Shielding Induced Safe Reinforcement Learning For Drone Navigation

Daniel Bramblett, Vivek Sahukar

Group # 1

CSE 598: AI Safety and Assessment

Introduction

- Autonomous drones increasingly used in farms, surveillance, etc.
- Safety concerns training drones to navigate real world environments.
- RL requires performing unsafe behavior to learn it is unsafe.
- Unsafe behavior puts both the drone and all users/agents at risk.
- Not possible to hardcode policies to handle all real world situations.

Objective of this project:

Apply shielding to allow for safer and faster reinforcement learning training for drone navigation while avoiding obstacles.

Image Sources:
https://www.technologyreview.com/2016/07/20/158748/six-ways-drones-are-revolutionizing-agriculture/
https://www.unmannedairspace.info/latest-news-and-information/drone-delivery-operations-underway-in-26-countries/

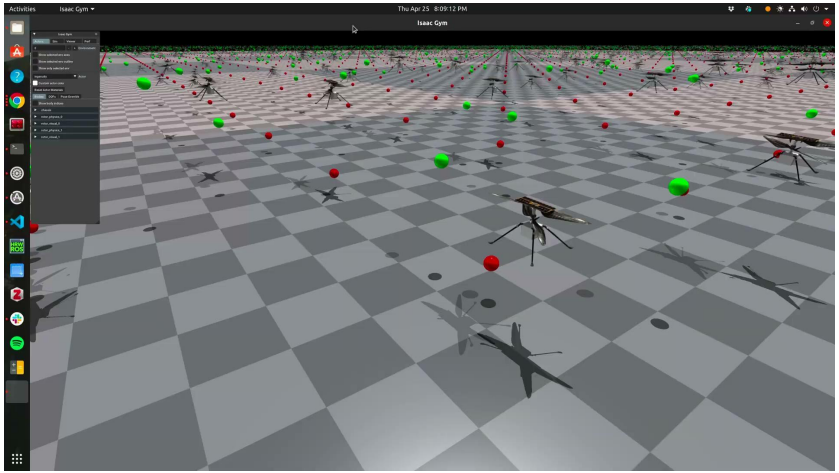
Approach

Evaluate different types of shields for safer and faster training of drones in a risk-prone environment:

1. Find an existing drone simulator.
2. Modify the environment by adding an obstacle that needs to be avoided.
3. Implement different approaches for shielding.
4. Train a reinforcement learning agent with and without a shield.
5. Evaluate the learning behavior and convergence speed while training.

Isaac Gym

- End-to-end parallel training on GPU (faster training and rendering)
- Ingenuity Domain Task: reach the random target location as quick as possible.
    - The problem is fully observable.
    - The agents has a dense reward function that rewards it for:
        - Distance to the target.
        - How upright the drone is.
        - Not spinning.

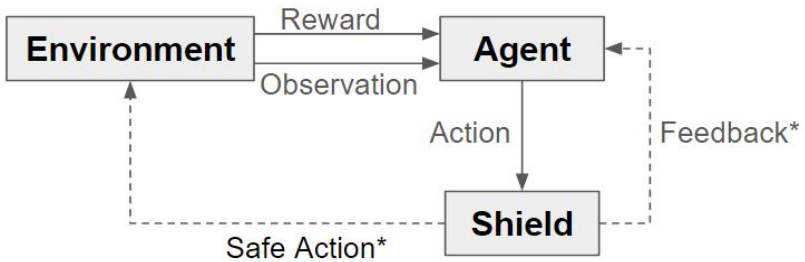


Makoviychuk, V., Wawrzyniak, L., Guo, Y., Lu, M., Storey, K., Macklin, M., ... & State, G. (2021). Isaac gym: High performance gpu-based physics simulation for robot learning. *arXiv preprint arXiv:2108.10470*.

Methodology - Shielding

Shielding: evaluates the agent's action before passing it to the environment.

- If an action does not satisfy the shield the shield can replace the action and/or give feedback to the agent.
- Unique challenge of the Ingenuity environment: continuous action space
  - Ingenuity Helicopter has 6 degrees of freedom.



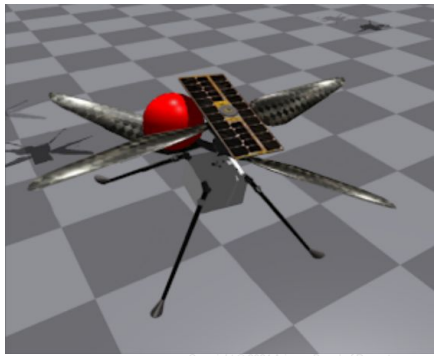
Referenced image source: Alshiekh, Mohammed, et al. "Safe reinforcement learning via shielding." *Proceedings of the AAAI conference on artificial intelligence*. Vol. 32. No. 1. 2018.

Methodology - Adding Risk Into Ingenuity

We added in an obstacle into the Ingenuity environment.

- If the agent crashed into the object, they take a -10,000 reward and the environment resets.
    - Any crashes will result in a negative expected value signifying unsafe behavior.
- The agent receives an observation of its relative location from the obstacle.

Updated Task: Reach the target location while safely circumnavigating the obstacle.



Methodology - 3 Shielding Variations

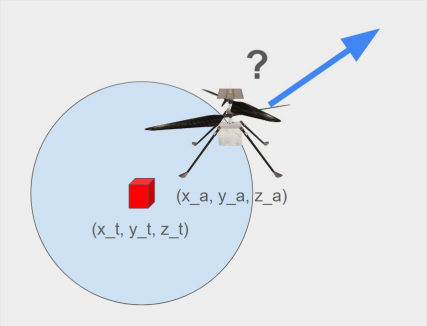
Actions fail the shield if the agent is within a certain distance of the obstacle.

Safe Action: Move the agent out of the shield as fast as possible

Feedback: Take away the positive reward

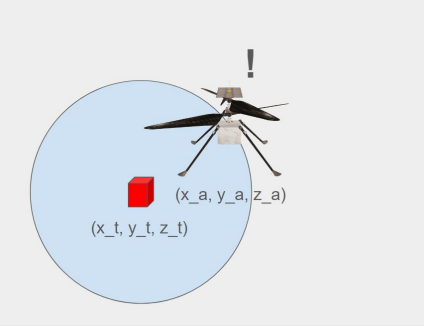
Hard Shield

Replaces the action with safe action.



Soft Shield

Gives Feedback

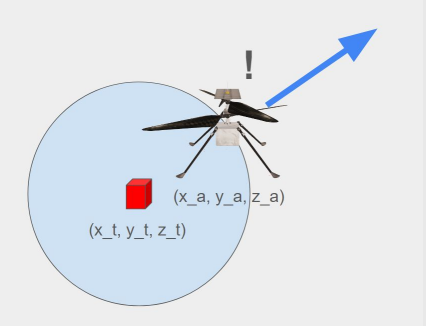


Hybrid Shield

Replaces action with safe action

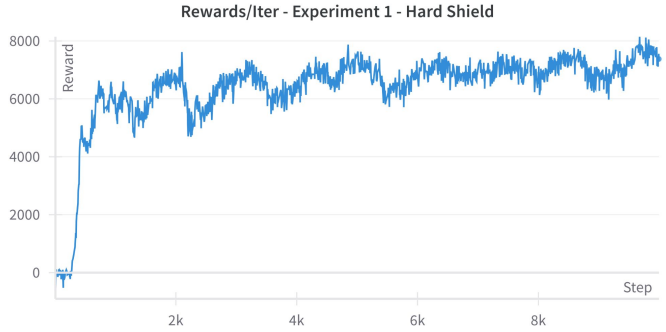
+

Gives Feedback

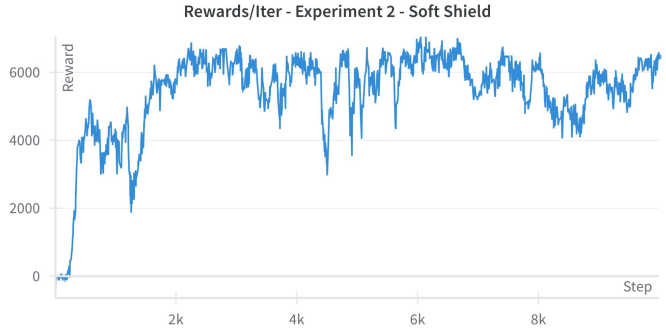


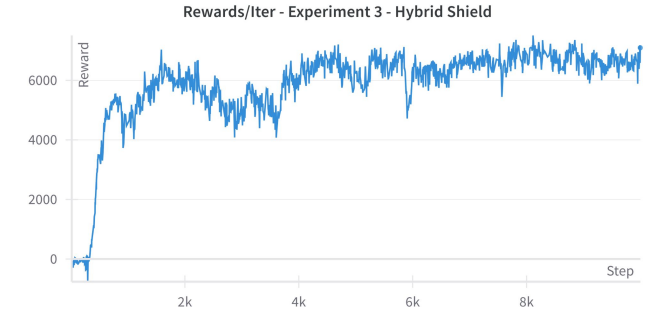
Ingenuity image source: https://science.nasa.gov/resource/mars-ingenuity-helicopter-3d-model

Results



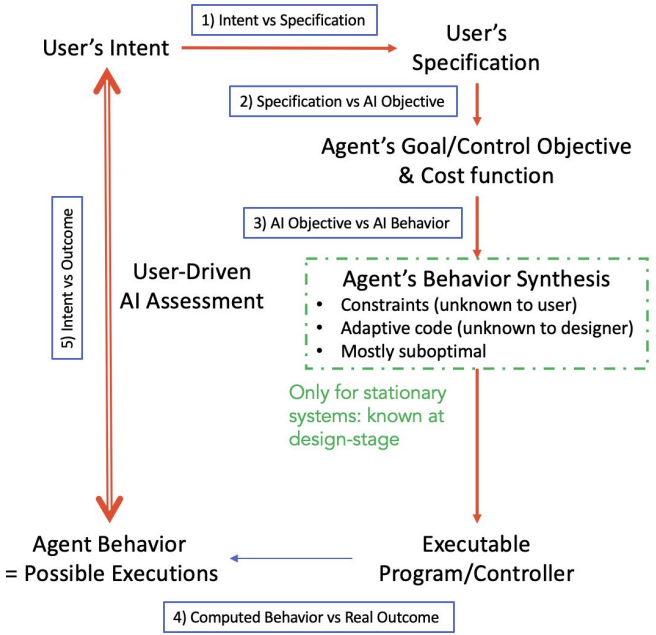
Experiment (Shield)	Reward at last step	Learning	Stability
0-No	-6567	No	No
1-Hard	**7283**	**Yes**	**Yes**
2-Soft	6581	Yes	No
3-Hybrid	6897	Yes	Yes





Shield → faster learning & convergence → safe behavior

Conclusion & Future work



Possible next project steps:

- Create more dynamic environments:
  - Different and moving types of obstacles and targets.
  - Environmental effects (eg. wind)
- Partial observability:
  - Without knowing the ground state, how can the shield be used to prevent unsafe behavior?
- Creating shields for multiple sources of risk.
  - How do you find actions that satisfy all safety requirements in a continuous action space.

Contributions

Daniel Bramblett

- Explored and tested other repositories: OpenAI Gym, gym-pybullet-drones
- Constructed the algorithm for redirecting the agent.

Vivek Sahukar

- Set up and modified Isaac Gym.
- Ran the experiments.
- Researched existing work on Safe AI with regards to drones.

Pair programming was used for coding the shielding experiments.

Thank you

Questions?