UnityFlexML: Training Reinforcement Learning Agents in a Simulated Surgical Environment*

Eleonora Tagliabue*, Ameya Pore*, Diego Dall'Alba*, Marco Piccinelli* and Paolo Fiorini*
*Dept. Computer Science, University of Verona, Verona, Italy
email: *eleonora.tagliabue@univr.it*

***Abstract*—Sim-to-real Deep Reinforcement Learning (DRL) has shown promising in subtasks automation for surgical robotic systems, since it allows to safely perform all the trial and error attempts needed to learn the optimal control policy. However, a realistic simulation environment is essential to guarantee direct transfer of the learnt policy from the simulated to the real system. In this work, we introduce *UnityFlexML*, an open-source framework providing support for soft bodies simulation and state-of-the-art DRL methods. We demonstrate that a DRL agent can be successfully trained within *UnityFlexML* to manipulate deformable fat tissues for tumor exposure during a nephrectomy procedure. Furthermore, we show that the learned policy can be directly deployed on the da Vinci Research Kit, which is able to execute the trajectories generated by the DRL agent. The proposed framework represents an essential component for the development of autonomous robotic systems, where the interaction with the deformable anatomical environment is involved.**

***Index Terms*—Deformable simulation; Sim-to-real reinforcement learning; Autonomous robotic surgery**

I. Introduction

Recent trends in surgical robotics have focused on automation of some common subtasks that take place in many different procedures, to enhance precision and reduce surgeons' workload. The main challenge to face during surgical task automation relies on accounting for the dynamic and deformable behavior of the anatomical environment, which makes the design of model-based control policies prone to failure [1]. Explicit modelling of the highly deformable anatomical environments can be avoided using a Learning From Demonstrations (LfD) approach, where the robot learns the task from a set of expert demonstrations [2]. However, collecting a wide enough dataset of example trajectories is often impractical in real clinical settings. Deep Reinforcement Learning (DRL) represents an alternative approach to the automation of robotic tasks, without the need to design ad-hoc control strategies [3]. The main drawback of model-free DRL is the huge amount of trial and error attempts required to learn even simple behaviors, which makes its direct application to real robotic system unfeasible. To tackle this issue, autonomous agents can be trained in a realistic simulation of the environment and the learned policies can be later transferred to a real system [4]. Sim-to-real DRL seems particularly promising for robotic surgery, since it allows to avoid all the limitations related to the organization of clinical data collection. However, to exploit sim-to-real methods in the surgical robotic field, it is essential to have realistic and fast simulation environments where to prototype and test the algorithms. In this work, we present *UnityFlexML*, an open-source framework to train DRL methods in simulated surgical environments which involve deformable objects. We further show that *UnityFlexML* can be successfully employed to train an end-to-end reinforcement learning algorithm to accomplish a tissue manipulation task, which consists in the retraction of perirenal fat tissue to expose the underlying kidney tumor, during robotic assisted partial nephrectomy (Fig.1). Eventually, we demonstrate that the learned policy translates directly to the real surgical robotic system thanks to the da Vinci Research Kit (dVRK).

*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").

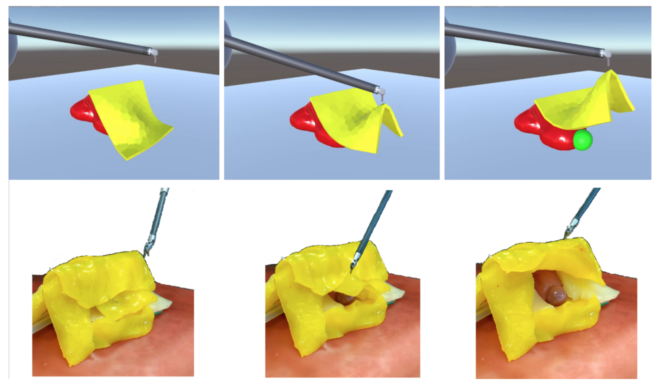

Fig. 1. Sequence of action frames for task completion in *UnityFlexML* (top) and real environment (bottom). From left to right: approach, grasp, retract.

II. Method

A. *UnityFlexML*

UnityFlexML[1] is based on Unity3D, a general and modular development platform which allows to easily edit and customize the virtual scene. It relies on two main Unity plugins: the Machine Learning Agents Toolkit (ML-Agents), for training intelligent agents, and NVIDIA FleX, for modelling soft objects. In particular, NVIDIA FleX provides a highly optimized implementation of the Position Based Dynamics (PBD) simulation approach, which has already proved able to model soft tissue deformations both realistically and fast

[1] https://gitlab.com/altairLab/unityflexml

[5]. In addition, *UnityFlexML* implements the simulation of dVRK slave arms, called Patient Side Manipulators (PSMs). We provide a closed form inverse kinematics of the PSM to enable the Cartesian space control of the manipulator and implementation of the grasping as an atomic event triggered when the relative distance with the end-effector (EE) is less that $2\,mm$.

B. Learning Soft Tissue Retraction

A DRL algorithm based on Proximal Policy Optimisation (PPO) is trained within *UnityFlexML* to perform fat tissue retraction and expose the underlying kidney tumor (Fig.1). The specific task to learn consists in moving the PSM arm from an initial position $\mathbf{p}_0$ to a position close to the tumour $\mathbf{t}_0$, grasp the tissue and lift it to a predefined target position $\mathbf{p}_F$ (Fig.1). The considered state space leverages on kinematics information only, defining the current robot and the environment states. The action space is defined by an increment motion of $0.5\,mm$ in each spatial dimension. The reward function is based on the distance between the current PSM position and the tumor, promoting the approach behaviour towards the tumour before the tissue is grasped and tissue retraction to a fixed target point after grasping.

III. Experiments

Our experimental setup consists of a synthetic kidney phantom surrounded by silicone fat tissue (Fig.1, bottom line). The agent is first trained on a realistic virtual replica of the real scenario (Fig.1, top line), where the fat tissue deformation parameters have been optimized to minimize the discrepancy between simulation and reality [6]. To evaluate if the task has been successfully learnt, we assess tumor exposure (TE) once the agent has reached $\mathbf{p}_F$. TE is computed as percentage of tumor pixels seen from a simulated endoscope placed in front of the kidney in simulation. Afterwards, the trained policies are transferred and tested on the real system. Policy transfer is evaluated by counting the number of successful task executions with the PSM arm starting from 25 different initial positions distributed on a regular grid above the real fat tissue.

IV. Results

The agent takes 3 million steps to learn the task in simulation, divided into 500 thousand, 1.5 million and 1 million steps to learn the approach behavior towards the fat, the interaction with the fat and the retract behavior respectively. Fig.2a shows the obtained TE in simulation when considering different $\mathbf{p}_0$. Whenever the agent starts from the farthest part of the tissue from the fixed region, the tumor becomes at least partially visible from the camera, demonstrating that the learned policy can account for different initial EE configurations. When transferring the learn policy to the real system, the PSM successfully gets in contact with the fat tissue for all the different initial positions and it is always able to reach the target point. However, Fig.2b shows that the PSM is able to grasp the fat tissue in only 9 cases out of 25 (plus 2 cases when the tissue is grasped but lost during the task). These results suggest that the tissue is more likely to be grasped when the EE starts close to the tumor area. The presence of failed grasps is due to the inaccurate modelling of the grasping action in the simulated environment, where grasping is triggered based on the relative distance between the tissue and the EE. If the same condition could be precisely replicated in the real setup, the total number of accomplished grasps would increase, due to the fact that the PSM is able to touch the fat surface in all the attempts.

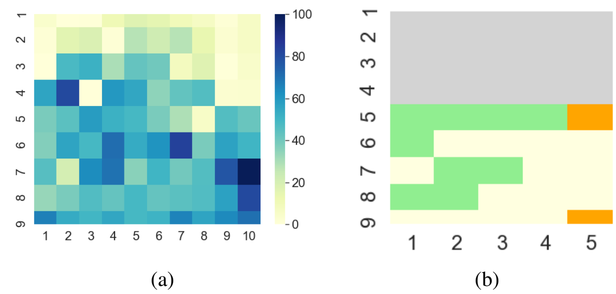

Fig. 2. (a) Simulation results: color of each cell represents the TE from the simulated camera at different $\mathbf{p}_0$. (b) Real world results: color of each cell represents the outcome of the grasping task performed with the dVRK at different $\mathbf{p}_0$. Green, orange and yellow stand for successful, semi-successful and failed grasps, respectively. Gray region is not considered for the experiments.

V. Conclusion

We present *UnityFlexML*, a flexible simulation environment suitable for DRL training in surgical robotic applications. The developed framework demonstrates adapt for training an autonomous agent to perform fat tissue manipulation for the exposure of tumor during robotic assisted nephrectomy procedure. We demonstrate that a DRL agent trained in simulation is able to generalize to real scenario. The proposed framework is an essential component in development of autonomous agents for controlling surgical tools and manipulating soft tissues. Future work would be focused on autonomous control using visual cues.

References

[1] E. Tagliabue, A. Pore, D. Dall'Alba, E. Magnabosco, M. Piccinelli, and P. Fiorini, "Soft tissue simulation environment to learn manipulation tasks in autonomous robotic surgery," in *2020 IEEE International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020.

[2] A. Murali, S. Sen, B. Kehoe, A. Garg, S. McFarland, S. Patil, and al., "Learning by observation for surgical subtasks: Multilateral cutting of 3d viscoelastic and 2d orthotropic tissue phantoms," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2015, pp. 1202–1209.

[3] J. Kober, J. A. Bagnell, and J. Peters, "Reinforcement learning in robotics: A survey," *The International Journal of Robotics Research*, vol. 32, no. 11, pp. 1238–1274, 2013.

[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] E. Tagliabue, D. Dall'Alba, E. Magnabosco, C. Tenga, I. Peterlik, and P. Fiorini, "Position-based modeling of lesion displacement in ultrasound-guided breast biopsy." *IJCARS*, 2019.

[6] M. Piccinelli, E. Tagliabue, D. Dall'Alba, and P. Fiorini, "Vision-based estimation of deformation properties for autonomous soft tissues manipulation," in *10th Joint Workshop on New Technologies for Computer/Robot Assisted Surgery*, 2020.