école normale supérieure paris–saclay

MATHÉMATIQUES VISION APPRENTISSAGE

# The DayDreamer: World Models for Physical Robot Learning

Research poster by: Ali Ahmadi & Manal Meftah
ali.ahmadi@ens-paris-saclay.fr
manal.meftah@ens-paris-saclay.fr

By : Philipp Wu, Alejandro Escontrela, Danijar Hafner, Ken Goldberg, Pieter Abbeel

Robotics Course 2023
Supervised by :
Stéphane Caron

## Context and Motivation

This study explores using a world model for predicting action outcomes to streamline planning and minimize trial and error.

By capitalizing on Dreamer's ability to learn from past experiences and refine behaviors through real-world interactions, this study aims to assess its effectiveness in enhancing robot learning capabilities. Dreamer is tested on four robots for direct, online real-world learning, bypassing the need for simulators.

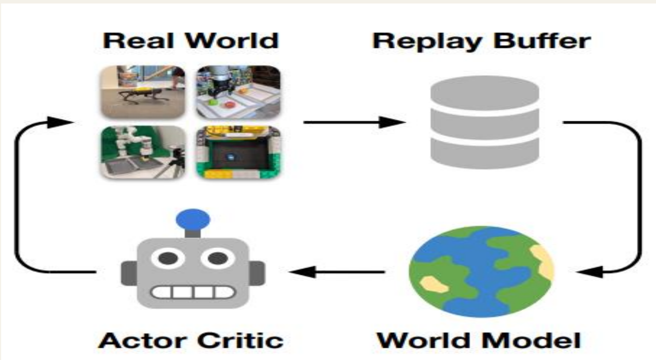


*Figure 1. Dreamer Pipeline*

## Method

***World Model Components:***

**Encoder**: This part takes in information from the robot's senses (like cameras or sensors) and combines it into a simplified, random representation.

**Dynamics** Network: This predicts what will happen next based on the current situation, creating a forecast of future scenarios.

**Decoder**: It reconstructs the sensory inputs from the predicted future representations, essentially trying to recreate what the robot should be seeing or sensing.

**Reward Network**: This learns to figure out how well the robot is doing in its tasks by predicting the rewards it might receive for certain actions.

***Behavior Model***:

The **actor-critic** algorithm is busy optimizing the policy network by running through **imagined scenarios** in a simplified version of the world model, called the latent space. This helps the robot get better at decision-making.

The data collected by the robot goes into a **replay buffer**, which is like a memory where the robot can later review and learn from past experiences.

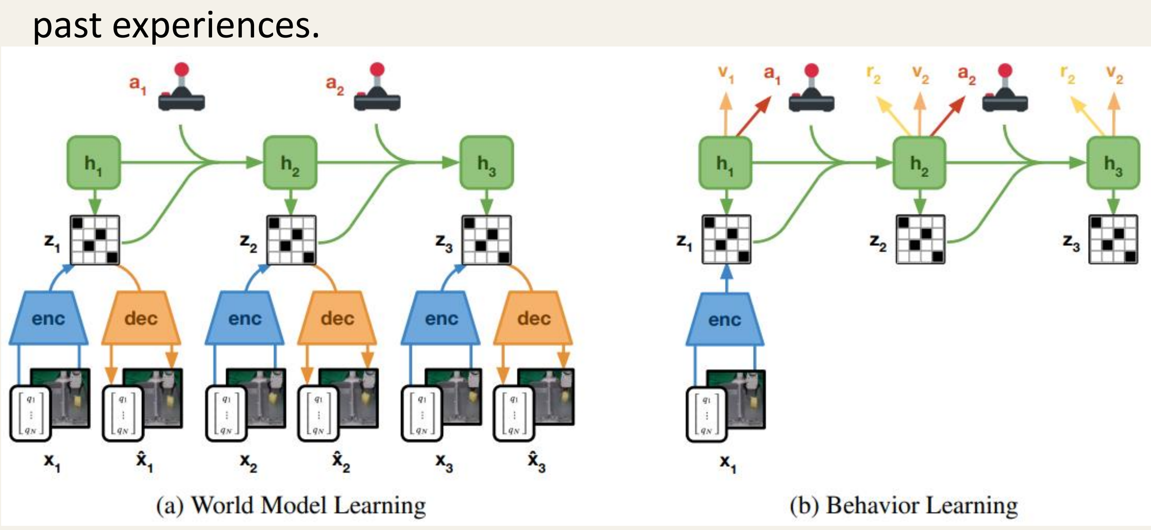


*Figure 2. Daydreamer Architecture*

## Experiments

***Objective :*** Exploring Physical Robot Learning Across Various Tasks and Platforms

***A1 Quadruped***

**Input Data:** Low-dimensional sensory inputs.

**Task:** Train a quadruped robot to autonomously stand up and walk starting from a position on its back.

**Challenge:** Rapidly learn complex locomotion patterns and physical adaptation to external forces.

**Outcome:** Achieved basic locomotion within one hour. Learned to resist pushes or recover from falls within 10 minutes.

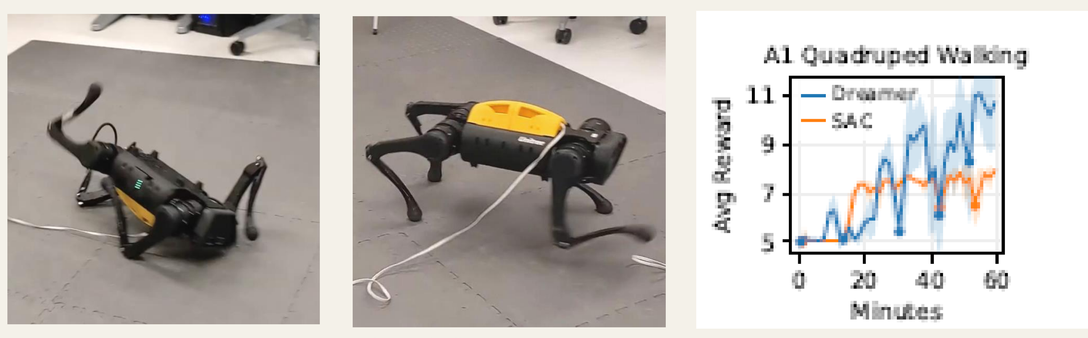


*Figure 3. A1 Quadruped Walking*

***UR5 Robotic Arm***

**Input Data:** Images and proprioceptive readings.

**Task:** Execute a multi-object visual pick and place operation in a simulated environment.

**Challenge:** Precisely manipulate objects based on visual cues and discrete action decisions.

**Outcome:** Achieved an average pick rate of 2.5 objects per minute within 8 hours of training.

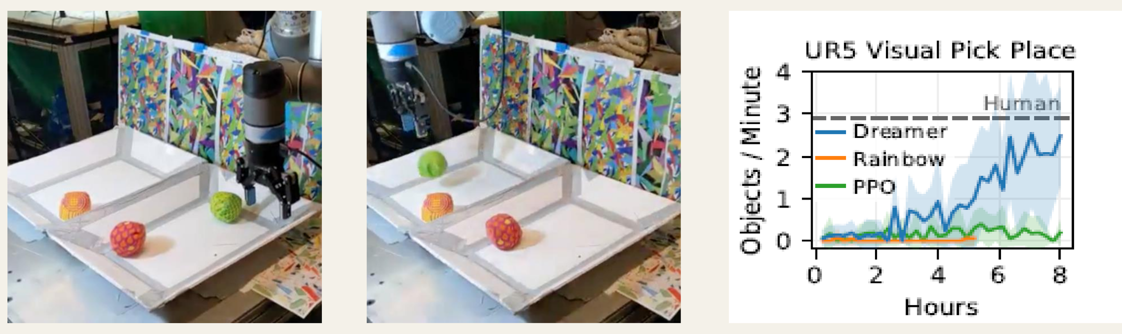


*Figure 4. UR5 Multi Object Visual Pick and Place*

***XArm Robotic Arm***

**Input Data:** Images combined with proprioceptive sensory data.

**Task:** Conduct a visual pick and place task with soft objects, moving them from one location to another.

**Challenge:** Adapt to the dynamic properties of soft objects and complex visual environments.

**Outcome:** Reached an average pick rate of 3.1 objects per minute over a 10-hour training period.

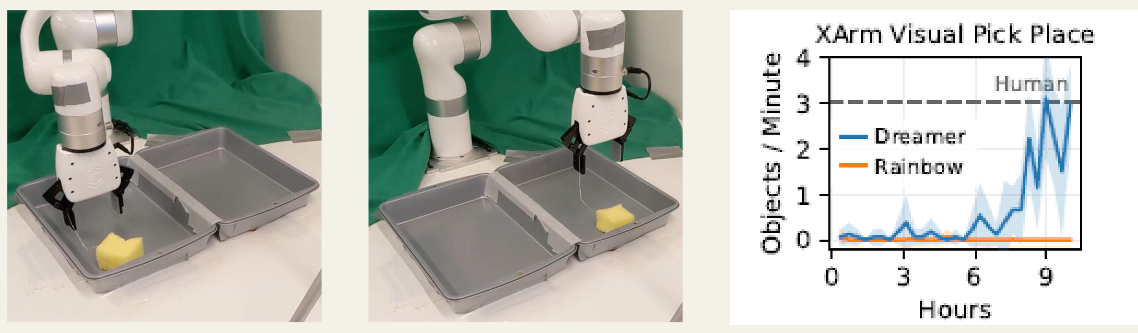


*Figure 5. XArm Visual Pick and Place*

***Sphero Robot***

**Input Data:** Camera images for navigation.

**Task:** Navigate through a complex environment with various obstacles.

**Challenge:** Efficiently process visual information for real-time navigation and obstacle avoidance.

**Outcome:** In 2 hours, Dreamer learned to quickly and consistently navigate to the goal, achieving an average distance to the goal of 0.15, demonstrating its effectiveness in complex navigation tasks.

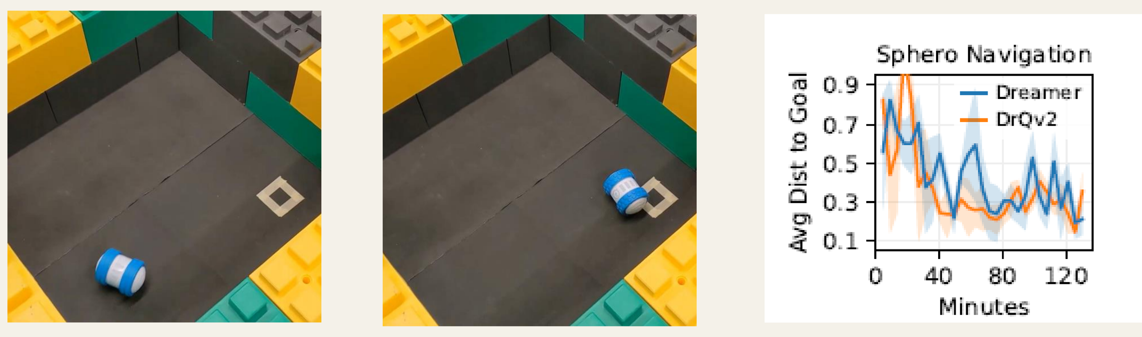


*Figure 6. Sphero Navigation*

## Results & Discussion

Dreamer demonstrated accelerated learning capabilities across various robotic platforms. It trained an A1 quadruped robot to walk within one hour—a process notably quicker than traditional methods requiring extensive simulation. In manipulation tasks, Dreamer showed rapid progress, achieving 2.5 and 3.1 objects per minute pick rates with UR5 and XArm robot arms, respectively, within 8-10 hours. Additionally, in a visual navigation task with a Sphero robot, Dreamer successfully learned to navigate to a goal in under 2 hours. These results illustrate Dreamer's proficiency in adapting to different tasks and conditions, performing on par with or exceeding human efficiency in specific tasks. In conclusion, the dreamer :

- **Enhanced Learning Efficiency :** Dreamer reduces robotic learning time across diverse platforms.
- **Performs Well in Sparse Rewards:** Outperforms traditional methods in challenging environments.
- **Has rapid Progress:** Notably fast learning observed in quadruped locomotion and robotic arms.
- **Has wide Applications:** Promising for industrial use, assistive robotics, and complex AI tasks.

## Limitations

**Physical Wear and Tear:** Continuous real-world learning may lead to wear and tear on the robots. Unlike simulations, the real world can cause physical damage to the robots over time.

**Safety Concerns:** Learning directly in the real world without simulators may pose safety risks, especially if the robot makes mistakes or encounters unexpected situations.

**Efficiency and Resource Usage:** Training in the real world can be time-consuming and resource-intensive. Simulators are often faster and more cost-effective for initial learning stages.

**Limited Exploration:** The real world might have limitations in terms of exploration compared to virtual environments. Simulators allow for more diverse scenarios and training conditions.

**Transferability:** The learned behaviors may not easily transfer to different environments or robot models. Simulators often provide more transferable knowledge.

Data Collection Challenges:

**Real-world data collection can be noisy and unpredictable**. This variability may affect the quality of data used for training the robot.

**Task Complexity:** While Dreamer shows success in simpler tasks, its effectiveness in handling highly complex tasks is not guaranteed. Some tasks may still require a combination of real-world and simulator training.

**Need for Calibration:** Robots may need frequent recalibration due to changes in the environment or wear and tear, making continuous real-world learning more challenging.

**Generalization:** The ability of the robot to generalize its learned behaviors to new situations or environments is a concern. Simulators often provide a controlled environment for better generalization.

**Integration with Simulators:** Combining real-world learning with simulator training might be necessary for more complex tasks. Integrating both approaches could enhance the overall learning process.

## Ideas to Explore

***Improving Algorithm by Adding Two Learning Styles:***

**Concept:** We're suggesting a way to improve the algorithm by using not one, but two learning styles. Each style does its own thing, making the algorithm smarter.

**Good Sides:**

1. **Learn More Stuff:** With two learning styles, the algorithm can learn many different things, making it better at doing different tasks.
2. **Change and Learn:** The algorithm can adapt better to new things because each learning style can focus on a special job.
3. **Become an Expert:** Each learning style can become good at one part of the job, so the algorithm becomes an expert in different areas.
4. **Learn Faster:** By using both learning styles at the same time, the algorithm might learn and make decisions faster.

**Not-So-Good Sides:**

1. **More Complicated:** Having two learning styles makes things more complicated. It might be harder to set up and make sure it keeps working well.
2. **Styles might Fight:** The two learning styles might not always work together perfectly. We have to be careful they don't interfere with each other and cause problems.

## Conclusion

The Dreamer approach has revolutionized robot learning, demonstrating a remarkable ability to teach robots diverse tasks efficiently in real-world environments. Its success in rapidly training a quadruped robot to stand and walk from scratch in just one hour is a testament to its effectiveness, bypassing the traditional reliance on simulators.

While Dreamer's achievements are notable, its application is not without challenges. Continuous operation leads to physical wear on robot components, raising concerns about durability and the need for maintenance. Furthermore, the approach's full capabilities and boundaries, particularly over longer training periods, are yet to be fully understood and explored.

Looking ahead, the integration of Dreamer's real-world learning efficiencies with the controlled environments of simulators presents an exciting frontier. This synergy could potentially unlock new levels of complexity in robot learning, paving the way for more advanced and robust robotic applications in various fields.