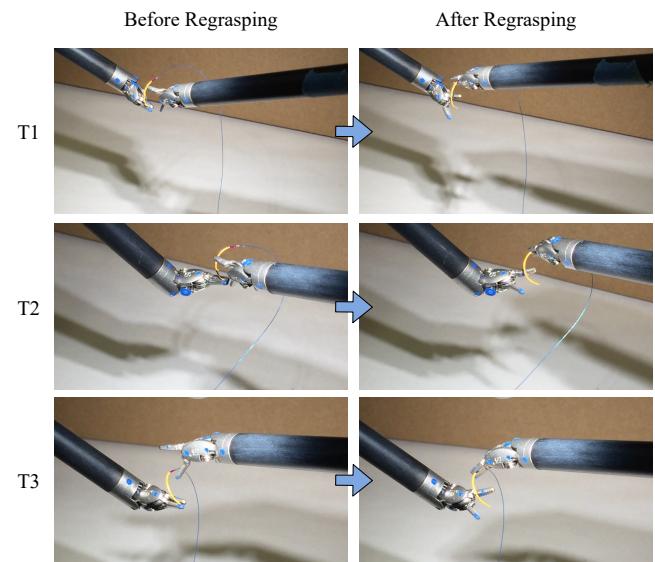


Fig. 1: Three trials (T1, T2, T3) of successful needle regrasping. The pictures on the right show a da Vinci surgical robot arm successfully regrasps the suture needle at the pose where a surgeon often holds before suturing.

needle at a proper pose using dual robot arms. Fontanelli et al. cite that 74% of suture throws from the in-vivo JIGSAW dataset [10] required regrasping which took an expert on average 7.9s to complete [11]. Previous work has also shown that proper grasping can enhance the effectiveness of suture throws [8], [12]. In fact, many planning techniques for suture needle throwing have even used task-specific gripper mechanisms to ensure proper grasping of a needle [13]–[16]. While these mechanisms are effective in highlighting suture throwing capabilities, they are too task-specific to be deployed on a surgical robot.

On the other hand, previous work has approached the proper needle grasping problem by visual servoing [17] or learning from demonstrations [18]. However, these works might not avoid needle regrasping after picking it up if the needle is displaced or reoriented due to collision. Also, these methods are designed under some specific robot configuration and might require new calibration or training in another working space. Therefore, regrasping suture needles to the optimal location for throwing on standard RAMIS tooling is still of great importance.

Standard practice for surgeons is to regrasp the suture

needle roughly one third needle length from the suture thread and over 30° off the needle as shown in Fig. 1. This gives an end-goal for the robot to achieve that can be given to a motion planner as a terminal state in order to find a collision free trajectory for regrasping. However, our experiments show that every available state-of-the-art motion planner was too slow for this task and often times-out. This problem is not unexpected; planning times typically can take up to 3 minutes for constrained environments and neural planning methods are required to breach this limitation [19]. These methods however are not directly developed for dual arm robots which are typical for RAMIS. In particular, the bases of the robotic arms are regularly adjusted between procedures so the learning based motion planner needs to generalize to all potential workspaces of the surgical robot.

To this end, we present a novel method for bimanual regrasping for suture needles. We use an RL approach integrated with rapid replanning to demonstrate real-time performance and success in regrasping. We structure the problem to be ego-centric to the needle and the end-effectors, making it far more generalizable than a typical configuration-space approach.

To plan for the needle passing task, the states and actions are related to the pose of the two end-effectors. The reference frames for these poses are not set to be the world frame or the base frame of an arm. Instead, the reference frames can be the frames of the end-effectors, which *change as they move*. This setting is referred as *the ego-centric setting*, since the information that the end-effectors get at each time step is centered on themselves.

There are several benefits of using the ego-centric setting. First, this setting allows the policy to successfully work in the global space without requiring the agent to explore the wide range of space. Instead, it only needs to observe the states relative to the goal, which are included in a local space. Hence, no matter where the agent and the goal are in the world frame, as long as their relation has been learned by the policy, the agent is able to reach the goal. Also, since the policy does not consider the configuration of a robot, it can be directly used for other robot arms. These benefits suggest that the ego-centric setting is very data efficient for learning a policy that can work in many different scenarios.

To sum up, we specifically present the following contributions in this work:

- 1) fast trajectory generation for bimanual needle regrasping for suture throwing via RL,
- 2) RL training strategy which incorporates intermittent, targeted exploration to guide the policy while still allowing for it to generalize, and
- 3) ego-centric parameterization of the needle and surgical tools to generalize the trajectory for non-specific end-effector nor robotic base positioning.

While this work is directly aimed towards reducing the time needed to suture when conducting RAMIS, we also foresee it being an important step towards making other automated suture throw techniques deployable by relieving them of task-specific suture needle grasping mechanisms.

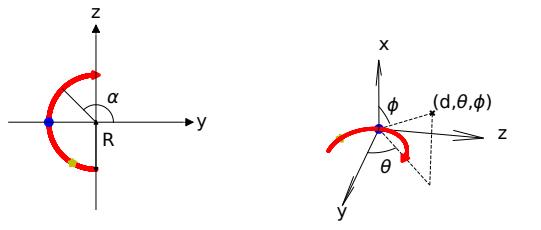


Fig. 2: Coordinate systems on the needle (red semi-circle: the suture needle, red arrow: the tip of the suture needle, blue dot: an example of a grasping point, yellow dot: an example of a goal grasping point)

II. REINFORCEMENT LEARNING BACKGROUND

The suture needle passing task in this work is formulated as an RL problem and is modelled as a Markov Decision Process (MDP). The goal is to find a deterministic policy π that solves this MDP. The following notations and definitions are used for the RL formulation. Let $\mathcal{M} = (\mathbb{S}, \mathbb{A}, P, r, \rho_0, \gamma, T)$ be the tuple for the discrete-time finite-horizon MDP, where \mathbb{S} is the state space, \mathbb{A} is the action space, P is the transition probability, r is the reward function, ρ_0 is the initial state distribution, γ is the discount factor, and T is the task horizon. A policy π is learned to maximize the expected return $\mathbb{E}_{\mathbf{s}_0 \sim \rho_0, P} [\sum_{t=0}^T \gamma^t r(\mathbf{s}_t, \mathbf{a}_t)]$, where $\mathbf{s}_t \in \mathbb{S}$ is the state at time step t , $\mathbf{a}_t = \pi(\mathbf{s}_t) \in \mathbb{A}$ is the action at t , and $\mathbf{s}_{t+1} \sim P(\mathbf{s}_{t+1} | \mathbf{s}_t, \mathbf{a}_t)$.

III. METHODS

In order to develop and test suture needle passing, a simulated environment is developed in V-REP based on Fontanelli's et al. previous work [20].

A. Initialization and Goal Generation for Needle Grasping

At the beginning of conducting needle passing, the needle can be grasped at a variety of points and directions. Therefore, the RL environment will be initialized with a random sample of an initial grasping point and direction such that the trained policy for motion planning will generalize well. In addition, the grasping point must be well defined. To model this mathematically, the following two coordinate frames are built on the needle.

1) Needle frame: A Cartesian coordinate system is defined on the suture needle as shown in Fig. 2a. This frame is referred as *the needle frame*. The most common suture needle used in RAMIS are of a semicircle shape, hence these will be considered in this work. The equation for the semicircle suture needle in the needle frame is defined as:

$$\begin{cases} x_n = 0 \\ y_n = R \cos \alpha \\ z_n = R \sin \alpha. \end{cases} \quad (1)$$

where R is the radius of the needle and $\alpha \in [\frac{\pi}{2}, \frac{3\pi}{2}]$ is the angle on the needle. By randomly sampling an α , a grasping point on the suture needle can be sampled. The pose of this

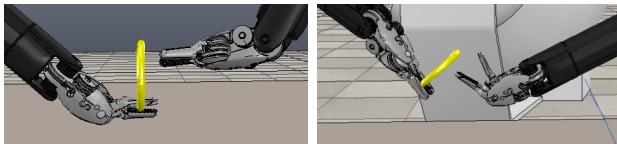


Fig. 3: Two different configurations of the robot arm that have the same state. Since a state is defined with the ego-centric setting, no matter how the robot configuration changes, as long as the relative poses of the two end-effectors and the needle remain the same, the state will not be changed.

grasping point in the needle frame is denoted as $(\mathbf{p}_g^n, \mathbf{q}_g^n)$, where $\mathbf{p}_g^n = [x_n \ y_n \ z_n]^\top$ and is calculated by equation (1). Here, $\mathbf{q}_g^n = [1 \ 0 \ 0 \ 0]^\top$ is the identity rotation represented in the quaternion form. Likewise, the goal grasping point can be defined by setting α .

2) Grasping point frame: In order to sample a grasping direction pointing to a grasping point for initialization, a spherical coordinate system is defined with the origin at the grasping point as shown in Fig. 2b. This frame is referred as *the grasping point frame*. A point using Cartesian representation in this frame can be calculated as

$$\begin{cases} x_g = d \cos \phi \\ y_g = d \sin \phi \cos \theta \\ z_g = d \sin \phi \sin \theta. \end{cases} \quad (2)$$

where (d, θ, ϕ) are the radius, azimuth, and inclination respectively from the grasping point frame. This parameterization is beneficial for setting grasping direction of a robotic gripper. d defines the depth of the grasp with regards to the gripper while θ and ϕ give the grasping angle relative to the needle. Therefore, by randomly sampling a (d, θ, ϕ) , the target position of the end-effector will be set to $\mathbf{p}_e^g = [x_g \ y_g \ z_g]^\top$, and its orientation will become \mathbf{q}_e^g such that the gripper points from $[x_g \ y_g \ z_g]^\top$ to the origin in the grasping point frame. $(\mathbf{p}_e^g, \mathbf{q}_e^g)$ can then be transformed to the needle frame by

$$\mathbf{H}_e^n(\mathbf{p}_e^n, \mathbf{q}_e^n) = \mathbf{H}_g^n(\mathbf{p}_g^n, \mathbf{q}_g^n) \mathbf{H}_e^g(\mathbf{p}_e^g, \mathbf{q}_e^g). \quad (3)$$

where $\mathbf{H}(\cdot, \cdot) \in SE(3)$ is the homogeneous representation of a pose. Then the end-effector can be set to reach $(\mathbf{p}_e^n, \mathbf{q}_e^n)$, hence grasping the needle, through inverse kinematics. The goal grasping direction can be set in a similar fashion. In the following sections, the end-effector that is initialized to hold a needle is referred as the grasping end-effector, and the one that approaches the goal is referred as the regrasping end-effector.

B. Ego-Centric State and Action Space

With the ego-centric setting, the states in \mathbb{S} and actions in \mathbb{A} for the needle passing task are defined as follows. A state includes

- 1) the position and quaternion of the needle measured in the grasping end-effector frame $\mathbf{p}_n^{e_g}, \mathbf{q}_n^{e_g}$, and

- 2) the position and quaternion of the regrasping end-effector measured in the grasping end-effector frame $\mathbf{p}_{e_r}^{e_g}, \mathbf{q}_{e_r}^{e_g}$, and
- 3) the position and quaternion of the needle measured in the regrasping end-effector frame $\mathbf{p}_n^{e_r}, \mathbf{q}_n^{e_r}$.

After initializing the environment, $(\mathbf{p}_n^{e_g}, \mathbf{q}_n^{e_g})$ is available by calculating the inverse of equation (3). For simplicity, in this work it is assumed that only the regrasping arm moves during planning. Yet, the ego-centric setting can be directly applied to two moving arms. Moreover, since a state has nothing to do with the joint angles or the base frame of the robot, two very different configurations of a robot arm can have the same state as shown in Fig. 3.

An action is defined as the variation of the position and quaternion of the regrasping end-effector measured in the grasping end-effector frame $\Delta \mathbf{p}_{e_r}^{e_g}, \Delta \mathbf{q}_{e_r}^{e_g}$. Given that the regrasping end-effector moves from $(\mathbf{p}_{e_r,t-1}^{e_g}, \mathbf{q}_{e_r,t-1}^{e_g})$ to $(\mathbf{p}_{e_r,t}^{e_g}, \mathbf{q}_{e_r,t}^{e_g})$ at time step t , the action would be

$$\Delta \mathbf{p}_{e_r}^{e_g} = \mathbf{p}_{e_r,t}^{e_g} - \mathbf{p}_{e_r,t-1}^{e_g} \quad (4)$$

$$\Delta \mathbf{q}_{e_r}^{e_g} = \mathbf{q}_{e_r,t}^{e_g} (\mathbf{q}_{e_r,t-1}^{e_g})^{-1} \quad (5)$$

which would eventually be used as the control commands to plan with in this ego-centric setting.

C. Learning a Policy by Deep Deterministic Policy Gradients with Demonstrations

To solve the suture needle passing task, we train a policy by Deep Deterministic Policy Gradients (DDPG) with behavior cloning (BC) [21]. DDPG is one of the most widely used RL algorithms for continuous control, and its combination with BC helps guide the exploration of an agent. Since the process of passing a needle is very sophisticated, especially when the two end-effectors are close to each other, it is reasonable to incorporate demonstrations into learning to motivate precise motions and speed up training.

The states and actions of this task are defined by the ego-centric setting as mentioned in the previous section. Also, the rewards are defined as follows [22]:

$$r = \begin{cases} -1, & \text{if collision happens} \\ 1, & \text{if equation (7) and (8) are satisfied} \\ \beta, & \text{otherwise} \end{cases}. \quad (6)$$

$$\|\mathbf{p}_{n,*}^{e_r} - \mathbf{p}_{n,t}^{e_r}\|_2 \leq \varepsilon_p \quad (7)$$

$$\|\text{AxisAngle}(\mathbf{q}_{n,*}^{e_r} (\mathbf{q}_{n,t}^{e_r})^{-1})\|_2 \leq \varepsilon_q \quad (8)$$

Equation (7) and (8) describe the distance between the current pose $(\mathbf{p}_{n,t}^{e_r}, \mathbf{q}_{n,t}^{e_r})$ and the goal grasping pose $(\mathbf{p}_{n,*}^{e_r}, \mathbf{q}_{n,*}^{e_r})$, where $\varepsilon_p \geq 0$ and $\varepsilon_q \geq 0$ are the maximum values of this distance. $\text{AxisAngle}(\cdot)$ transforms a quaternion to its axis-angle representation. $\beta < 0$ is a value that can be tuned. This value should be small enough to prevent the agent from not approaching the needle, while it cannot be too small, or the agent might choose to collide with the needle instead of roaming to find a feasible path. Usually, the value of β is tuned according to the reward of collision, which is -1 here, and the maximum time steps that the agent is allowed to run in the environment.



Fig. 4: Feasible space and scenario when the regrasping end-effector approaches the needle. As being closer to the needle, the collision-free space decreases in a funnel like fashion as shown by the plot on the left. Therefore, forward planning is very challenging, and instead we use reverse planning.

1) *Generating expert demonstrations:* Generating expert demonstrations can be time consuming if we ask a real expert to do so. However, the goal position and orientation of holding a suture needle can be calculated from equation (3) by setting some target $(\mathbf{p}_{e_r}^n, \mathbf{q}_{e_r}^n)$. Therefore, the expert demonstrations can be generated automatically in the simulation environment.

An intuitive way to do this is to apply some motion planning algorithm to a given start and goal configurations. Yet, in this way, the feasible space is narrower near the goal, so a gripper is more likely to collide with the needle or another gripper. This leads to failure for planning algorithms to find a feasible path in a given time period.

Fig. 4 shows the scenario and probability of no collision when the regrasping gripper moves closer to the goal. From this figure, it can be observed that the feasible space looks like a funnel with the narrow part near the goal. Note that this space shrinks a lot within a small distance hence making the planning require high precision. This makes a motion planning algorithm more difficult to find a path without collision. To overcome this difficulty, instead of planning from a starting point to a goal grasping point, a path can be planned in reverse, i.e., escaping a funnel. Based on this concept, the expert demonstrations are generated by applying a motion planning algorithm from the goal grasping point to a randomly sampled starting point in the free space. The planned path is then reversed to be used for learning a policy.

2) *Applying demonstrations via an Active Learning approach:* Active learning involves collecting demonstrations during training. It is used to demonstrate successful episodes only when the RL policy fails, which helps guide the exploration of the robot and significantly reduce random exploration [22].

This strategy, referred as target exploration in [22], can be easily applied to our work. However, in our experiments it is observed that eliminating random exploration slows down the learning. The main reason is that the expert, which is a motion planning algorithm, is not perfect. It is not guaranteed that this algorithm can always provide a feasible path, and relying almost entirely on targeted exploration can instead

Algorithm 1: Generate Episodes with Mixed Exploration

```

Input: env, number of episodes  $N$ , policy  $\pi$ , random
exploration probability  $\mathcal{E}_R$ , target exploration
probability  $\mathcal{E}_T$ , motion planning algorithm
MP

episodes  $\leftarrow [ ]$ 
for  $n \leftarrow 1$  to  $N$  do
    epi  $\leftarrow [ ]$ 
    s = env.reset()
    while s is not a terminal state do
        a =  $\begin{cases} \pi(\mathbf{s}) + \text{noise} & \text{with probability } \mathcal{E}_R \\ \pi(\mathbf{s}) & \text{with probability } 1 - \mathcal{E}_R \end{cases}$ 
        s', r, done = env.step(a)
        epi.append([s, a, r, done, s'])
        s  $\leftarrow s'
    end
    episodes.append(epi)
    if epi fails and with probability  $\mathcal{E}_T$  then
        demo.epi  $\leftarrow [ ]$ 
        cs, cg  $\leftarrow$  Start and goal configurations of the
robot from epi
        demo.epi  $\leftarrow$  MP(cs, cg)
        episodes.append(demo.epi)
    end
end
Output: episodes$ 
```

hurt the performance. Therefore, to make an agent explore in a more effective way, random exploration is kept, and target exploration is gradually introduced to training. This exploration strategy is referred as *mixed exploration*.

When generating episodes with mixed exploration, if an episode fails, then with probability \mathcal{E}_T , a motion planning algorithm will generate a demonstration for this episode. \mathcal{E}_T is the probability of applying target exploration and will gradually increase from \mathcal{E}_T^0 to 1. Meanwhile, to stabilize training, the probability of random exploration \mathcal{E}_R will gradually decrease from \mathcal{E}_R^0 to 0.

Algorithm 1 describes the process of generating episodes with mixed exploration. After generating N episodes, \mathcal{E}_R and \mathcal{E}_T will be updated by

$$\mathcal{E}_R = \max (\mathcal{E}_R - \Delta\mathcal{E}_R, 0) \quad (9)$$

$$\mathcal{E}_T = \min (\mathcal{E}_T + \Delta\mathcal{E}_T, 1), \quad (10)$$

where $\Delta\mathcal{E}_R > 0$ and $\Delta\mathcal{E}_T > 0$ are the changes applied to \mathcal{E}_R and \mathcal{E}_T respectively.

IV. EXPERIMENTAL SETUP

A series of experiments are conducted in order to evaluate the performance of the proposed methods. First the training and testing settings are defined followed by a comparison of different exploration strategies for RL training. The learned motion planner from RL is also compared against classical motion planners. Lastly, the policy is implemented and tested

on a real robot to pass a 5.4mm suture needle, and all regrasps are done in a single pass.

The needle selected for simulation is of radius $R = 5.4\text{mm}$. To resemble the preferred grasping position and orientation of surgeons before throwing the suture needle, the goal is generated by setting $\alpha = \frac{4\pi}{3}$ and $d \in [6.5\text{mm}, 8\text{mm}]$, $\theta = 0$, $\phi = \pi$ in the needle and grasping point frame respectively. Meanwhile, the randomized initialization for the initial grasping is generated by sampling uniformly from $\alpha \in [\frac{11}{18}\pi, \frac{13}{18}\pi]$ and $d \in [6.5\text{mm}, 8\text{mm}]$, $\theta \in [0, 2\pi]$, $\phi \in [0, 0.4\pi]$ in the needle and grasping point frame respectively. The regrasping end-effector is randomly initialized to be 13mm away from the center of the needle with a position variation $\sim \mathcal{U}(-1\text{ mm}, 1\text{ mm})$ and a orientation variation $\sim \mathcal{U}(-5^\circ, 5^\circ)$ along some rotation axis. $\mathcal{U}(\cdot, \cdot)$ is a uniform distribution. Expert demonstrations for both the behavioral cloning and mixed exploration are generated using batch informed trees (BIT*) [23] due to its good performance in planning time. The maximum horizon T of the RL environment is 100, β in the reward function is tuned to be -0.02 , and $(\varepsilon_p, \varepsilon_q) = (1\text{mm}, 5^\circ)$. The values of $(\varepsilon_p, \varepsilon_q)$ are empirically set to ensure a proper grasp. For the upcoming comparison studies, a test set of 300 randomized initialized grasps is generated and used for all the corresponding results.

The implementation of DDPG + BC is based on OpenAI Baselines [24]. The actor and critic neural networks in DDPG are both multilayer perceptrons with 3 layers and 512 neurons per layer. They are trained for a total of 500 epochs. Each epoch contains 10 iterations, where in each iteration, 5 episodes are collected in parallel to fill a replay buffer with a set size of 10^6 . Per iteration, the actor and critic networks are updated a total of 200 times with a batch size of 256 and learning rate of 10^{-4} for both the actor and critic. Meanwhile, the target actor and critic networks are updated only once every 40 of those updates with a coefficient value of 0.95 for polyak-averaging. The discount factor, γ , is set to 0.99, and an additional quadratic penalty for the actions is with a coefficient of 1. The coefficient for the primary and cloning loss of the actor are both set to 10^{-3} . The total number of demonstration episodes generated for BC is 9900.

V. RESULTS

A. Solving Needle Passing with Mixed Exploration

To compare the proposed RL exploration algorithm with other methods, three separate policies are trained with the following settings in the needle passing environment:

- 1) DDPG + BC [21] where: $\mathcal{E}_R^0 = 0.3$, $\Delta\mathcal{E}_R = 6 \times 10^{-4}$
- 2) DDPG + BC with targeted exploration [22] where: $\mathcal{E}_R^0 = 0.02$, $\Delta\mathcal{E}_R = 0$, $\mathcal{E}_T^0 = 0.5$, $\Delta\mathcal{E}_T = 0$
- 3) DDPG + BC with mixed exploration where: $\mathcal{E}_R^0 = 0.3$, $\Delta\mathcal{E}_R = 6 \times 10^{-4}$, $\mathcal{E}_T^0 = 0.1$, $\Delta\mathcal{E}_T = 4.5 \times 10^{-3}$

The DDPG + BC portion in all three methods uses the same network architecture and training hyper-parameters as previously listed. Also, each setting is run three times with different random seeds.

The training curves are shown in Fig. 5, and performance of the best trained policies are shown in Table I. DDPG + BC

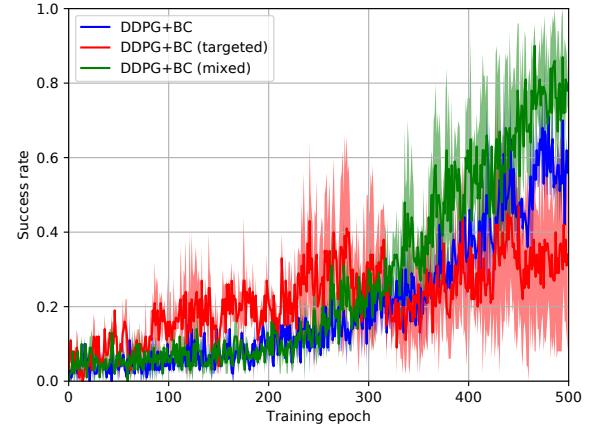


Fig. 5: Training curves of different exploration strategies to solve the needle passing environment.

Algorithm	Success rate	Planning time (s)	Path length (mm)
DDPG+BC	0.71	0.0262	4.2
DDPG+BC (targeted)	0.64	0.0233	4.2
DDPG+BC (mixed)	0.97	0.0212	4.2

TABLE I: Success rate, average planning time and path length of different RL algorithms on the needle passing task

(mixed) outperforms the other two methods with regards to success rate and planning time. DDPG + BC (targeted) does reach a higher success rate at the beginning, but becomes unstable as the training goes on due to more heavily relying on imperfect experts. Therefore, DDPG + BC is able to eventually surpass DDPG + BC (targeted) for success rate.

B. Comparing Against Sampling-Based Motion Planners

To show that DDPG + BC (mixed) can solve the suture needle passing task effectively and efficiently, a comparison against other sampling-based motion planning algorithms is conducted. These algorithms include probabilistic roadmaps with the star strategy (PRM*) [25], [26], rapidly exploring random trees with the star strategy (RRT*) [26], bidirectional fast marching trees (BFMT*) [27], and BIT* [23]. All of these algorithms are implemented on the needle passing environment using The Open Motion Planning Library [28]. These are top performing planners in OMPL. Table II summarizes their respective performances alongside the proposed learning based motion planning strategy in the needle passing environment. The proposed method performs significantly better with regards to success rate and planning time. The planners are also tested in reverse to highlight the speed up and improved success rate when doing so, hence showing why BIT* in reverse, which performed the best of the sampling-based planners, is used to generate demonstrations for the RL training.

C. Real World Experiment

The best trained RL policy is tested in the real world on a da Vinci Research Kit (dVRK) [5] with a suture needle

Algorithm	Success rate	Planning time (s)	Path length (mm)
PRM* (F)	0.66	103.342	5.2
PRM* (R)	0.72	89.9856	5
RRT* (F)	0.18	186.279	5.1
RRT* (R)	0.2	147.6574	4.7
BFMT* (F)	0	—	—
BFMT* (R)	0	—	—
BIT* (F)	0.66	3.8377	5.2
BIT* (R)	0.72	2.5905	5
DDPG+BC (mixed)	0.97	0.0212	4.2

TABLE II: Success rate, average planning time and path length of different motion planning algorithms on the needle passing task (F: forward planning, R: reverse planning)

of radius 5.4mm. The needle is initially grasped in one of the Patient Side Manipulators (PSM) arms from the dVRK using a Large Needle Driver (LND). The end-effector of this PSM arm is the grasping end-effector. The end-effector of another PSM arm with an LND will act as the regrasping end-effector by following the RL policy to regrasp the needle. In order to provide the state information for the RL policy, dVRK’s stereo-endoscope is used which are 1080p and run at 30fps. Both PSM arms are tracked from the stereo-endoscope using our previous work [29] which gives the pose of end-effectors in the camera frame. Let $(\mathbf{p}_{e_g}^c, \mathbf{q}_{e_g}^c)$ and $(\mathbf{p}_{e_r}^c, \mathbf{q}_{e_r}^c)$ be the pose of the real-time tracked grasping and regrasping end-effectors respectively in the camera frame. Then the egocentric state for the regrasping end-effector can be computed as:

$$\mathbf{H}_{e_r}^{e_g}(\mathbf{p}_{e_r}^{e_g}, \mathbf{q}_{e_r}^{e_g}) = \mathbf{H}_{e_g}^c(\mathbf{p}_{e_g}^c, \mathbf{q}_{e_g}^c)^{-1} \mathbf{H}_{e_r}^c(\mathbf{p}_{e_r}^c, \mathbf{q}_{e_r}^c) \quad (11)$$

In our previous work, we also showed effective control of the end-effector in the camera frame [29]. Therefore, each action generated by the policy is converted to a target pose in the camera frame by solving $(\mathbf{p}_{e_r,t}^c, \mathbf{q}_{e_r,t}^c)$ using equations (4), (5), and (11). Then, following our previously developed controller, the end-effector of the regrasping arm in its own base frame is set to the pose $(\mathbf{p}_{e_r,t}^b, \mathbf{q}_{e_r,t}^b)$ via dVRK’s built in inverse kinematic, which is computed by:

$$\mathbf{H}_{e_r,t}^b(\mathbf{p}_{e_r,t}^b, \mathbf{q}_{e_r,t}^b) = \mathbf{H}_c^b(\mathbf{p}_c^b, \mathbf{q}_c^b) \mathbf{H}_{e_r,t}^c(\mathbf{p}_{e_r,t}^c, \mathbf{q}_{e_r,t}^c) \quad (12)$$

where $\mathbf{H}_c^b(\mathbf{p}_c^b, \mathbf{q}_c^b)$ is updated in real-time by our previous surgical tool tracking method [29].

Similarly, the needle is reconstructed using the technique proposed in [30], which gives the pose of the needle in the camera frame. Let $(\mathbf{p}_n^c, \mathbf{q}_n^c)$ be the reconstructed pose of the needle. Combined with the previously described end-effector tracking, the reconstructed needle pose can be transformed to the grasping and regrasping end-effector frames by:

$$\mathbf{H}_n^{e_g}(\mathbf{p}_n^{e_g}, \mathbf{q}_n^{e_g}) = \mathbf{H}_{e_g}^c(\mathbf{p}_{e_g}^c, \mathbf{q}_{e_g}^c)^{-1} \mathbf{H}_n^c(\mathbf{p}_n^c, \mathbf{q}_n^c) \quad (13)$$

$$\mathbf{H}_n^{e_r}(\mathbf{p}_n^{e_r}, \mathbf{q}_n^{e_r}) = \mathbf{H}_{e_r}^c(\mathbf{p}_{e_r}^c, \mathbf{q}_{e_r}^c)^{-1} \mathbf{H}_n^c(\mathbf{p}_n^c, \mathbf{q}_n^c) \quad (14)$$

hence giving the last components of the state information for the policy from the real world. However, we experimentally found the reconstruction to be too inaccurate to directly apply as shown in Fig. 6. This highlights that needle reconstruction is the weak point of

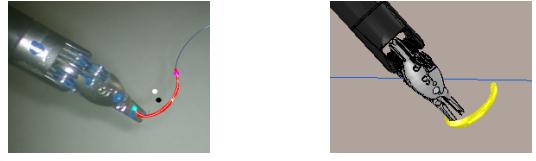


Fig. 6: Detected (left) and reconstructed (right) pose of a needle. Since the end-effector and needle are detected separately, their reconstructed poses might not form a feasible grasp.



Fig. 7: Failure cases of needle passing which are hitting a joint limit (left) and grasped too far inside (right)

transferring the RL policy to the real-world. Therefore, to map the reconstructed pose $(\mathbf{p}_n^{e_g}, \mathbf{q}_n^{e_g})$ to a valid initial grasp, this pose will be compared to a set of 1000 valid initial grasps $\{(\mathbf{p}_n^{e_g}, \mathbf{q}_n^{e_g})_1, (\mathbf{p}_n^{e_g}, \mathbf{q}_n^{e_g})_2, \dots, (\mathbf{p}_n^{e_g}, \mathbf{q}_n^{e_g})_{1000}\}$ collected from the simulated scene beforehand. The pose in this set that is closest to $(\mathbf{p}_n^{e_g}, \mathbf{q}_n^{e_g})$ will be used for the policy. The distance is calculated in the same manner as equation (7) and (8).

In the real-world experiments, the poses of an initial needle grasp and initial end-effectors are manually randomized. The complete success rate for the needle passing is **73.3%** from 15 trials, with an average planning time of 0.0846s and an average run time of 5.1454s. Several example regrasps are shown in Fig. 1. In these trials, the main failures are caused by the needle reconstruction. To test this, a secondary experiment is conducted where the initial grasp is preset to some known position and orientation. During this experiment, the success rate of the needle passing is **90.5%** from 21 trials, with an average planning time of 0.0807s and an average run time of 2.8801s. In this case, the main failures come from the arm reaching a joint limit or needle regrasped too inside of the gripper, which are shown in Fig. 7.

VI. DISCUSSION AND CONCLUSION

In this work, we present a novel method for trajectory generation to conduct suture needle regrasping. The task of regrasping is a critical and time consuming task during RAMIS. It is critical since the suture needle needs to be properly orientated and positioned to conduct an effective throw. Given the proposed work, one can combine it with the litany of work in automatic needle throwing to complete the autonomous suturing procedure. Moving forward, the largest contributor to failed grasps is from inaccurate needle reconstruction. This will be solved by applying Bayesian estimation under constraints [31] where the constraint will be feasible grasping positions and orientations as described in this work.

REFERENCES

- [1] M. Yip and N. Das, *ROBOT AUTONOMY FOR SURGERY*. World Scientific, 2018, ch. Chapter 10, pp. 281–313. [Online]. Available: https://www.worldscientific.com/doi/abs/10.1142/9789813232266_0010
- [2] G. Hubens, H. Coveliers, L. Balliu, M. Ruppert, and W. Vaneerdeweg, “A performance study comparing manual and robotically assisted laparoscopic surgery using the da vinci system,” *Surgical Endoscopy and other interventional techniques*, vol. 17, no. 10, pp. 1595–1599, 2003.
- [3] F. Corcione, C. Esposito, D. Cuccurullo, A. Settembre, N. Miranda, F. Amato, F. Pirozzi, and P. Caiazzo, “Advantages and limits of robot-assisted laparoscopic surgery: preliminary experience,” *Surgical Endoscopy and Other Interventional Techniques*, vol. 19, no. 1, pp. 117–119, 2005.
- [4] F. Richter, R. K. Orosco, and M. C. Yip, “Open-sourced reinforcement learning environments for surgical robotics,” *arXiv preprint arXiv:1903.02090*, 2019.
- [5] P. Kazanzides, Z. Chen, A. Deguet, G. S. Fischer, R. H. Taylor, and S. P. DiMaio, “An open-source research kit for the da vinci® surgical system,” in *2014 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2014, pp. 6434–6439.
- [6] P. Sundaresan, B. Thananjeyan, J. Chiu, D. Fer, and K. Goldberg, “Automated extraction of surgical needles from tissue phantoms,” in *2019 IEEE 15th International Conference on Automation Science and Engineering (CASE)*. IEEE, 2019, pp. 170–177.
- [7] H. Mayer, F. Gomez, D. Wierstra, I. Nagy, A. Knoll, and J. Schmidhuber, “A system for robotic heart surgery that learns to tie knots using recurrent neural networks,” *Advanced Robotics*, vol. 22, no. 13-14, pp. 1521–1537, 2008.
- [8] T. Liu and M. C. Cavusoglu, “Needle grasp and entry port selection for automatic execution of suturing tasks in robotic minimally invasive surgery,” *IEEE Transactions on Automation Science and Engineering*, vol. 13, no. 2, pp. 552–563, 2016.
- [9] K. Watanabe, T. Kanno, K. Ito, and K. Kawashima, “Single-master dual-slave surgical robot with automated relay of suture needle,” *IEEE Transactions on Industrial Electronics*, vol. 65, no. 8, pp. 6343–6351, 2017.
- [10] Y. Gao, S. S. Vedula, C. E. Reiley, N. Ahmadi, 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.
- [11] G. A. Fontanelli, M. Selvaggio, L. R. Buonocore, F. Ficuciello, L. Villani, and B. Siciliano, “A new laparoscopic tool with in-hand rolling capabilities for needle reorientation,” *IEEE Robotics and Automation Letters*, vol. 3, no. 3, pp. 2354–2361, 2018.
- [12] T. Liu and M. C. Çavuşoğlu, “Optimal needle grasp selection for automatic execution of suturing tasks in robotic minimally invasive surgery,” in *2015 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2015, pp. 2894–2900.
- [13] R. C. Jackson and M. C. Çavuşoğlu, “Needle path planning for autonomous robotic surgical suturing,” in *2013 IEEE International Conference on Robotics and Automation*. IEEE, 2013, pp. 1669–1675.
- [14] S. Leonard, K. L. Wu, Y. Kim, A. Krieger, and P. C. Kim, “Smart tissue anastomosis robot (star): A vision-guided robotics system for laparoscopic suturing,” *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 4, pp. 1305–1317, 2014.
- [15] S. Sen, A. Garg, D. V. Gealy, S. McKinley, Y. Jen, and K. Goldberg, “Automating multi-throw multilateral surgical suturing with a mechanical needle guide and sequential convex optimization,” in *2016 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2016, pp. 4178–4185.
- [16] S. A. Pedram, P. Ferguson, J. Ma, E. Dutson, and J. Rosen, “Autonomous suturing via surgical robot: An algorithm for optimal selection of needle diameter, shape, and path,” in *2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2017, pp. 2391–2398.
- [17] C. D’Ettorre, G. Dwyer, X. Du, F. Chadebecq, F. Vasconcelos, E. De Momi, and D. Stoyanov, “Automated pick-up of suturing needles for robotic surgical assistance,” in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 1370–1377.
- [18] V. M. Varier, D. K. Rajamani, N. Goldfarb, F. Tavakkolmoghaddam, A. Munawar, and G. S. Fischer, “Collaborative suturing: A reinforcement learning approach to automate hand-off task in suturing for surgical robots,” in *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, pp. 1380–1386.
- [19] A. H. Qureshi, Y. Miao, A. Simeonov, and M. C. Yip, “Motion planning networks: Bridging the gap between learning-based and classical motion planners,” *IEEE Transactions on Robotics*, 2020.
- [20] G. A. Fontanelli, M. Selvaggio, M. Ferro, F. Ficuciello, M. Vendittelli, and B. Siciliano, “A v-rep simulator for the da vinci research kit robotic platform,” in *BioRob*, 2018.
- [21] A. Nair, B. McGrew, M. Andrychowicz, W. Zaremba, and P. Abbeel, “Overcoming exploration in reinforcement learning with demonstrations,” in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 6292–6299.
- [22] T. Jurgensohn and A. Tamar, “Harnessing reinforcement learning for neural motion planning,” *arXiv preprint arXiv:1906.00214*, 2019.
- [23] J. D. Gammell, S. S. Srinivasa, and T. D. Barfoot, “Batch informed trees (bit*): Sampling-based optimal planning via the heuristically guided search of implicit random geometric graphs,” in *2015 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2015, pp. 3067–3074.
- [24] P. Dhariwal, C. Hesse, O. Klimov, A. Nichol, M. Plappert, A. Radford, J. Schulman, S. Sidor, Y. Wu, and P. Zhokhov, “Openai baselines,” <https://github.com/openai/baselines>, 2017.
- [25] L. E. Kavraki, P. Svestka, J.-C. Latombe, and M. H. Overmars, “Probabilistic roadmaps for path planning in high-dimensional configuration spaces,” *IEEE transactions on Robotics and Automation*, vol. 12, no. 4, pp. 566–580, 1996.
- [26] S. Karaman and E. Frazzoli, “Sampling-based algorithms for optimal motion planning,” *The international journal of robotics research*, vol. 30, no. 7, pp. 846–894, 2011.
- [27] J. A. Starek, J. V. Gomez, E. Schmerling, L. Janson, L. Moreno, and M. Pavone, “An asymptotically-optimal sampling-based algorithm for bi-directional motion planning,” in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2015, pp. 2072–2078.
- [28] I. A. Şucan, M. Moll, and L. E. Kavraki, “The Open Motion Planning Library,” *IEEE Robotics & Automation Magazine*, vol. 19, no. 4, pp. 72–82, December 2012, <https://ompl.kavrakilab.org>.
- [29] Y. Li, F. Richter, J. Lu, E. K. Funk, R. K. Orosco, J. Zhu, and M. C. Yip, “Super: A surgical perception framework for endoscopic tissue manipulation with surgical robotics,” *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 2294–2301, 2020.
- [30] D. Lo, P. R. Mendonça, A. Hopper *et al.*, “Trip: A low-cost vision-based location system for ubiquitous computing,” *Personal and Ubiquitous Computing*, vol. 6, no. 3, pp. 206–219, 2002.
- [31] D. Simon, “Kalman filtering with state constraints: a survey of linear and nonlinear algorithms,” *IET Control Theory & Applications*, vol. 4, no. 8, pp. 1303–1318, 2010.