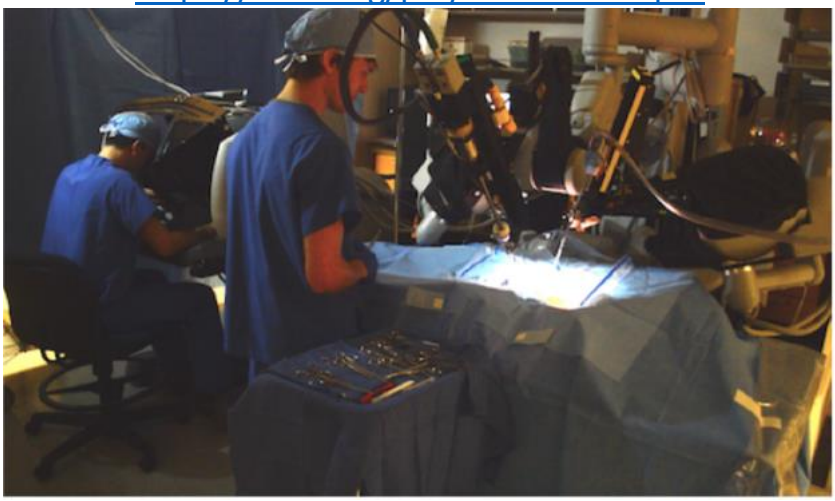
Open-Sourced Reinforcement Learning Environments for Surgical Robotics

https://arxiv.org/pdf/1903.02090.pdf

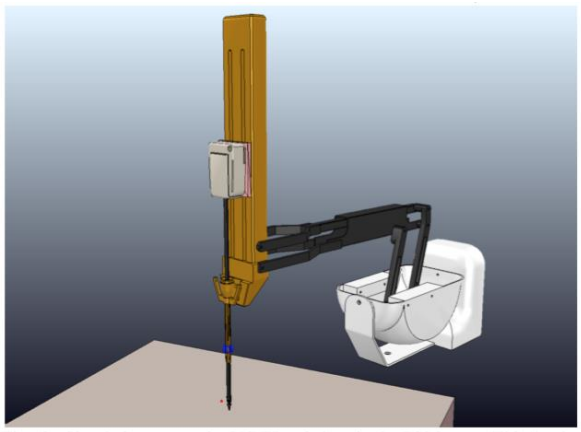


Problem and relevance:

Recent years has seen a surge of successes solving challenging games and smaller domain problems, including simple though non-specific robotic manipulation and grasping tasks. Rapid successes in RL have come in part due to the strong collaborative effort by the RL community to work on common, open-sourced environment simulators such as OpenAI's Gym that allow for expedited development and valid comparisons between different, state-of-art strategies. In this paper, authors aim ***to bridge the RL and the surgical robotics communities*** by presenting the first open-sourced reinforcement learning environments for surgical robotics, called dVRL. The authors ***present the first, open-sourced reinforcement learning environment for surgical robotics called dVRL***. dVRL provides a syntatically common RL environments to OpenAI Gym with a simulation of the da Vinci Surgical Robot system.

Enviroment:

The environments presented inherit from the OpenAI Gym Environments and utilize the V-REP physics simulator. the environments only utilize one slave arm from the da Vinci Surgical System, also known as a Patient Side Manipulator (PSM) arm. The current environments use the Large Needle Driver (LND). The environment contains 3 or for actions and 3 + has_object or 4 + has_object states where has_object is a boolean flag whether or not the environment has an object.



Solution:

The environments are solved in simulation using Deep Deterministic Policy Gradients (DDPG). DDPG is from the class of Actor-Crtic algorithms where it

approximates both the policy and Q-Function with separate neural networks. The Q-Function is optimized by minimizing the Bellman loss error:

$$\mathcal{L}_Q = (Q(s_t, a_t) - (r_t + \gamma Q(s_{t+1}, a_{t+1}))^2$$

and the policy is optimized by minimizing:

$$\mathcal{L}_\pi = -\mathbb{E}_{s_t}[Q(s_t, \pi(s_t)]$$

The reward function is different for *PSM Reach Environment and PSM Pick Environment, but they represent wether the arm is in the area with radius po around an item.*

Hindsight Experience Replay (HER) is used as well to generate new experiences for faster training. HER generates new experiences for the optimization of the policy and/or Q-Function where the goal portion of the state is replaced with previously achieved goals. This improves the sample efficiency of the algorithms.

The size of the state space relative to the distance the maximum action is very large in the presented environments. This makes exploration very challenging, especially for the PSM Pick environment. To overcome this, demonstrations $\{(s_i^d, a_i^d)\}_{i=0}^{N_d}$ which reach the goal, are generated in simulation and the behavioral cloning loss $\mathcal{L}_{BC} = \sum_{i=0}^{N_d} ||\pi(s_i^d) - a_i^d||^2$ is augmented with the DDPG policy loss. OpenAI Baselines implementation and hyper parameters of DDPG + HER, with the addition of the augmented behavioral cloning, was used.

Experiments:

State of the art RL algorithms are utilized to solve the environments in simulation. The learned polices are then transferred to the real da Vinci Surgical System using the da Vinci Research Kit (dVRK) running at 50Hz. The policy transfer is evaluated individually by replicating the simulated scene and completion of the surgical: tasks suction and debris removal. The training of the RL policies and dVRK ran on an Intel Core i9-7940X Processor and NVIDIA's GeForce RTX 2080.

Authors model of endoscopic stereo cameras with their uniquely tight disparities and narrow field of view would allow for visual servoing and visuomotor policy approaches to be explored. Learned policies are furthermore successfully transferable to a real robot. Finally, combining dVRL with the over 40+ network of da Vinci Surgical Research Kits in active use at academic institutions, we see dVRL as enabling the broad surgical robotics community to fully leverage the newest strategies in reinforcement learning, and for reinforcement learning scientists with no knowledge of surgical robotics to test and develop new algorithms that can solve the realworld, high-impact challenges in autonomous surgery.

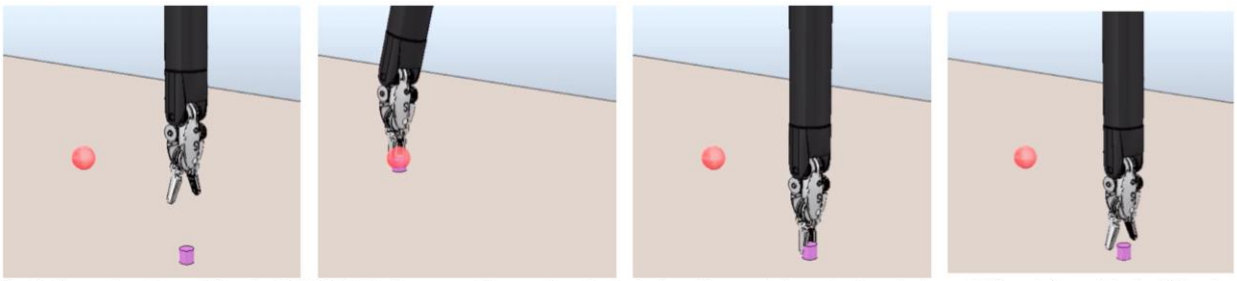


Fig. 3: Example policy solving the PSM Pick Environment. The purple cylinder is the object, and the red sphere is the goal. From left to right the following is done: move to the object, grasp the object, transport the object to the goal.

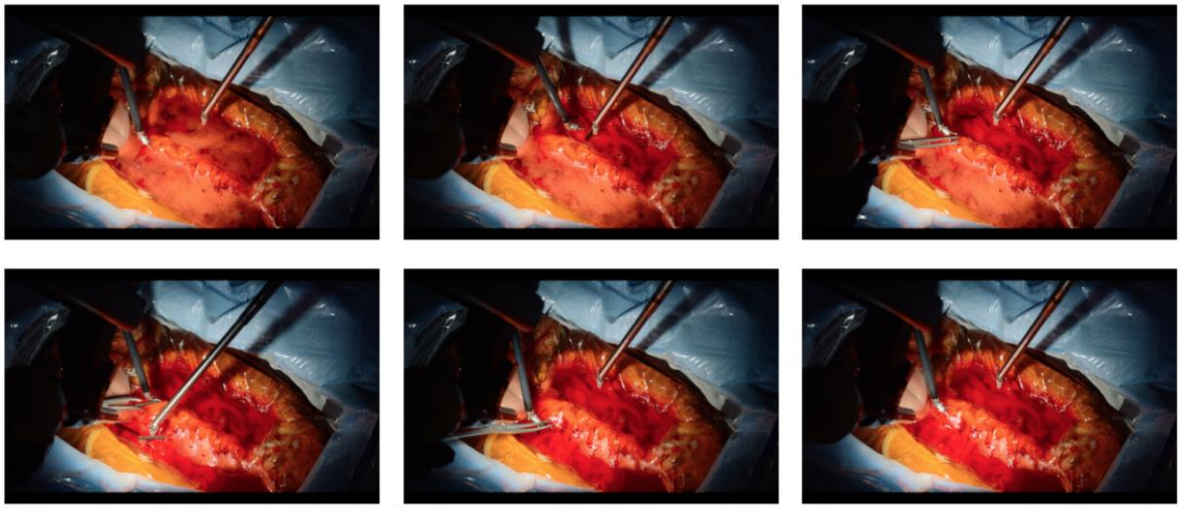


Fig. 8: The suction tool using a trained PSM Reach policy to remove fake blood to reveal debris so the surgeon can remove them from a simulated abdomen. After located and removed by teleoperational control from the simulated abdomen, the debris is handed off to the first assistant.

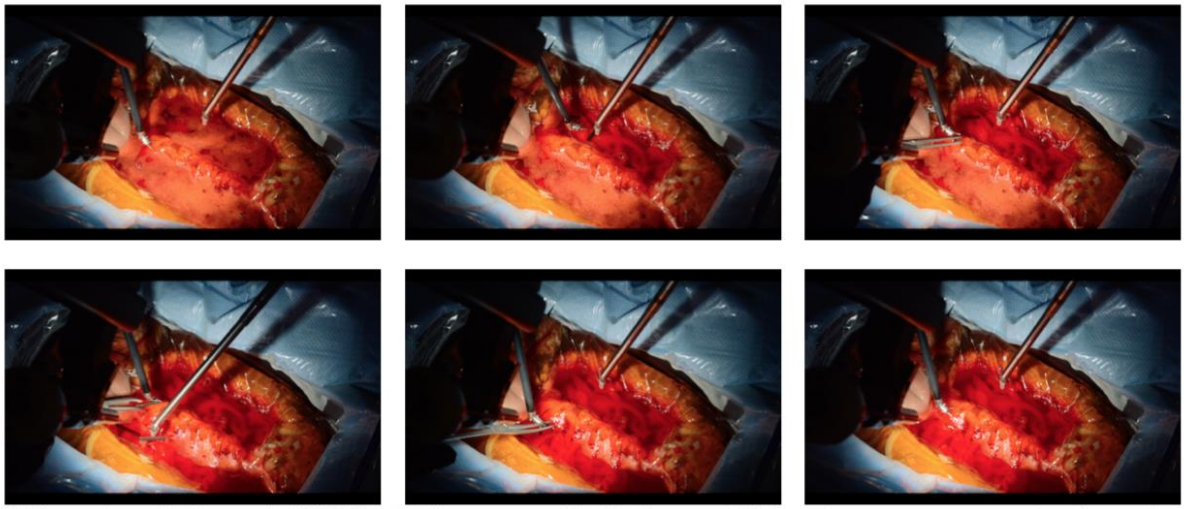


Fig. 8: The suction tool using a trained PSM Reach policy to remove fake blood to reveal debris so the surgeon can remove them from a simulated abdomen. After located and removed by teleoperational control from the simulated abdomen, the debris is handed off to the first assistant.