Framework for soft tissue manipulation and control using Deep Reinforcement Learning

Ameya Pore[1], Eleonora Tagliabue[1], Diego Dall'Alba[1], Paolo Fiorini[1]

[1]*Department of Computer Science, University of Verona, Italy,*
ameya.pore@univr.it

INTRODUCTION

The future generation of surgical robotic systems will provide an increased level of autonomy. All minimally invasive surgical (MIS) procedures are composed of a series of repetitive subtasks like dissection, suturing and knot tying, that extensively deal with the manipulation of soft tissues. Automation of these subtasks would bring several benefits to MIS, among which enhanced precision in tool control and increased dexterity, which would eventually lead to improved operator comfort. Some attempts to automate these repetitive sub-tasks have been already proposed [1]. However, these approaches use hand-crafted control policies which work well in laboratory conditions but perform poorly within the highly-variable surgical environment. Hence, autonomous surgical systems should be able to manipulate highly deformable and dynamic soft tissues in a robust and efficient way.

Deep Reinforcement Learning (DRL) offers a good framework to model such hard-to-engineer behaviours [2]. DRL has shown promising success in industrial robotic systems to extract useful feature representations from images and learn to perform an optimal action [2]. Most of the works based on DRL for soft-tissue manipulation have focused on 2D deformations such as clothing articles [1], [3]. To tackle issues such as clinical safety, ethics, economical and hardware risks, various studies have proposed to train autonomous agents in a simulated environment. Then the learned policies can be transferred to a real system based on a sim-to-real approach [4].

In the context of clinical applications, soft tissues are usually simulated using finite element models (FEM) due to their high accuracy in simulating soft tissues dynamics [5]. However, high accuracy implies high computational time, which makes it impractical to employ FEMs in DRL framework, where a huge number of trial and error attempts are needed. A possible alternative to FEM is represented by non-physics based methods, such as Position-Based Dynamics (PBD). PBD has been demonstrated to model anatomical tissues deformations to an acceptable degree of accuracy while keeping the computation time low [6].

In this work, we introduce a soft-tissue simulation framework that replicates the preliminary steps of partial nephrectomy procedure. Furthermore, we show that an end-to-end reinforcement learning algorithm can be trained in the simulation without any user demonstration to accomplish a tissue manipulation task. To the best of our knowledge, this is one of the first attempts of using DRL agents to manipulate soft tissues for autonomous surgical action execution.

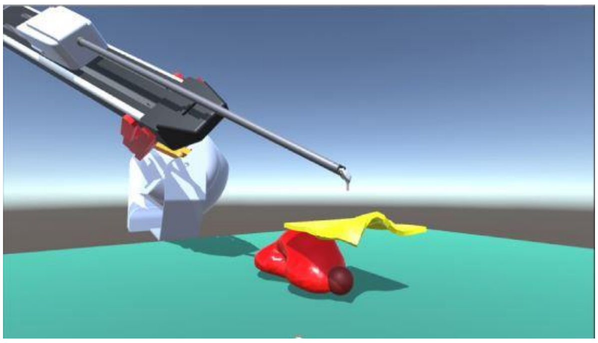

Figure. 1. Environment scene developed in Unity3D for robot-assisted nephrectomy procedure. The scene consists of a single PSM arm, kidney, tumour and renal fascia (fat tissue). For visual simplicity, high contrast colours are used to depict various objects

MATERIALS AND METHODS

Simulation details: Our simulation environment relies on the Unity3D engine (Fig 1). In addition to the high flexibility and modularity offered by this development platform, which allows adding custom objects and sensors to the scene, Unity enables the modelling of soft objects exploiting the efficient PBD implementation provided by NVIDIA FleX. All the deformable structures present in our environment (kidney, tumour and surrounding fat tissue) have been modelled using FleX, and deformation parameters have been tuned with an ad-hoc experiment to provide realistic behaviour [6]. In the proposed virtual environment, we modelled a single slave arm of the da Vinci Research Kit (dVRK) setup, that is, a Patient Side Manipulator (PSM) equipped with a Large Needle Driver (LND) instrument having jaw gripper to grasp objects. We implemented a closed-form inverse kinematics of the PSM to enable the Cartesian space control of the manipulator. We define the tissue grasping condition based on proximity; in particular, we check when the distance between LND tooltip and the fat surface is less than 2mm. We set a bound to the maximum limit of movement of the end-effector (EE) by a predefined delta movement Δ parameter.

Reinforcement learning: RL is a trial and error optimisation technique that defines a Markov decision process (MDP) where an agent learns by interacting with the environment. An MDP is characterised by a state space S, action space A, and a reward function R : $S \times A \rightarrow R$ that encapsulates the goal. At each time step t, the environment produces a state observation $s_t \in S$, the agent then samples an action $a_t \sim \mu(s_t)$, $a_t \in A$ and $\mu: S \rightarrow A$ represents a behaviour policy. As a consequence, the environment then yields a reward r_t sampled from R and the agent moves to a new state s_{t+1} sampled from a transition probability function

$P(S' = s_{t+1}|S = s_t, A = a_t)$. The agent's goal is to maximise the expected future discounted reward:

$$G_t = E_{s_{0\to T}, a_{0\to T-1}, R_{0\to T-1}}\left[\sum_{i=0}^{T-1} \gamma^i R_i\right]$$

where T is the time horizon and $\gamma \in [0,1]$ is the discount factor.

Experiment: The experiment aim is to move the PSM arm from an initial position P_0 to a position close to the tumour T_0, grasp the fat and retract it to a predefined target position P_F. The state and action space of the environment at timestep t is $S_t = [P_t, P_F, T_0, G]$, $A_t = [-1,0,1] * \Delta$, where P_t, T_0, G are the position of the EE, tumour and gripper state, respectively. The reward function is:

$$\mathrm{r}(s_t) = \begin{cases} -d(P_t, T_0) * k & if \text{ gripper close} \\ -d(P_t, P_f) * \mathrm{k} & \text{if gripper open} \end{cases}$$

Where d denotes the distance function, and k is a constant that depends on the volume in which PSM can move. The gripper state changes automatically when a grasping event is detected. Each episode consists of 1500 steps. We use Proximal Policy Optimisation (PPO), which is the state-of-the-art DRL algorithm supported by MLAgents framework provided by Unity [7].

RESULTS

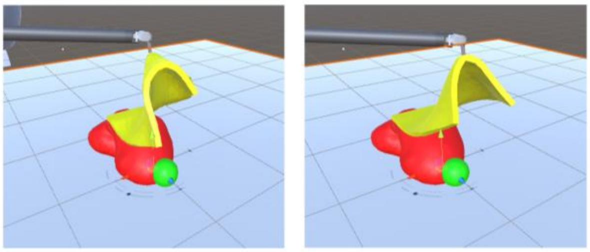

Figure. 2. Learnt policy from two different starting position of the PSM arm.

Fig 2. highlights the final state of the learnt behaviour where the PSM arm approaches, grasps the fat tissue and consequently exposes the underlying tumour. The figure depicts the dynamic and flexible nature of the soft tissue that is rendered in the simulator.

Fig 3. shows the learning curve, that is, the number of steps taken to reach a higher cumulative reward. The agent requires 3 million steps to learn the manipulation task. This result is comparable to the training steps required to grasp rigid objects [8]. From the reward trend, it emerges that the agents require 500 thousand steps to learn the approach behaviour towards the fat, the subsequent 1.5 million steps to learn the interaction with the fat and finally 1 million steps to learn the retract behaviour. Note that the behaviour is learnt without any visual cues and purely on the basis of positional information. The training is carried out on an NVIDIA TitanX GPU with 16 parallel workers and takes 3hrs in real-time.

CONCLUSION AND DISCUSSION

This work is a preliminary step towards developing autonomous agents for controlling surgical tools and manipulating soft tissues. Future work would be focused on autonomous control using visual cues and developing a metric to evaluate the trajectory/behaviour learnt. We

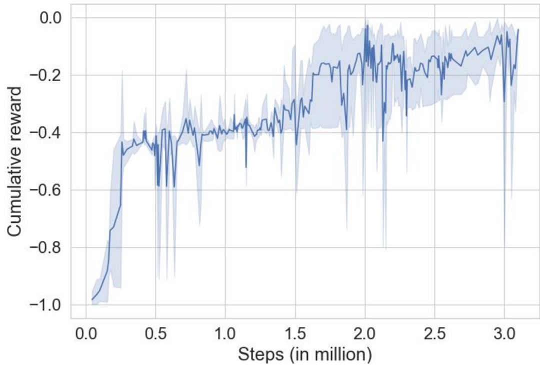

Figure. 3. The obtained learning curve. The cumulative reward is normalised in the range $[-1,0]$. The shaded area spans the range of values obtained when training the agent starting from three different initialisation seeds.

plan to extend the work towards using model-based approaches and incorporating imitation based learning using expert demonstration to bootstrap the number of training steps. Subsequent directions would be carried out to transfer the learned knowledge to a real-robotic system.

ACKNOWLEDGEMENTS

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 742671 "ARS") and under the Marie Sklodowska-Curie (grant agreement No. 813782 "ATLAS"). We acknowledge the support of NVIDIA Corp. for the donation of the Titan X GPU.

REFERENCES

[1] Murali, Adithyavairavan, et al. "Learning by observation for surgical subtasks: Multilateral cutting of 3d viscoelastic and 2d orthotropic tissue phantoms." *2015 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2015.

[2] Arulkumaran, Kai, et al. "Deep reinforcement learning: A brief survey." *IEEE Signal Processing Magazine* 34.6 (2017): 26-38.

[3] Tsurumine, Yoshihisa, et al. "Deep reinforcement learning with smooth policy update: Application to robotic cloth manipulation." *Robotics and Autonomous Systems* 112 (2019): 72-83.

[4] Peng, Xue Bin, et al. "Sim-to-real transfer of robotic control with dynamics randomisation." *2018 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2018.

[5] Zhang, Jinao, Yongmin Zhong, and Chengfan Gu. "Deformable models for surgical simulation: a survey." *IEEE reviews in biomedical engineering* 11 (2017): 143-164.

[6] Tagliabue, Eleonora, et al. "Position-based modeling of lesion displacement in Ultrasound-guided breast biopsy." *International journal of computer assisted radiology and surgery* 14.8 (2019): 1329-1339.

[7] Juliani, Arthur, et al. "Unity: A general platform for intelligent agents." *arXiv preprint arXiv:1809.02627* (2018).

[8] Richter, Florian, Ryan K. Orosco, and Michael C. Yip. "Open-sourced reinforcement learning environments for surgical robotics." *arXiv preprint arXiv:1903.02090* (2019).