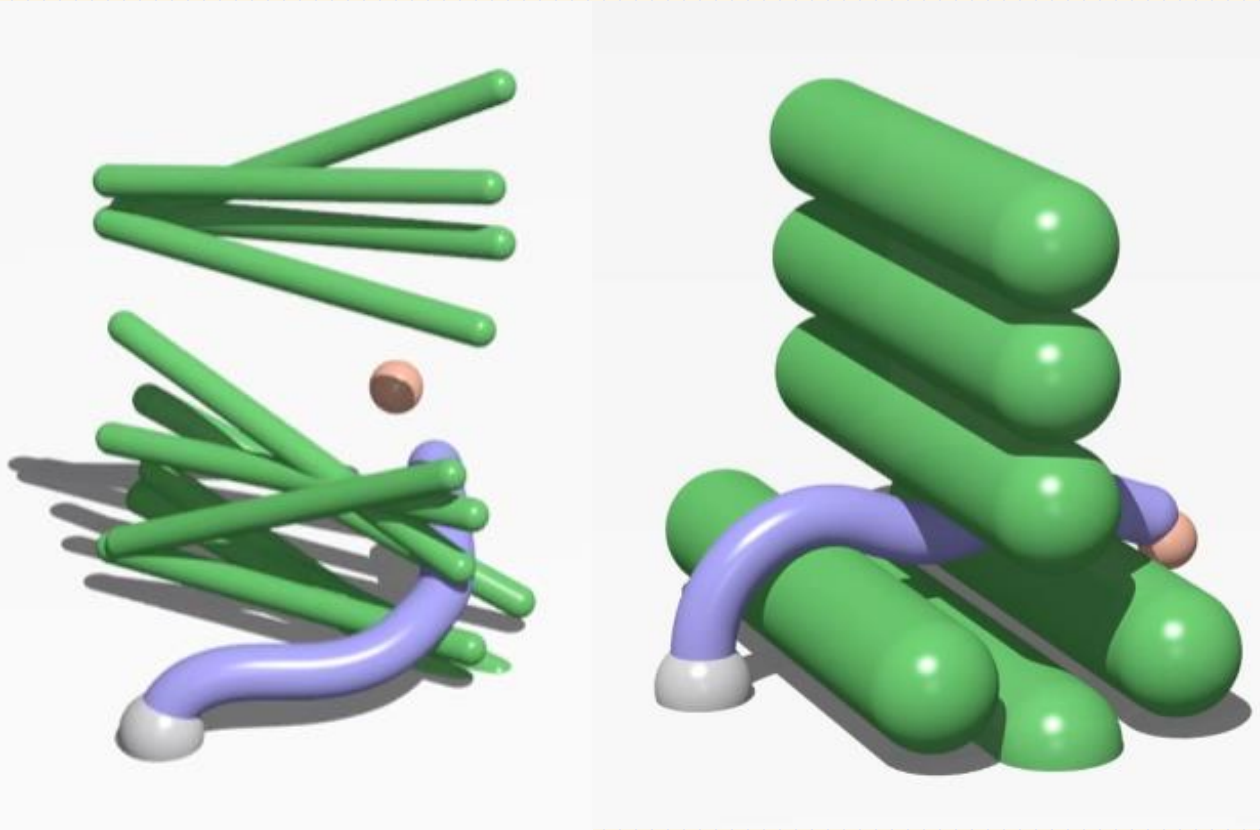
NUS National University of Singapore

Soft Robot RL-Control based on Elastica Simulator

NEURO-ROBOTICS PROJECT 18

ZHAO YIMIN LIU XIANGGE WANG PENGYU LIN YI

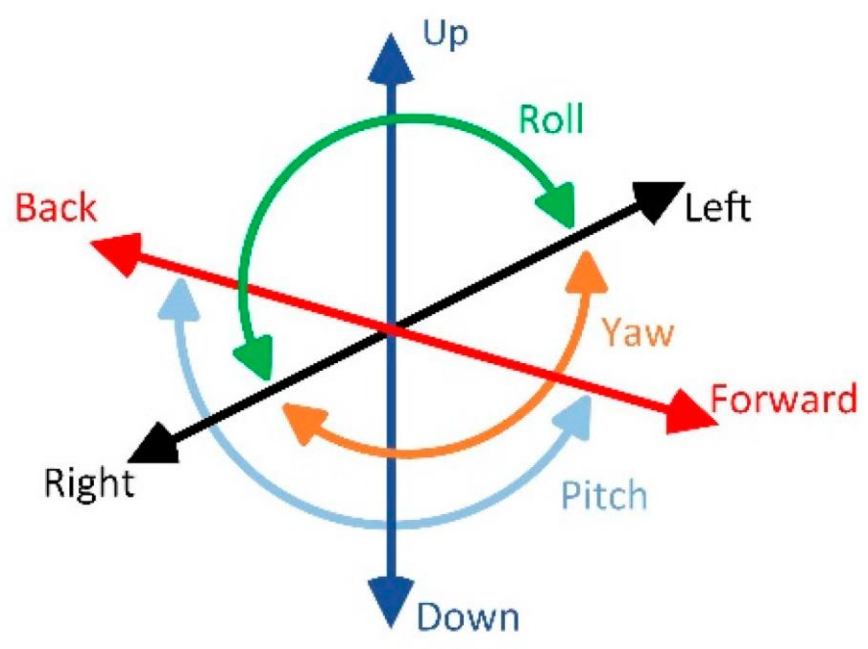


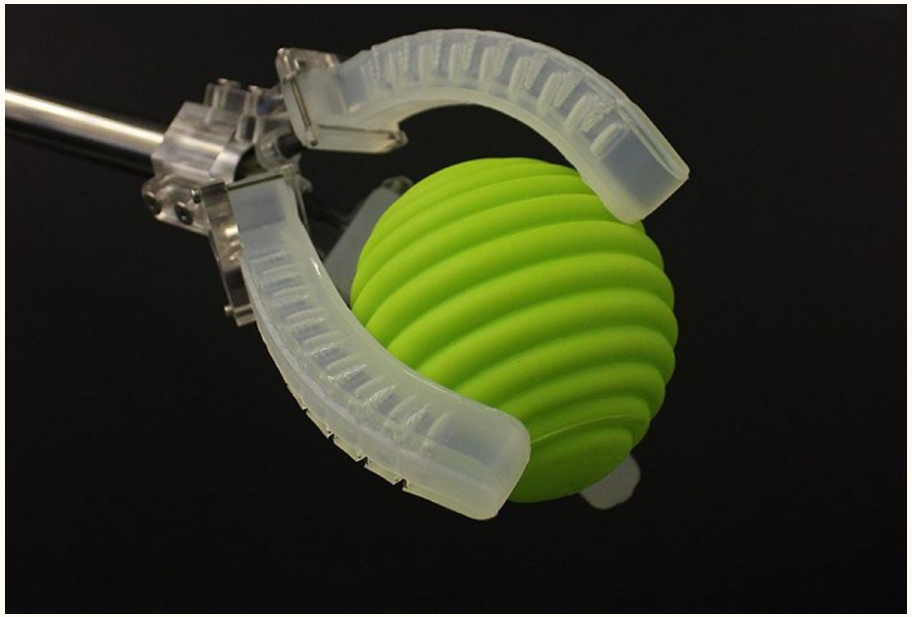
Contents

NUS National University of Singapore

Introduction: Challenge

- The control challenge of soft robots primarily arises from two aspects



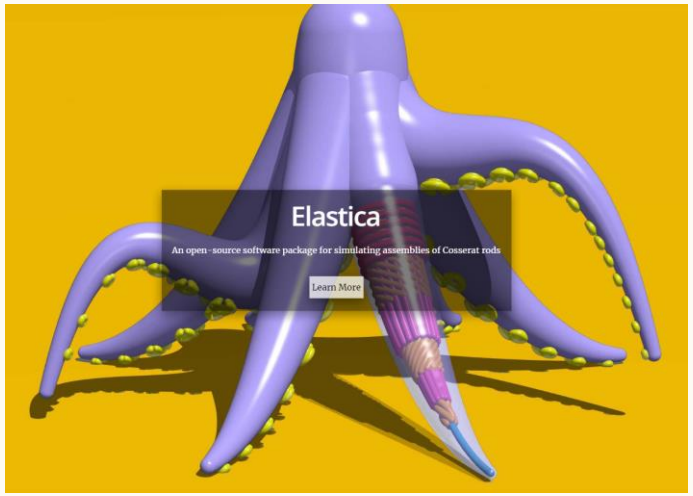


- Degree of Freedom(DOF)
- Long Range Stress

NUS National University of Singapore

Introduction: Solution

- At present, due to the lack of strict, accurate and efficient numerical models, it is difficult to model the mechanical behavior of soft robots, which adds challenges to the control problem.
- However, if the elastic effects of soft robots can be accurately captured, there is an opportunity to use them to simplify control problems.
- Therefore, the emergence of **Elastica** fills the gap between traditional rigid-body solvers and high-fidelity finite element methods, providing a new test platform for the control methods of soft robots

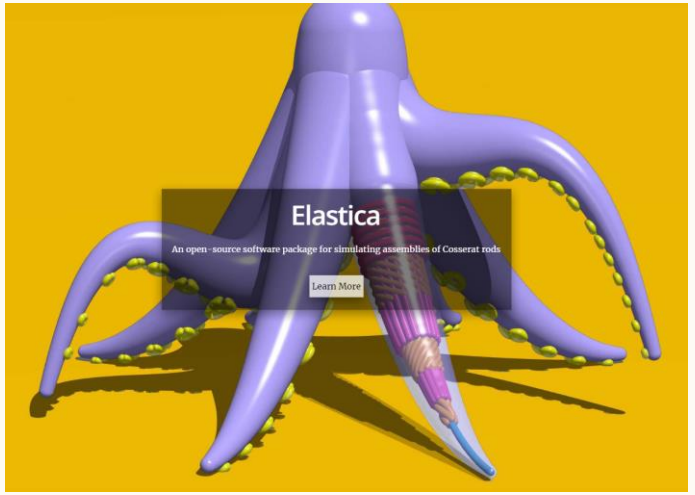


NUS National University of Singapore

Introduction: Toolkits

The Elastica Simulation Platform

This is an open-source software aiming to provide cutting-edge platform for simulating the behavior of elastic materials. The software supports a variety of elastic material models, including linear elastic, nonlinear elastic, plastic and other models.



- **PyElastica Package:** This is the python implementation of Elastica, which is used to simulate assemblies of slender, one-dimensional structures using Cosserat Rod theory. We used PyElastica to achieve robot simulation.

- **Stable Baselines:** Reinforcement learning algorithms toolkits based on OpenAI Baselines. Five different RL model-free algorithms are used.

NUS National University of Singapore

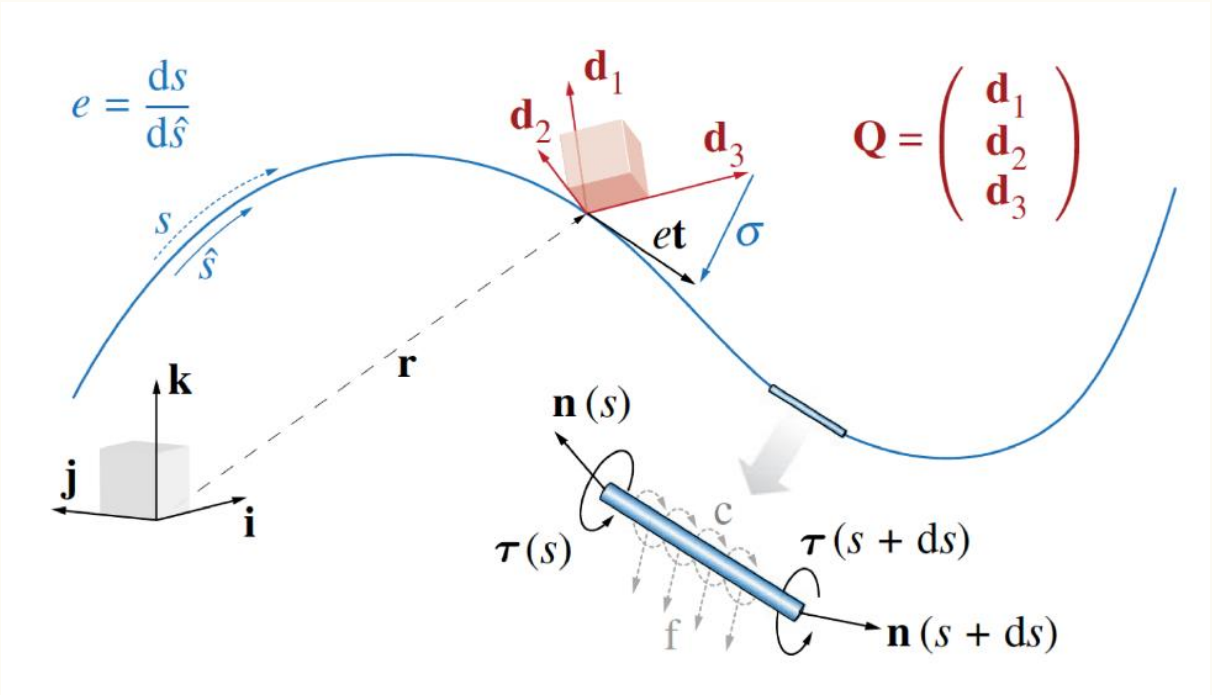
Introduction: Backgrounds

- Based on some open-source projects, we achieve simulation of controlling soft robot to touch the target point using its end-effector in 4 cases.
- Analyze others' experiments and tune better RL parameters.
- Fix the bugs in simulation and rendering codes.
- Train the new controller for each case.
- Construct the pipeline instruction for the process mentioned above and publish on GitHub: https://github.com/ztony0712/Elastica-RL-control-fix-improve

NUS National University of Singapore

Preliminary: Cosserat Rods

- **Theory:** Cosserat rods are a generalization of Kirchhoff rods, which model 1-d, slender rods incorporating only bend and twist. Cosserat rods add the ability to consider stretching and shearing, allowing all the possible modes of deformation of the system to be considered.

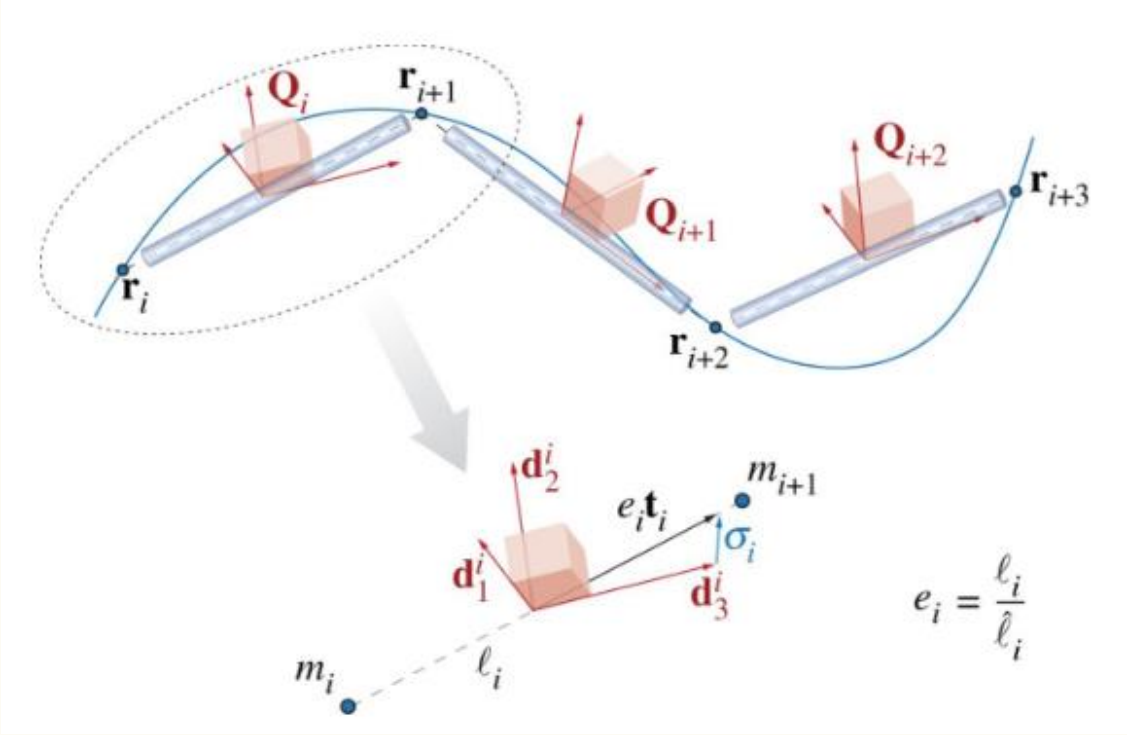


NUS National University of Singapore

Preliminary: Cosserat Rods

Numerics: There are three steps to solving this problem numerically

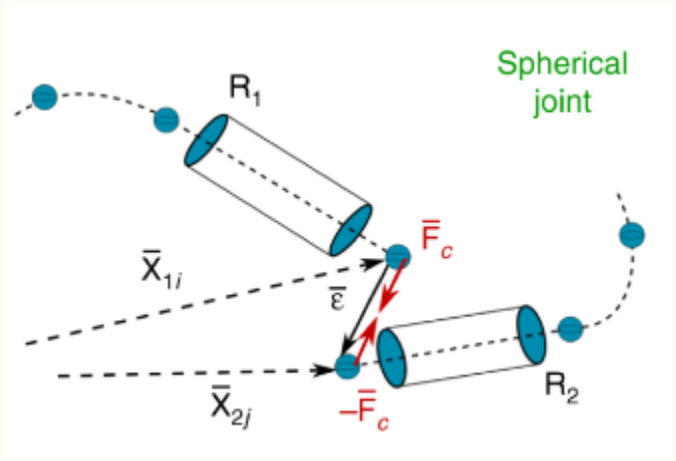
- Spatially discretizing the continuum Cosserat rod equations
- Selecting a time stepping algorithm
- Specifying boundary conditions and interaction forces

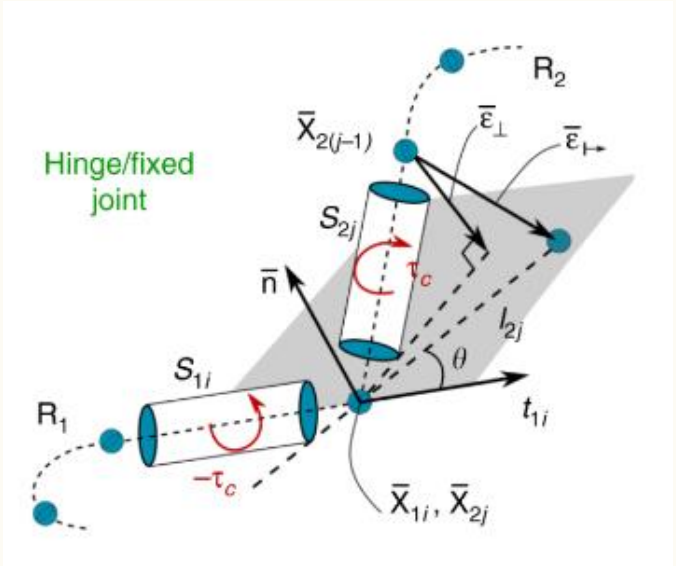


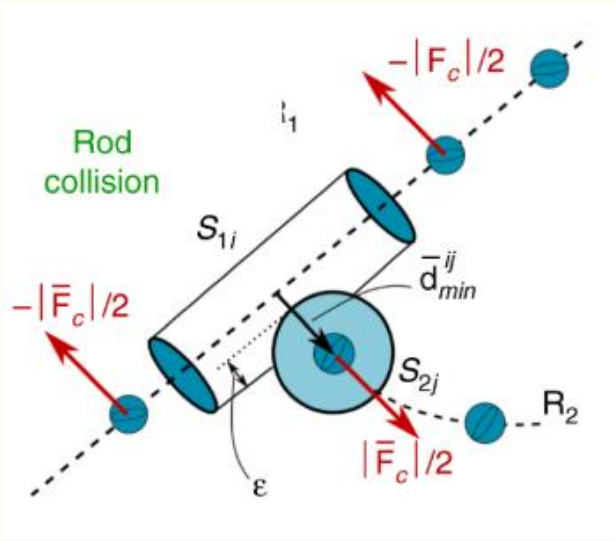
NUS National University of Singapore

Preliminary: Cosserat Rods

- **Multiple Rods:** Complex systems, such as musculoskeletal architectures, often require modeling assemblies of Cosserat rods. These assemblies can be a heterogeneous mix of active and passive rods that, when coupled together, allow modeling of dynamic structures. To assemble multiple active and passive rods into these dynamic architectures, it is first necessary to prescribe their rules of interaction.







NUS National University of Singapore

Preliminary: Reinforcement Learning

In the aspect of soft robot control in Elastica, we adopt five commonly used reinforcement learning algorithms

- **TRPO (Trust Region Policy Optimization)** is a reinforcement learning algorithm that optimizes strategies to maximize expected returns. The key idea of TRPO is to limit the size of each policy update, ensuring trust and improving stability. Although it has better convergence and stability, its computational cost is high.
- **PPO (Proximal Policy Optimization)** improves on this by using techniques such as shear ratio and shear advantage function to limit policy updates while maintaining reliability, resulting in better performance and lower computational cost than TRPO.

On-policy

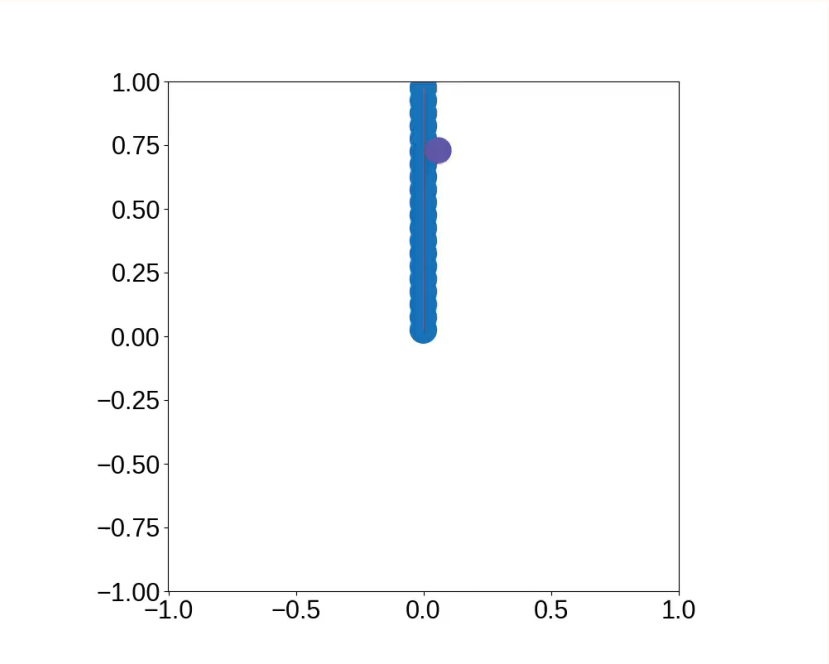
- **SAC (Soft Actor-Critic)** balances exploration and exploitation by maximizing strategy entropy, promoting more exploratory behavior for good performance especially in continuous motion space problems.
- **DDPG (Deep Deterministic Policy Gradient)** suitable for continuous action and state space problems, uses deep neural networks to represent policies and value functions with experiential playback and target network techniques for improved stability but may face challenges with convergence in some environments
- **TD3 (Twin Delayed DDPG)** further improves on this with dual Q networks and delayed policy updates.

Off-policy

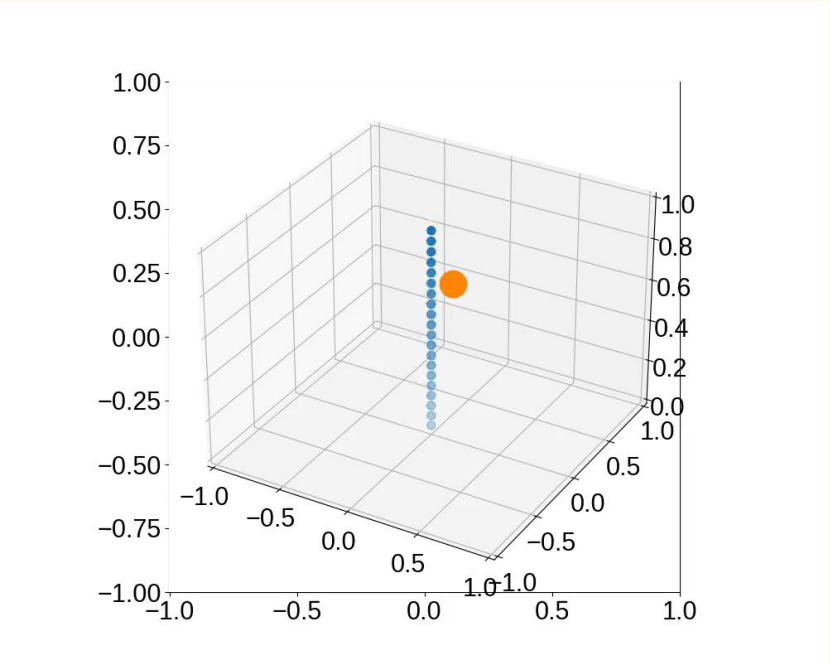
NUS National University of Singapore

Elastica Case 1-3D tracking of a randomly moving target

- **Problem Description:** The goal of this case is for the tip of the compliant arm to continuously track a randomly moving target in 3D space. The target moves with a constant velocity of 0.5 m/s while randomly changing directions every 0.7 seconds.



2D visualization



3D visualization

NUS National University of Singapore

Elastica Case 1-3D tracking of a randomly moving target

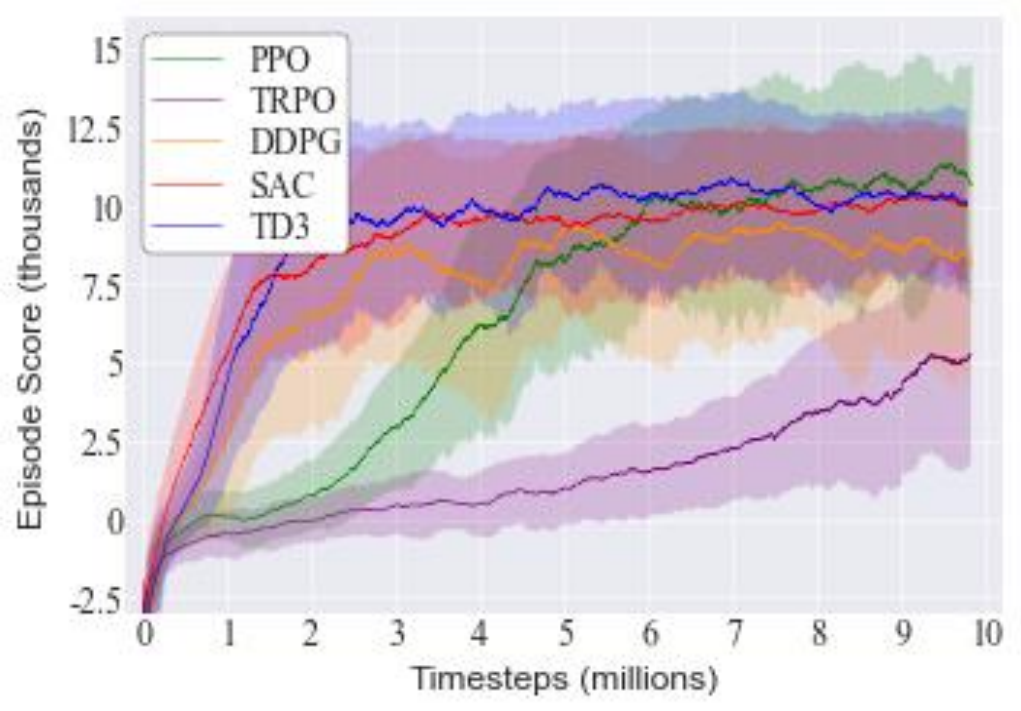


Fig. Learning results of the different algorithm

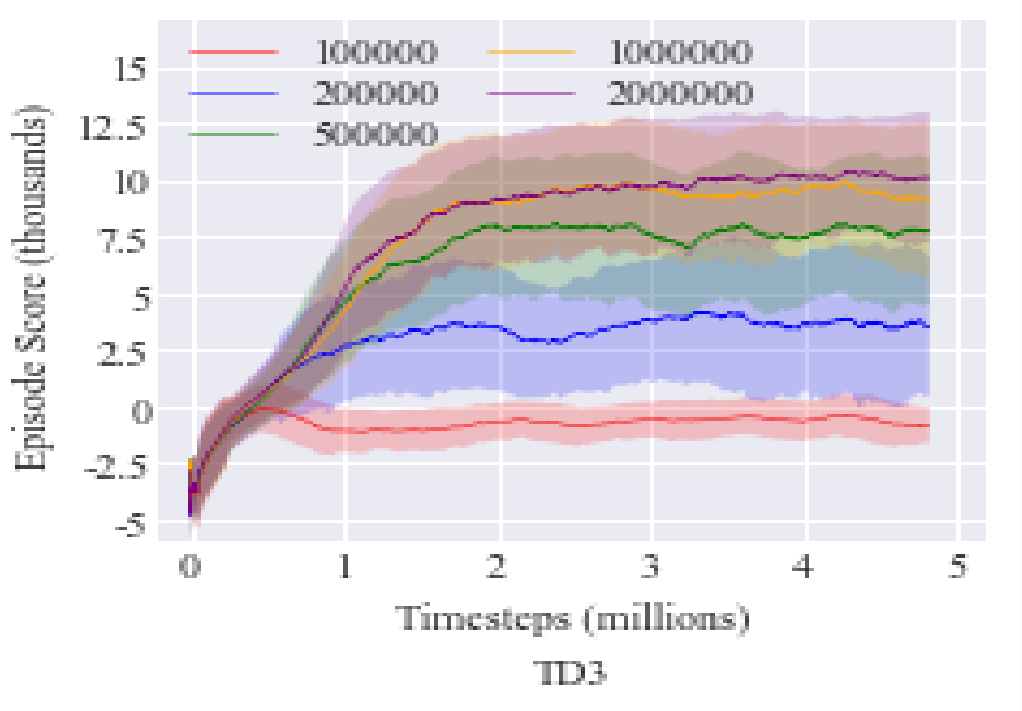


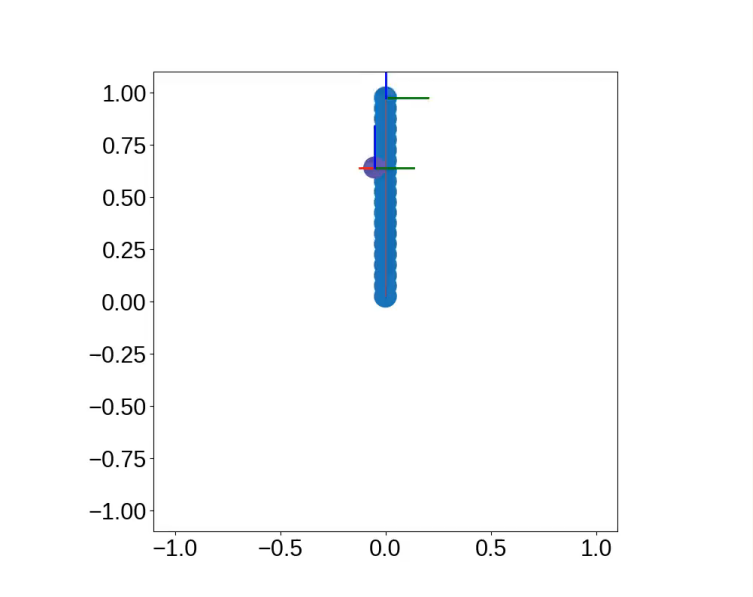
Fig. Different algorithm with different timesteps

- Choose TD3 cause its curve is stable. Choose 7e6 as total timesteps cause that's the peak of curve.
- These choices can balance the performance and training time.
- TD3 perform best when timesteps per batch are 2e6.

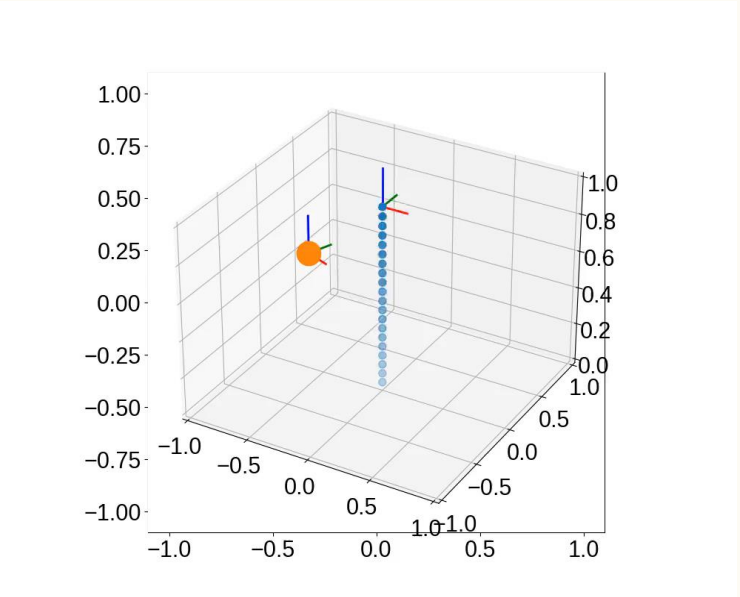
NUS National University of Singapore

Elastica Case 2-Reaching to random target location with defined orientation

- **Problem Description:** The goal of this case is to have the tip of the arm reach towards a stationary target location that is randomly positioned every episode. The target orientation is defined such that the tangent direction of the arm tip should be pointed vertically upwards while the normal-binormal vectors are rotated away from the global coordinate frame by a random amount between $-90°$ and $90°$



2D visualization



3D visualization

NUS National University of Singapore

Elastica Case 2-Reaching to random target location with defined orientation

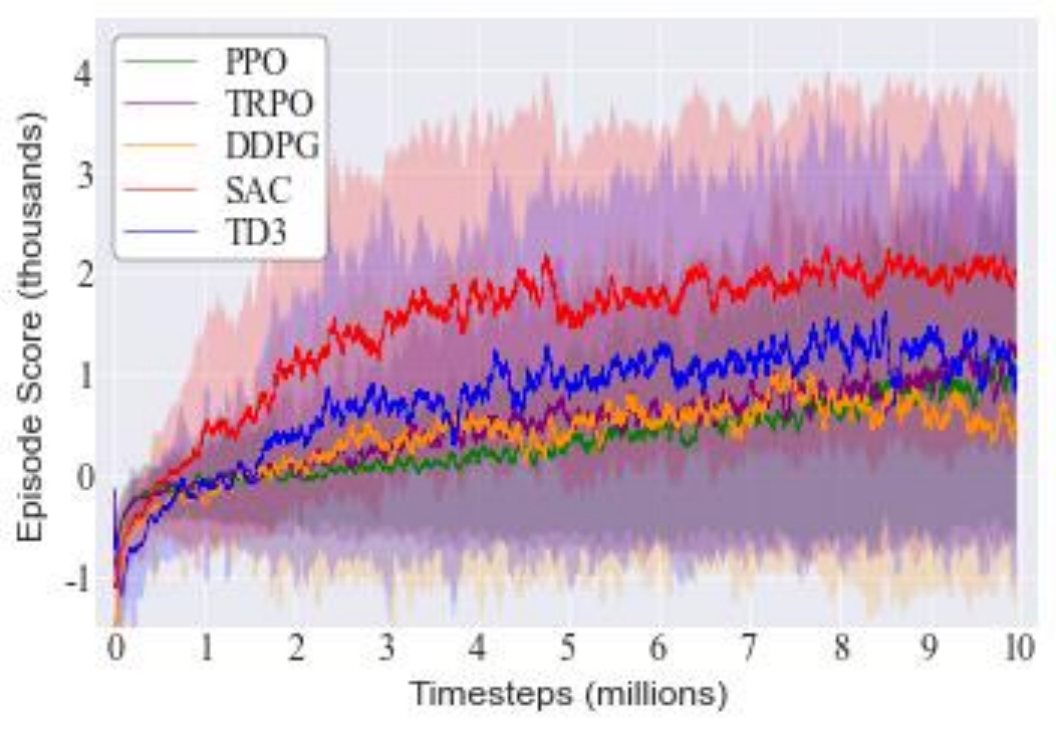


Fig. Learning results of the different algorithm

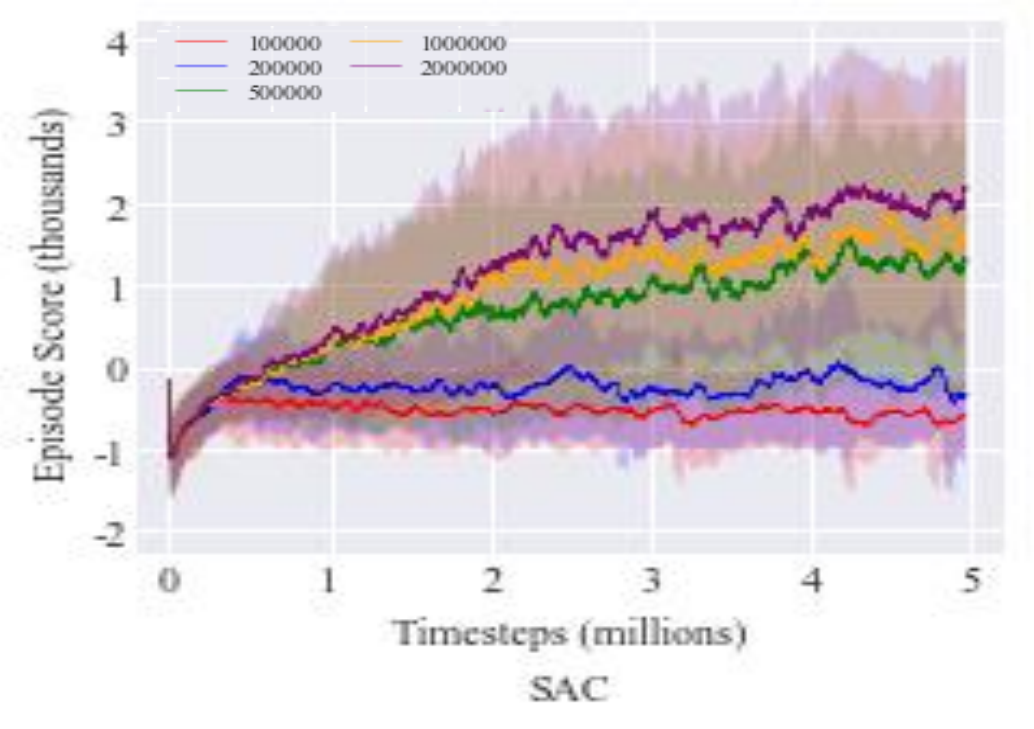


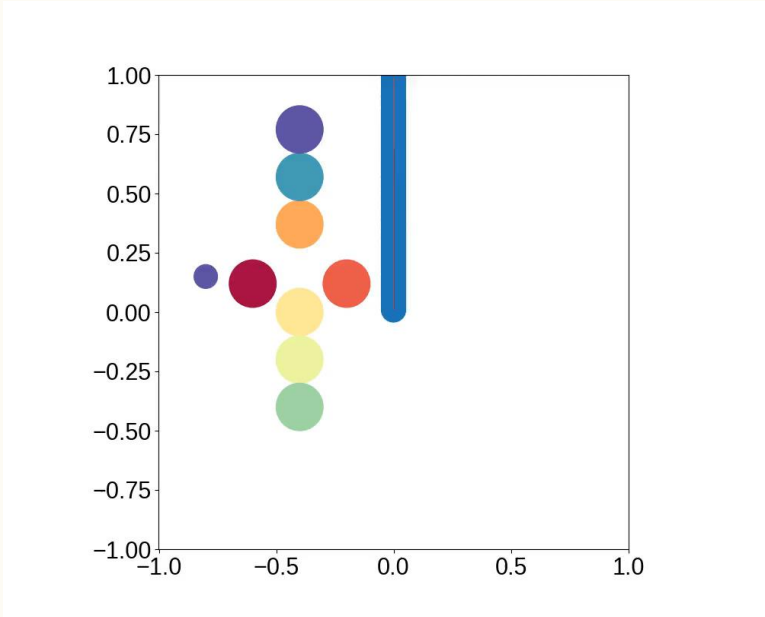
Fig. Different algorithm in different timesteps

- Choose SAC cause its curve is the best. The curve converge at 8e6, which is chosen as total timesteps.
- These choices can balance the performance and training time.
- SAC perform best when timesteps per batch are 2e6.

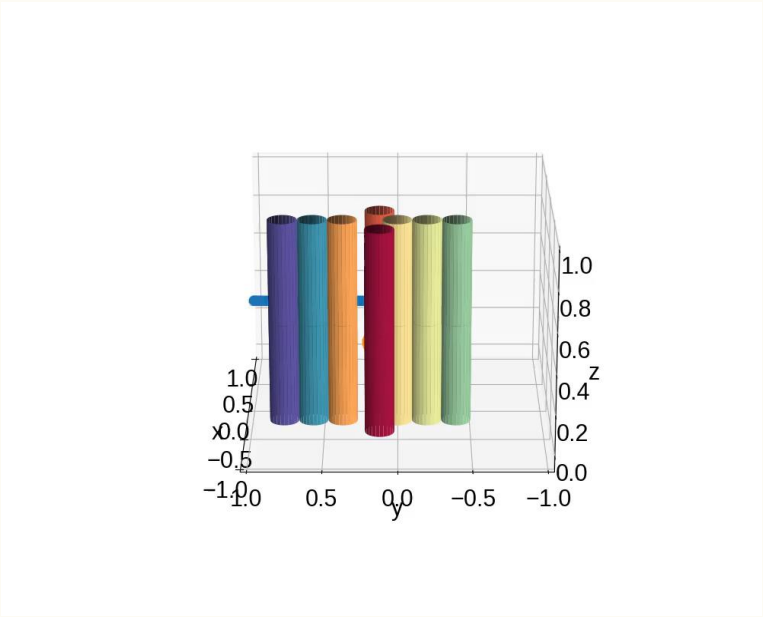
NUS National University of Singapore

Elastica Case 3-Underactuated maneuvering among structured obstacles

- **Problem Description:** In this case, a stationary target is placed behind an array of 8 obstacles with an opening through which the arm must maneuver to reach the target. The target is placed in the normal plane so that only in-plane actuation is required. Obstacles and target locations are located in the same location and configuration each episode.



2D visualization



3D visualization

NUS National University of Singapore

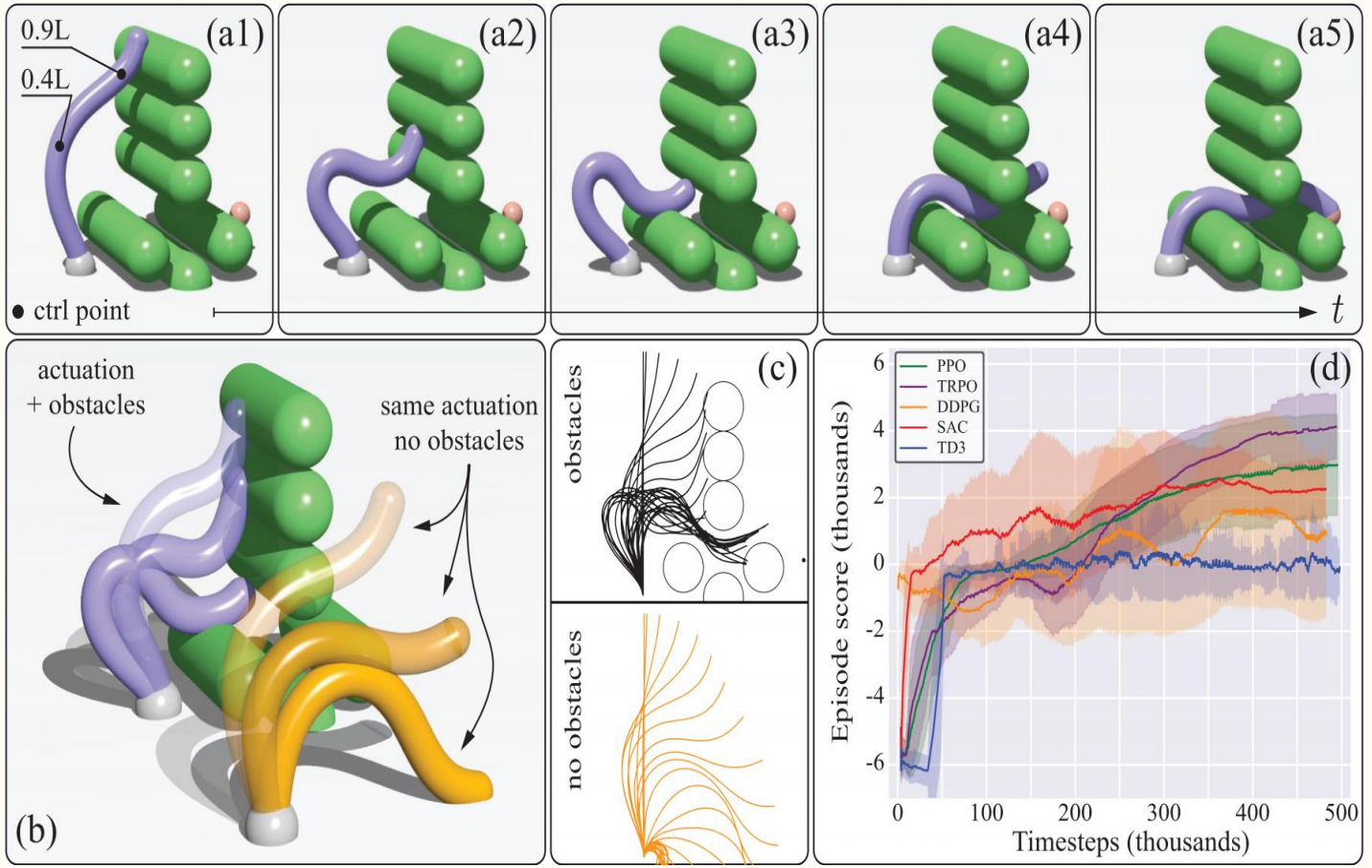
Elastica Case 3-Underactuated maneuvering among structured obstacles

- Choose TRPO cause its curve is the best. The peak of curve is at 500e6, which is chosen as total timesteps.

- All experiments were conducted by setting 16000 timesteps per batch, which was claimed as the best

- Case 4 is similar to Case 3, so we still chose to use 2-control points.

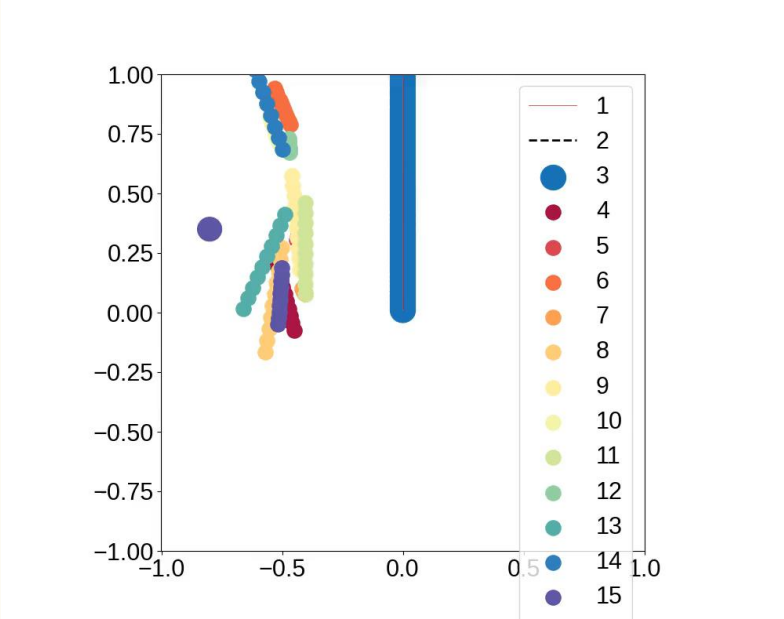
- 2 control points manually placed at locations 0.4L and 0.9L along the arm were used.



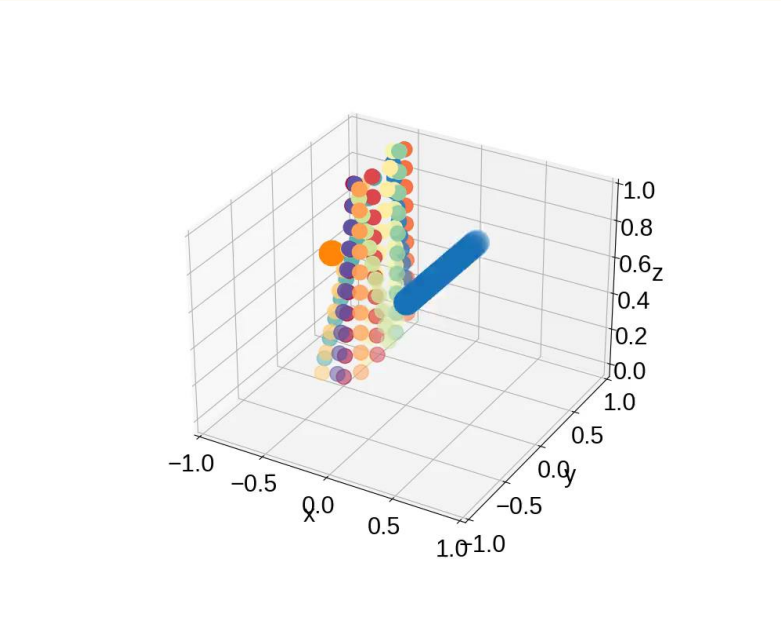
NUS National University of Singapore

Elastica Case 4-Underactuated maneuvering among unstructured obstacles

- **Problem Description:** The goal of this case is to have the tip of the arm reach towards a stationary target by maneuvering around an unstructured nest of 12 randomly located obstacles. Obstacles and target locations are located in the same place and configuration each episode.



2D visualization



3D visualization

NUS National University of Singapore

Elastica Case 4-Underactuated maneuvering among unstructured obstacles

Fig. Different algorithm in different timesteps

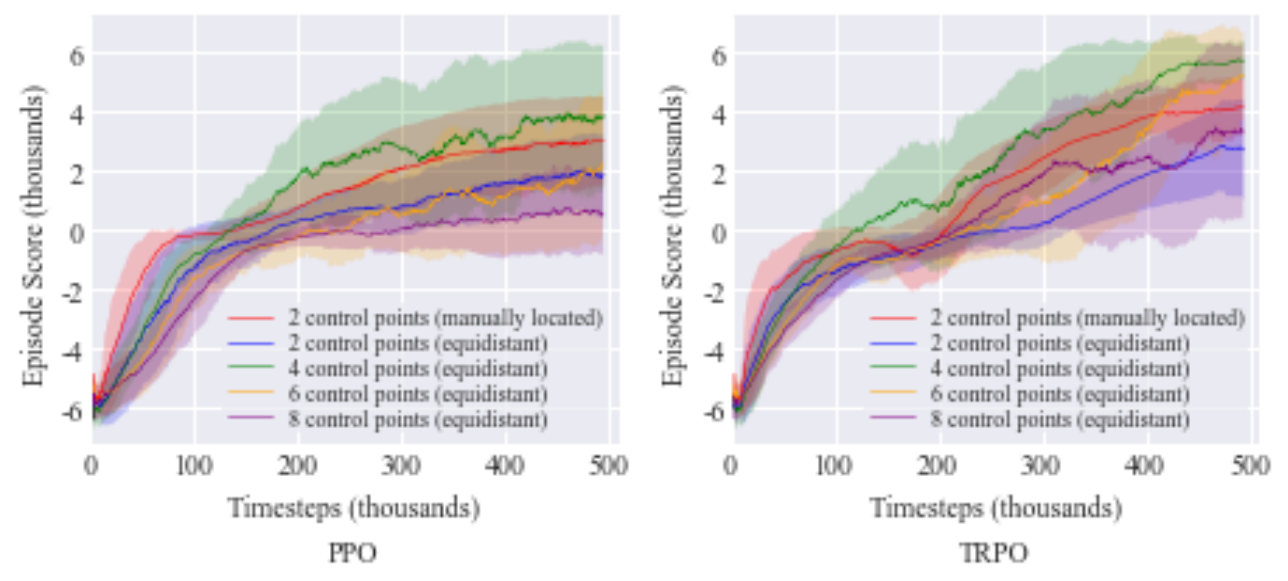
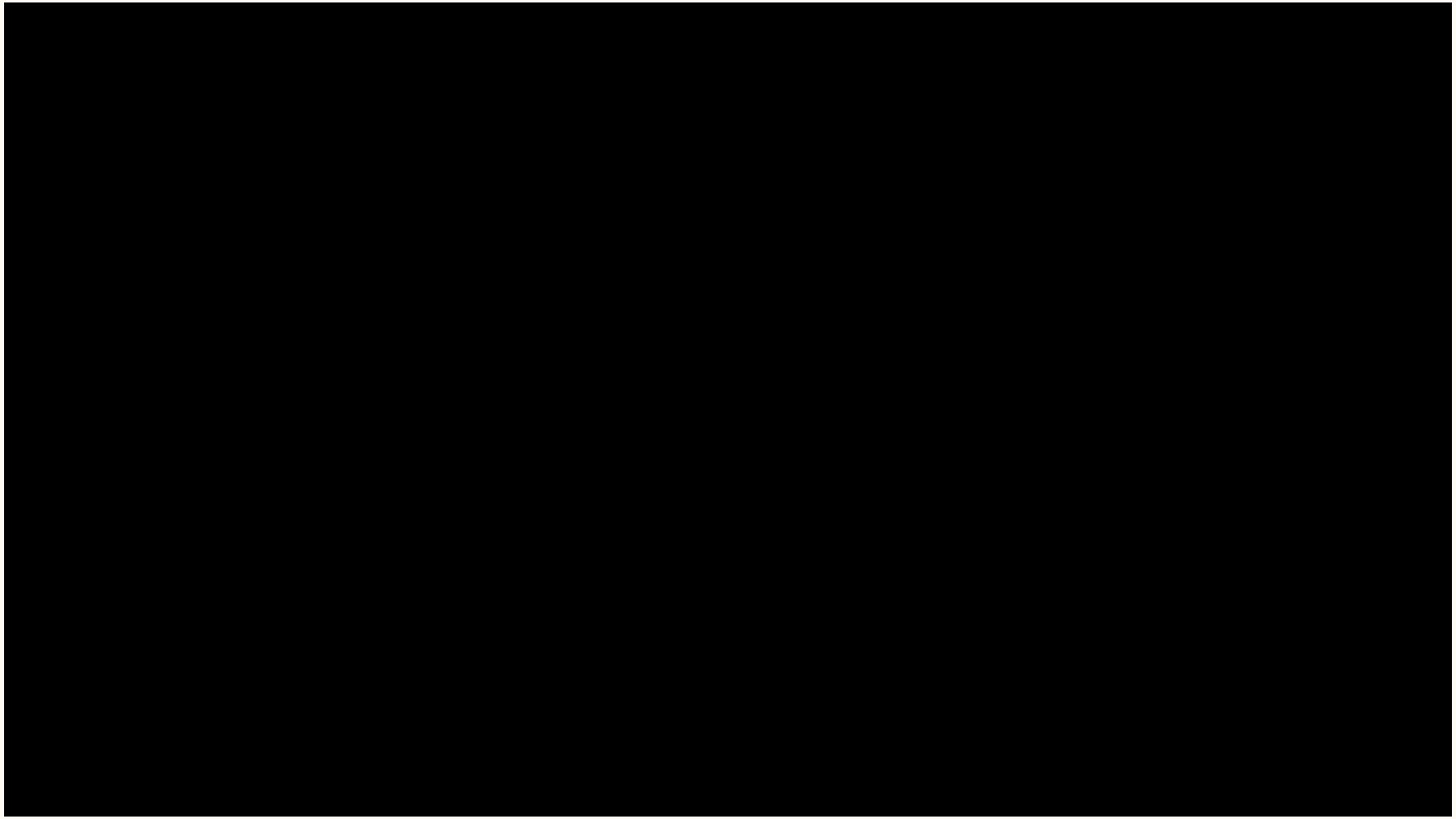


Fig. Different algorithm with different control points

- Choose PPO cause its curve is high and stable. The total timesteps is set to 1.0e6.
- All experiments were conducted by setting 16000 timesteps per batch, which was claimed as the best.
- To balance the performance and resource usage, we choose the 2-control points located manually.
- 2 control points manually placed at locations 0.4L and 0.9L along the arm were used.

NUS National University of Singapore

Elastica Case All – Video Presentation



NUS National University of Singapore

Results and Discussion

- ✓ For Case 1 and 2, all algorithm finished task because it is simple, but off-policy did better.
- ✓ For Case 3 and 4, off-policy even can't converge. The failure caused by numerical instabilities derived from large external contact forces from slamming arm into barriers.

Case	Algorithm	Score	Total TS (millions)	TS per batch (millions)
1	TD3	10	7	2
2	SAC	2	8	2
3	TRPO	4	0.5	0.016
4	PPO	2	1	0.016

Table 1: **Cases parameters comparison. Case** is the case number, **Algorithm** is the optimal algorithm used, **Score** represents performance, **TS** is the time steps.

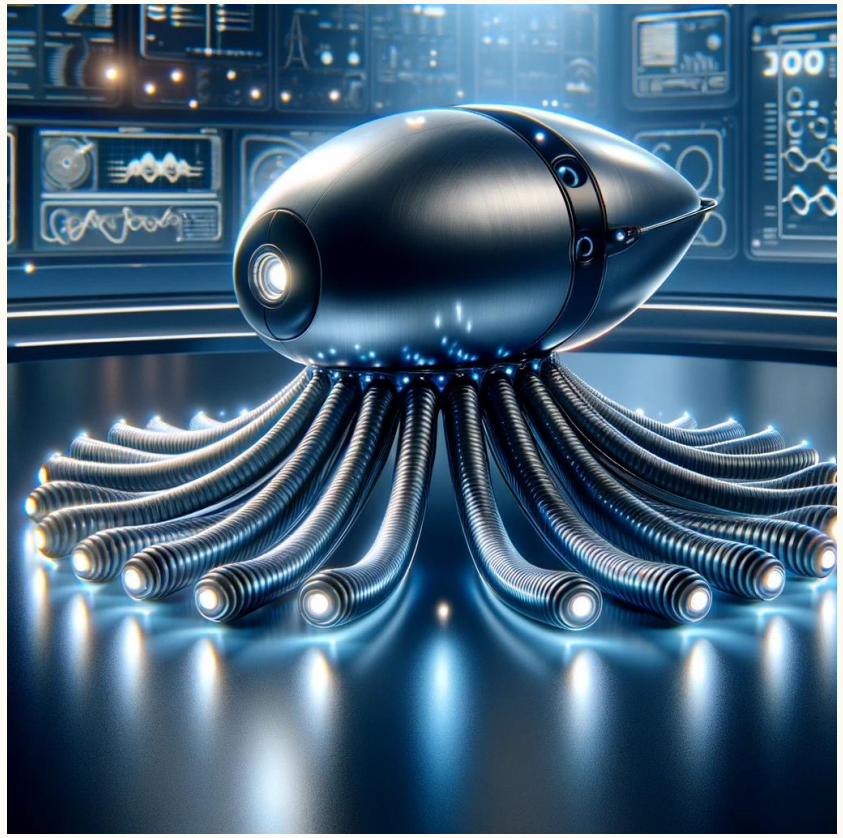
Practically, we need to combine advantages of on-policy and off-policy by classifying external environment. That is the development direction of the next generation of soft robot controller.

NUS National University of Singapore

Conclusion

In this project, we ……

- Introduce the Elastica and RL to solve soft robots' simulation
- Analyze former experiments results and tune RL parameters
- Train new controllers
- Visualize the results and render videos



Check our code by: https://github.com/ztony0712/Elastica-RL-control-fix-improve

NUS National University of Singapore

Reference

[1] Timothy P. Lillicrap, Jonathan J. Hunt, Alexander Pritzel, Nicolas Heess, Tom Erez, Yuval Tassa, David Silver, and Daan Wierstra. Continuous control with deep reinforcement learning, 2015.
[2] John Schulman, Sergey Levine, Philipp Moritz, Michael I. Jordan, and Pieter Abbeel. Trust region policy optimization. CoRR, abs/1502.05477, 2015.
[3] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. CoRR, abs/1707.06347, 2017.
[4] Scott Fujimoto, Herke van Hoof, and David Meger. Addressing function approximation error in actor-critic methods. CoRR, abs/1802.09477, 2018.
[5] Tuomas Haarnoja, Aurick Zhou, Pieter Abbeel, and Sergey Levine. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. CoRR, abs/1801.01290, 2018.
[6] Richard S. Sutton, David Mcallester, Satinder Singh, and Yishay Mansour. Policy gradient methods for reinforcement learning with function approximation. In In Advances in Neural Information Processing Systems 12, pages 1057–1063. MIT Press, 2000.
[7] Mattia Gazzola, LH Dudte, AG McCormick, and L Mahadevan. Forward and inverse problems in the mechanics of soft filaments. Royal Society Open Science, 5(6):171628, 2018.
[8] Xiaotian Zhang, Fan Kiat Chan, Tejaswin Parthasarathy, and Mattia Gazzola. Modeling and simulation of complex dynamic musculoskeletal architectures. Nature Communications, 10(1):1– 12, 2019.

Footnotes
https://github.com/hill-a/stable-baselines ↩
https://stable-baselines.readthedocs.io/en/master/modules/trpo.html ↩
https://stable-baselines.readthedocs.io/en/master/modules/ppo1.html ↩
https://stable-baselines.readthedocs.io/en/master/modules/ddpg.html ↩
https://stable-baselines.readthedocs.io/en/master/modules/td3.html ↩
https://stable-baselines.readthedocs.io/en/master/modules/sac.html ↩
https://gym.openai.com/docs/#installation ↩

NUS National University of Singapore