A reinforcement learning approach to robotic suturing

Background Reading

[2] 2021-2022 AccelNet Surgical Robotics Challenge

Group 20
Members: Walee Attia, Jocelyn Hsu, Jihoon Kim
Mentors: Dr. Anqi Liu, Dr. Adnan Munawar, Dr. Manish Sahu, Dr. Peter Kazanzides

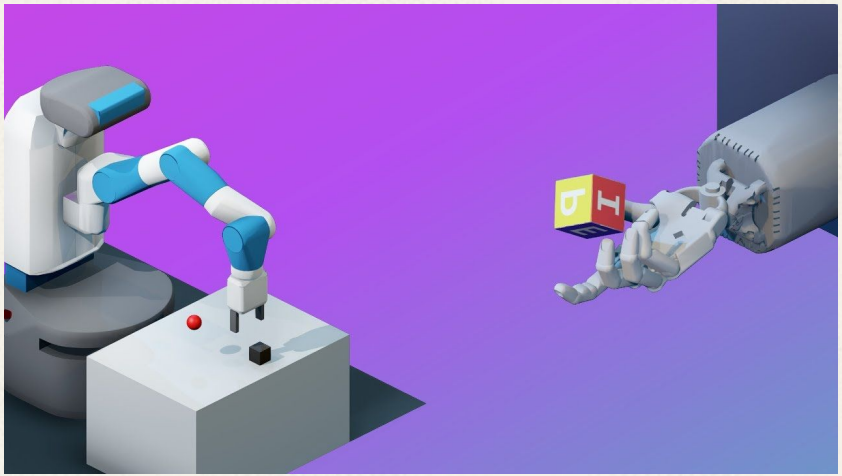
PROJECT OVERVIEW

Reinforcement Learning (RL)

RL techniques have seen significant progress in the robotics domain; however, there exist a lack of platforms which offer environments conducive to medical robotics.

OpenAI Gym

OpenAI Gym is an open-source RL framework that offer realistic simulation environments with easy integration.



[3] Gymnasium documentation

Surgical Robotics Challenge (SRC)

A simulation platform to develop algorithms to address various questions in surgical robotics automation with:

- Two 7-DOF instrument arms based on the da Vinci Surgical System large needle driver
- Controllable camera based on the da Vinci Endoscopic Camera Manipulator
- Suturing phantom
- Needle with a suture

Asynchronous Multibody Framework (AMBF)

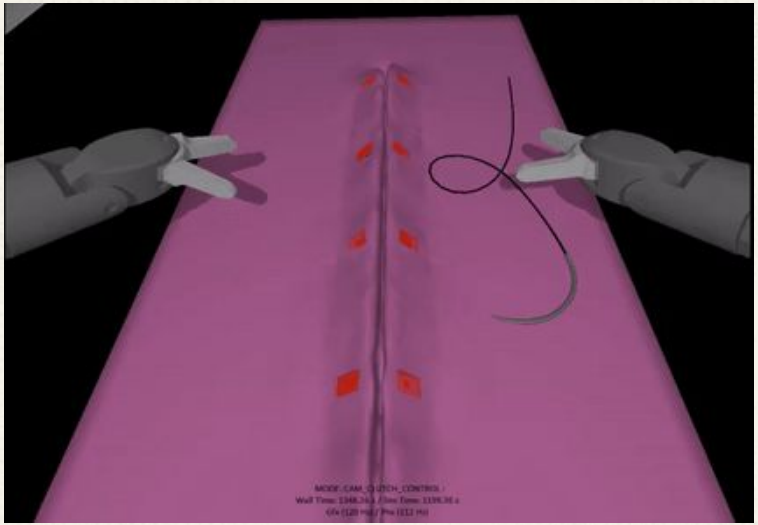
A real time dynamics simulator that serves as the backbone for the Surgical Robotics Challenge environment

[2] 2021-2022 AccelNet Surgical Robotics Challenge

BACKGROUND

Goal

- Building a Reinforcement Learning platform addressing Challenge 2
  - Grasp needle and drive through tissue



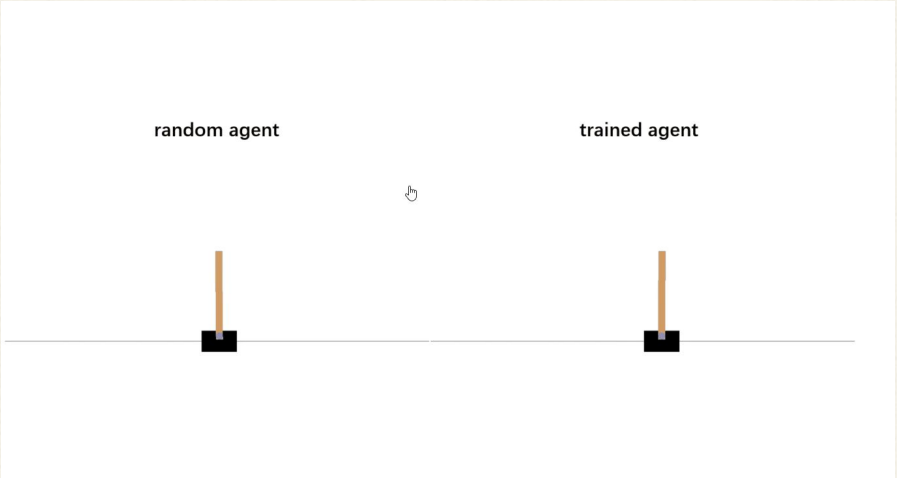
[2] 2021-2022 AccelNet Surgical Robotics Challenge

REINFORCEMENT LEARNING

[7] Introduction to reinforcement learning with David Silver.

REINFORCEMENT LEARNING (RL)

- A type of machine learning where an agent learns to make decisions by interacting with an environment to achieve a goal.
- Balances **exploration** (trying new actions) and **exploitation** (choosing the best-known action).

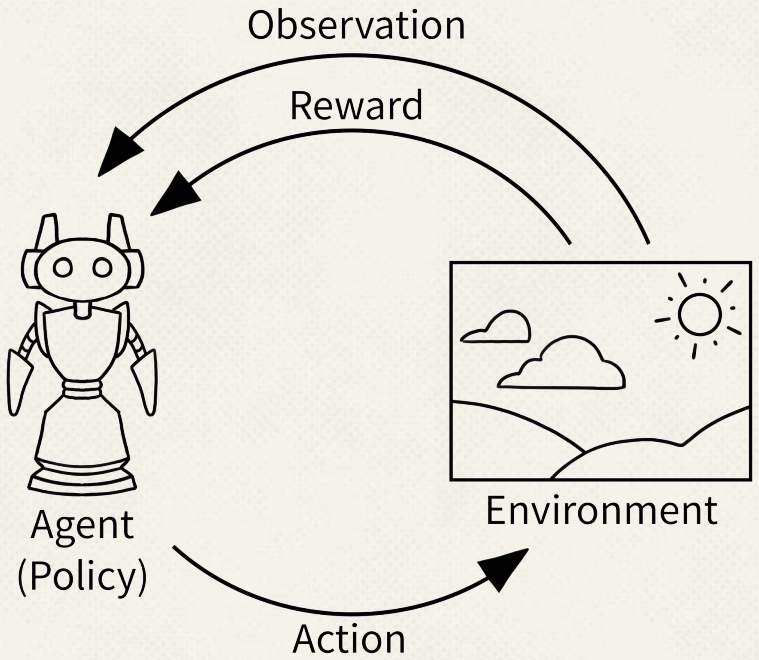


[3] Gymnasium documentation

RL KEY TERMS

Key Components

- **Agent**: The decision-maker that learns to interact with the environment.
- **Environment**: The world the agent interacts with, providing feedback based on the agent's actions.
- **State**: A representation of the current situation in the environment.
- **Action**: The possible moves or decisions the agent can make in a given state.
- **Reward**: Immediate feedback provided by the environment based on the agent's action.

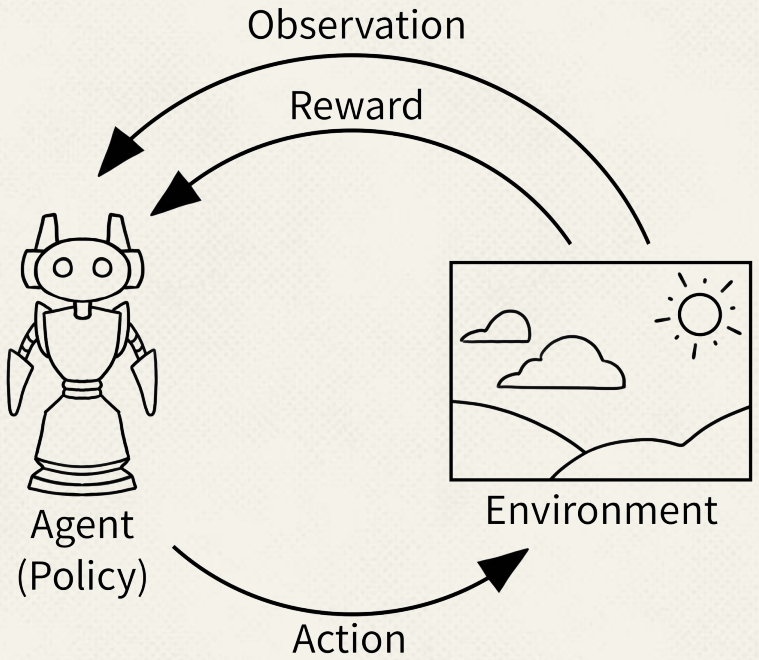


[7] Introduction to reinforcement learning with David Silver.

RL KEY TERMS

Learning Process

- The **agent learns** a **policy** (mapping from states to actions) to maximize the cumulative reward.
- The policy can be **deterministic** (choosing a specific action) or **stochastic** (selecting actions based on probabilities).
- **Q-value function:** Predicts expected reward for state-action pairs, guiding optimal decisions.



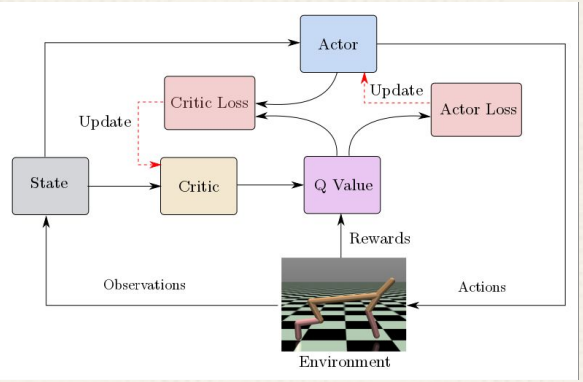
[7] Introduction to reinforcement learning with David Silver.

Paper 1: Deep Deterministic Policy Gradient + Hindsight Experience Replay

[6] M. Andrychowicz, F. Wolski, A. Ray, J. Schneider, R. Fong, P. Welinder, B. McGrew, J. Tobin, O. Pieter Abbeel, and W. Zaremba, Hindsight experience replay," Advances in neural information processing systems, vol. 30, 2017

Deep Deterministic Policy Gradient (DDPG)

- Technique for training RL agents in continuous control problems
- Deep neural networks for enabling learning of complex, nonlinear functions
- **Key components:**
  - Actor network: selects actions based on current state *(Policy Function)*
  - Critic network: estimating the expected future rewards for state-action pairs (*Q-value function)*
  - Replay Buffer: Stores experiences to enable batch learning
- **Training:**
  - Actor network: deterministic policy gradient
    - Improve decision making by learning to maximize Q-value function
  - Critic network: temporal difference (TD) learning
    - Target Q-value = Reward + (Discount factor * Value of next state)
    - Loss = (Predicted Q-value - Target Q-value)^2



[5] P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning,"

Hindsight Experience Replay (HER)

Overview

- HER is a technique that improves upon DDPG
- Main Idea: learn from failed experiences by treating them as successful

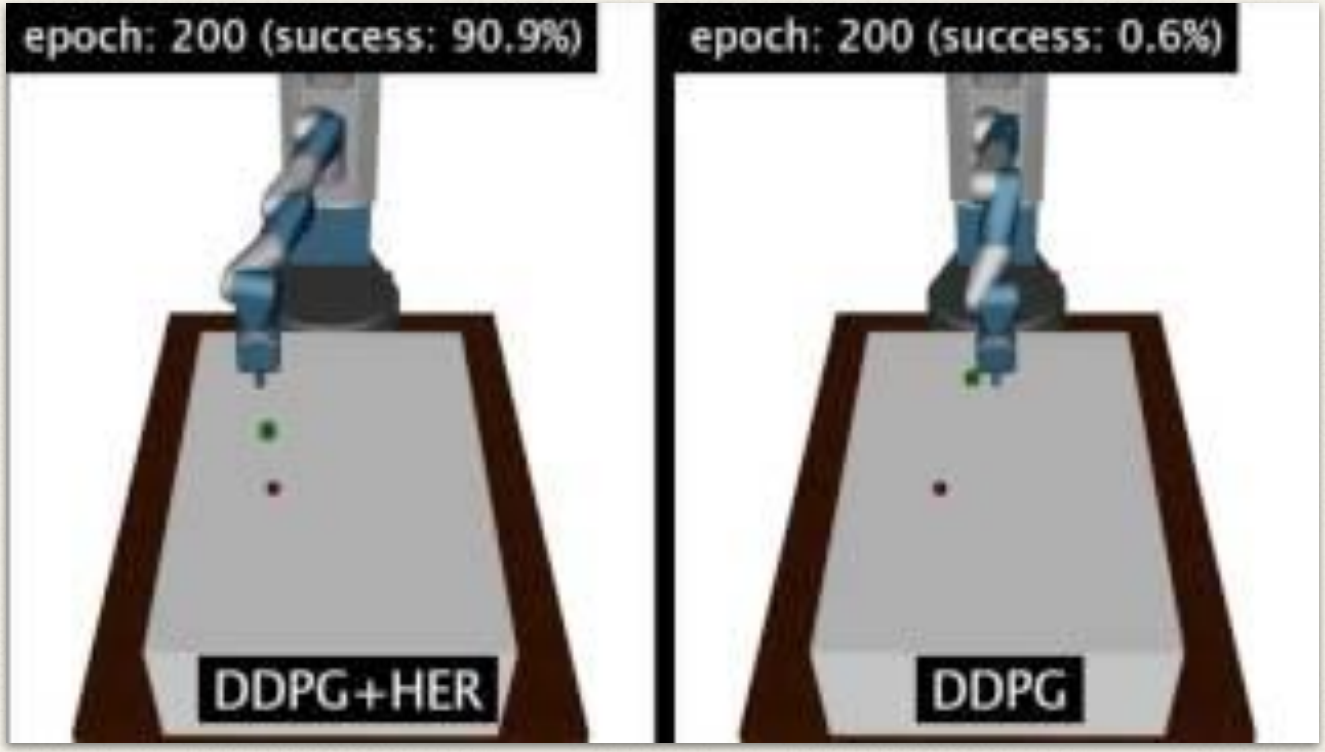
Key Concepts

- Sparse Rewards: the agent receives meaningful feedback only when it achieves the goal, making it difficult to learn from intermediate steps.
- Goals: The desired outcomes that the agent aims to achieve in the environment.

How HER Works

a. When an episode ends, the agent stores the experience tuples (state, action, reward, next state, goal) in the replay buffer.
b. The hindsight goal is chosen based on the actual outcome of the episode
c. Agent updates its model using the original and hindsight experiences.

Hindsight Experience Replay (HER)



Hindsight Experience Replay (HER)

Benefits of HER

- Improves sample efficiency by learning from failed experiences.
- Accelerates learning in tasks with sparse rewards by providing more diverse and informative feedback.
- Enables the agent to generalize its learning across multiple goals

Drawbacks of HER

- Increases computational complexity
- Requires discrete goal space, where reward algorithm are binarized and cannot model all environments
- Tune additional hyperparameters, such as ratio of hindsight to original experiences

MEDICAL APPLICATIONS OF RL

Paper 2: AMBF-RL

[4] A. Munawar, Y. Wang, R. Gondokaryono, and G. S. Fischer, “A real-time dynamic simulator and an associated front-end representation format for simulating complex robots and environments.”

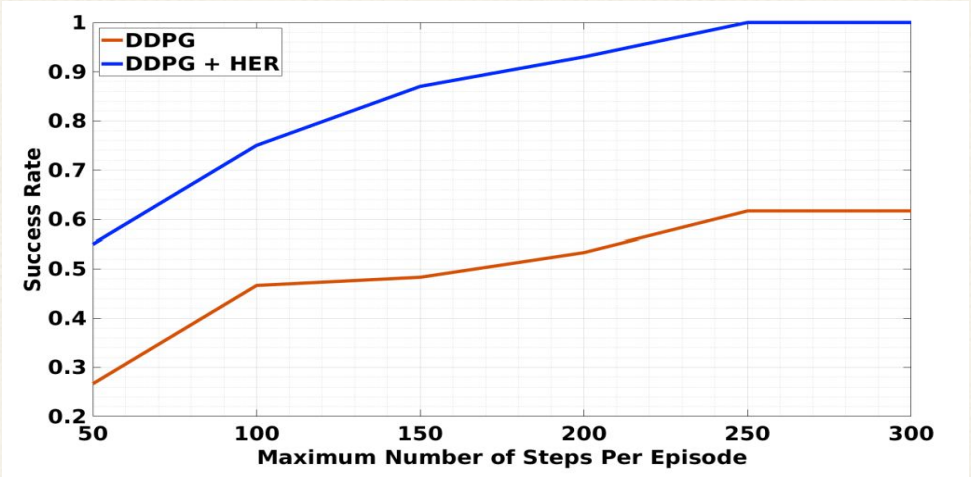
AMBF-RL

- **Need**: Robust simulation frameworks to enable RL techniques to solve medical robotics problems
- **AMBF** (Asynchronous Multibody Framework): real-time dynamics simulator
- **AMBF-RL**: toolkit that enables
  - the design of control algorithms
  - collection and processing of expert data from demonstration on real systems
- Toolkit to assist in designing control algorithms for medical robotics in AMBF simulator
- Demonstrated use of RL for debris removal on dVRK Patient Side Manipulator (PSM)
  - Deep Deterministic Policy Gradient (DDPG)
  - Hindsight Experience Replay (HER)
- Successfully transferred optimal RL policy to the physical system

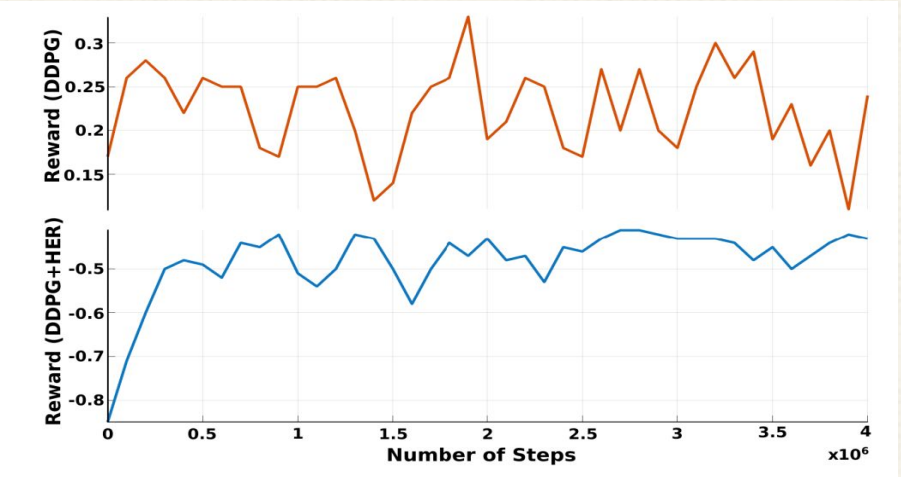
AMBF-RL

DDPG VS DDPG + HER

Success Rate vs Max Steps



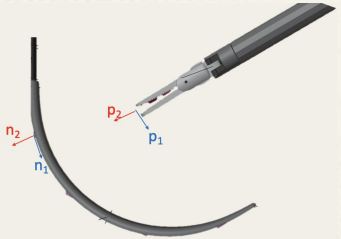
Reward vs Time

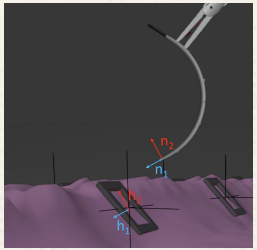


[4] A. Munawar, Y. Wang, R. Gondokaryono, and G. S. Fischer, “A real-time dynamic simulator and an associated front-end representation format for simulating complex robots and environments.”

AMBF-RL: Connection to project

- **Takeaways**:
  - Transferring environment to Open-AI Gym
    - Easy comparison amongst different RL techniques
    - Reward function iterability
  - Building off AMBF-RL Debris removal / Reach task
- **Pros**:
  - Clear documentation, linked codebase
- **Cons**:
  - Potentially overly simplistic reward function
  - Tackling Needle grasping & suturing requires more complex reward function to account for approach angle



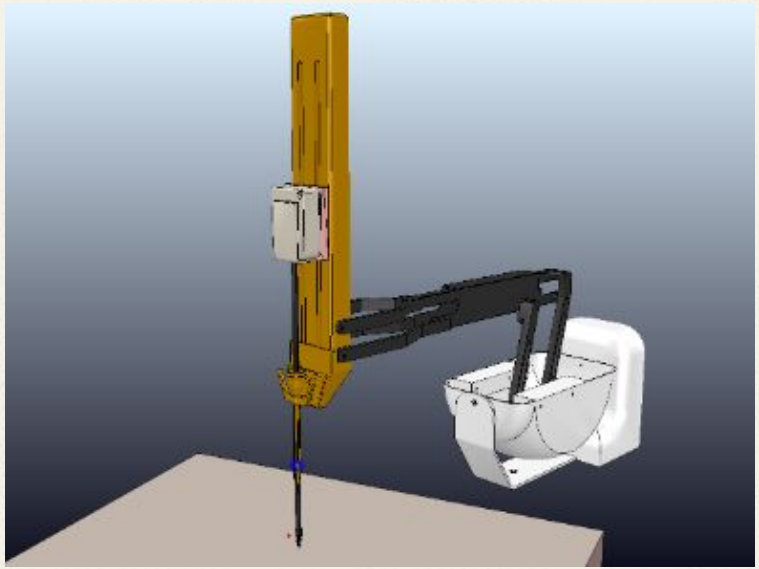


Paper 3: dVRL

[1] Richter, F., Orosco, R. K., & Yip, M. C. (2019). Open-sourced reinforcement learning environments for surgical robotics. arXiv preprint arXiv:1903.02090.

dVRL

- First open-sourced RL environment for surgical robots
  - Kinematic, not dynamic simulation
- Goals:
  - Provide an RL environment for training surgical robotics
  - Transfer the learned policies from simulation to an actual robot



[1] Richter, F., Orosco, R. K., & Yip, M. C. (2019).

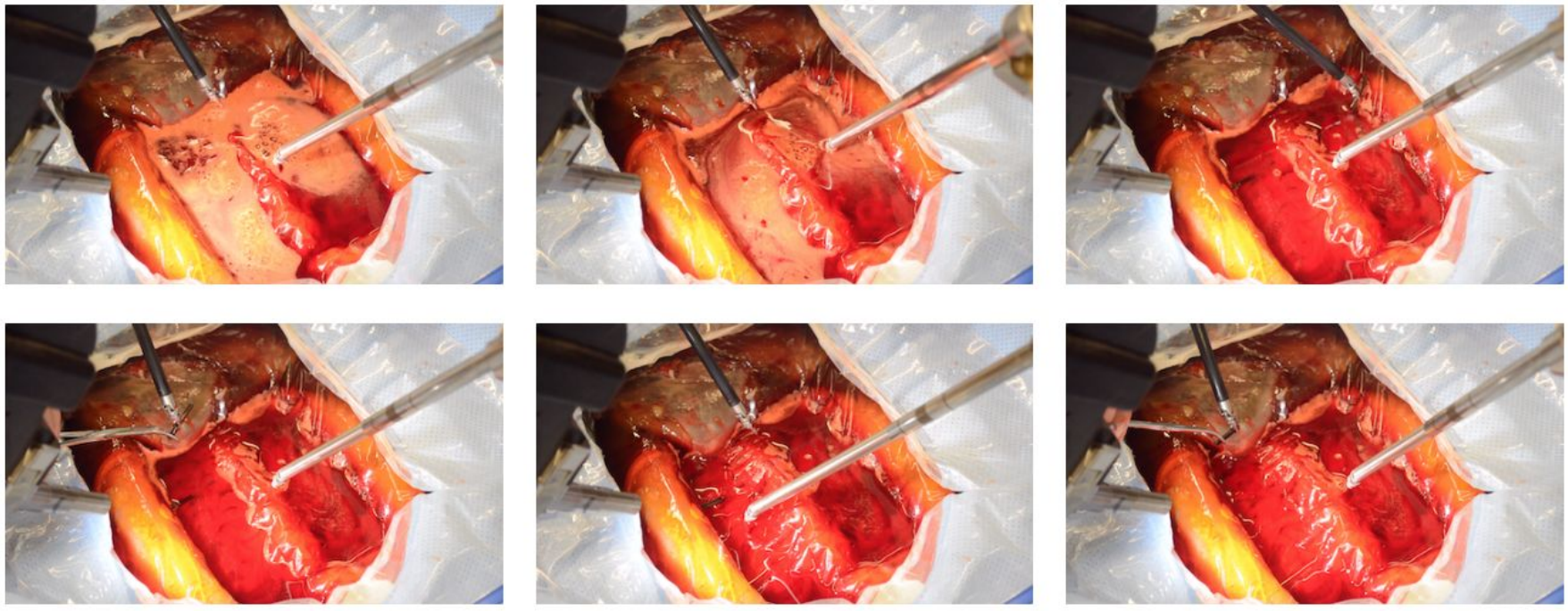
- Fetch task:

$$r(s_t) = \begin{cases} -1 & \text{if } \rho \parallel \tilde{\mathbf{p}}_\mathbf{t} - \tilde{\mathbf{g}}_\mathbf{t} \parallel > \delta \\ 0 & \text{otherwise} \end{cases}$$

- Pick task:

$$r(s_t) = \begin{cases} -1 & \text{if } \rho \parallel \tilde{\mathbf{o}}_\mathbf{t} - \tilde{\mathbf{g}}_\mathbf{t} \parallel > \delta \\ 0 & \text{otherwise} \end{cases}$$

dVRL Methods



Demonstration of Pick and Reach Policies used in mock operating room, removing fake blood to reveal debris [1] Richter, F., Orosco, R. K., & Yip, M. C. (2019).

dVRL Methods

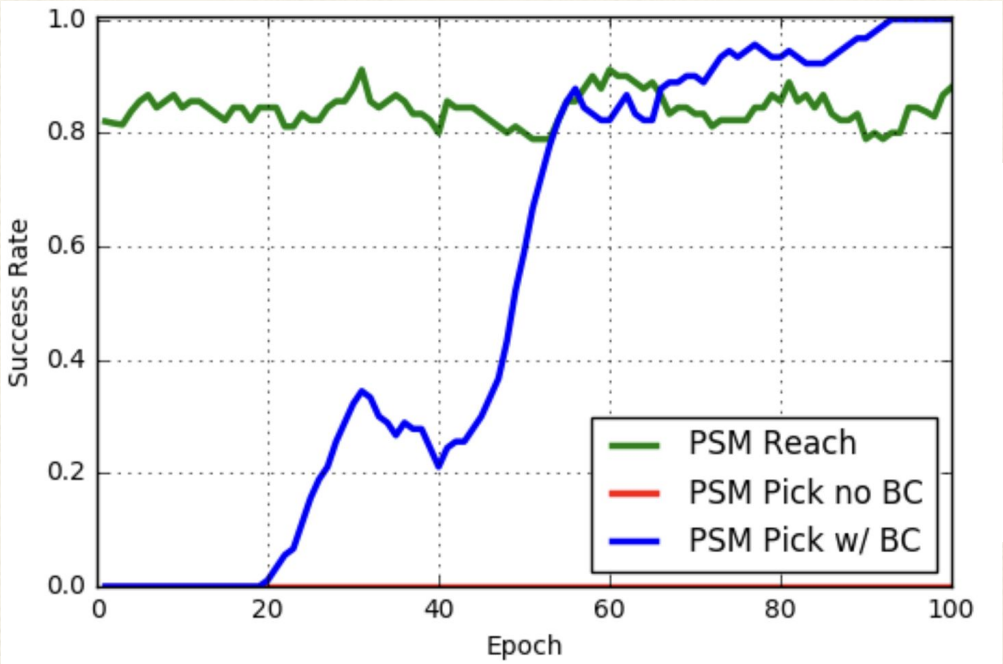


TABLE II:
Timing Results of one rollout per Environment

Num. of Env.	PSM Reach	PSM Pick
1	2.09 sec	2.09 sec
2	2.36 sec	2.35 sec
4	2.78 sec	2.78 sec
6	3.03 sec	3.02 sec
8	3.27 sec	3.26 sec

BC: behavioral cloning

[1] Richter, F., Orosco, R. K., & Yip, M. C. (2019).

dVRL: Takeaways

- Similar to the in-progress SRCEnv, with simpler tasks
- **Benchmarking**:
  - dVRL Fetch :: SRC Grasp
  - dVRL Pick :: SRC Insert & Target
- **Pros**:
  - Successful demonstration of policy in mock operation
  - Detailed calculations, algorithmic descriptions and analysis of experiments
- **Cons**:
  - Binary reward functions may be insufficient for capturing enough information about how “good” or “bad” a step is for more complex task

TAKEAWAYS

TAKEAWAYS

- SRCEnv project is critical to updating robotic suturing systems with SOTA technologies
  - dVRK
  - AMBF
  - HER algorithm
  - OpenAI Gymnasium
- Benchmark our algorithm’s accuracy with RL environments developed for similar tasks
- Make modifications to RL algorithm based on successful reward functions that others have developed

REFERENCES

[1] Richter, F., Orosco, R. K., & Yip, M. C. (2019). Open-sourced reinforcement learning environments for surgical robotics. arXiv preprint arXiv:1903.02090.

[2] 2021-2022 AccelNet Surgical Robotics Challenge (online). Collaborative Robotics Toolkit (CRTK). Retrieved February 19, 2023, from https://collaborative-robotics.github.io/surgical-robotics-challenge/challenge-2021.html

[3] Gymnasium documentation. Basic Usage. (n.d.). Retrieved February 19, 2023, from https://gymnasium.farama.org/content/basic_usage/

[4] A. Munawar, Y. Wang, R. Gondokaryono, and G. S. Fischer, “A real-time dynamic simulator and an associated front-end representation format for simulating complex robots and environments,” in 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, Nov. 2019. [Online]. Available: https://doi.org/10.1109/iros40897.2019.89685688

[5] P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, “Continuous control with deep reinforcement learning,” 2015. [Online]. Available: https://arxiv.org/abs/1509.02971

[6] M. Andrychowicz, F. Wolski, A. Ray, J. Schneider, R. Fong, P. Welinder, B. McGrew, J. Tobin, O. Pieter Abbeel, and W. Zaremba, Hindsight experience replay,” Advances in neural information processing systems, vol. 30, 2017

[6] J. Weng, H. Chen, D. Yan, K. You, A. Duburcq, M. Zhang, Y. Su, H. Su, and J. Zhu, “Tianshou: A highly modularized deep reinforcement learning library,” arXiv preprint arXiv:2107.14171, 2021. 9

[7] Introduction to reinforcement learning with David Silver. DeepMind. (n.d.). Retrieved February 19, 2023, from https://www.deepmind.com/learning-resources/introduction-to-reinforcement-learning-with-david-silver