# Critical Sumary for Action Priors for Large Action Spaces in Robotics

Wen Dong
University of Windsor
*#110057395*
Windsor, Ontario
dong23@uwindsor.ca

***Abstract*—This article is a critical summary for paper Critical Sumary for Action Priors for Large Action Spaces in Robotics.**

***Keywords*— AI, deep learning, action prior, Robotics, Action Spaces, Summary.**

## I. About the Authors

Ondrej Biza is a doctoral student in the Khoury College of Computer Science at Northeastern University, specifically interested in reinforcement learning framework for sequential decision making, and Robotics [1]; Robert Platt is an assistant professor at the Khoury College of computer science, a former NASA employee, work and research primarily on perception, planning, and control for robotic manipulation [2]; Dian Wang is a PHD student in computer science in the Northeastern University with a specialization in Machine Learning and Robotics, interested in applied reinforcement learning and imitation learning on Robotics [3]; Jan-Willem van de Meent is also an assistant professor at the Khoury College of Computer Science in Northeastern University, his research interests is interface programming languages and machine learning, with his team is working on combining probabilistic programming with deep learning to create models for data science, machine learning and artificial intelligence [4]; Lawson L.S. Wong is an assistant professor in Khoury College at the Northeastern University, focused on learning general abstracted real-word knowledge, representing, estimating, and applying the learned in robotic applications.

## II. What's the paper about

The paper introduced a novel approach to improve the model training process of robotics by summarizing and storing the policies and knowledge learned from previously solved tasks in a fully convolutional neural network to generate action priors to facilitate the learning of new tasks. In robotics, it's usually impossible to use model-free reinforcement learning to learn policies without reward shaping and curriculum learning due to the complexity of robotics, therefore, expert demonstration is often used for learning guidance which can be very expensive to acquire, this is the reason the paper proposes the solution taking advantage of previously learned tasks rather than expert resources. The paper also elaborates all the experiments it conducted to verify the solution.

## III. Hypothesises

Ideally, robotic agent should be able to learn by itself progressively without additional supervision.

## IV. approach

This paper uses policies generated from previously learned tasks to manipulate the robotic agent toward the goal state. Basically, the approach can be divided into policies summarizing from training tasks and leveraging the policies to instruct the agent.

### *A. Learning Action Priors*

It trains action priors in an environment of image states and pixel-wise action spaces, the action spaces can be very large, in this environment, it trained a set of training tasks and uses a set of separate but similar testing tasks to validate the learned action priors or policies from the training tasks. It uses imitation learning to train the tasks and store the learned expert policies in a fully convolutional neural network, and in the testing circle, mark each state-action pair with a reward that reflects how optimal the action is for the given state, it also trained task classifier from visited states and optional action pairs and uses the task classifier to identify relevant actions set for each state, as not all actions are applicable to a specific state, for example, removing a building block from a building will never happen in a building task.

### *B. Apply Action Priors to exploration*

Once the action prior set is built, we can use it for action prior exploration for new task, for each state in the exploration, the reinforcement learning agent built in the learning stage is responsible for recommending the optimal action based on the policies stored in the agent, and then applies the action to the state to drive it toward its goal, for state without a satisfying optimal action, use a random action instead to move forward, as optimal action might be not available in the learning agent due to the sample inefficiency or insufficiency.

## V. Tools and Development

The work used several existing libraries and framework in the implementation, it uses PyTorch for neural networks building and training, and PyBullet physics simulator for robotic simulation, it also used another third-party library "helping hands rl envs" to support PyBullet environment simulation. [6], and a real-world UR5 robotic arm.

It uses Python as the programming language to develop the project, explored a large amount of imitation training

dataset to training the learning agent convolutional neural networks, and created MongoDB database for storing intermediate and final learning results, for the computing resources, it used a machine with four NVIDIA RTX 2080 Ti GPUs, 256 GB memory of RAM and AMD Rizen Threaddripper CPU.

## VI. Proof of Hypothesise

It proved the hypothesis by experiments in two distinct domains, a proof-of-concept Fruits World and a block stacking robotic manipulation. Policies learned can be deployed in a real-world UR5 robotic arm in combination with high-depth top-down cameras to verify whether the action priors can be used to enable a deep Q-learning to perform a learning task that were previously impossible.
For the proof-of-concept Fruits World, experimental results shows that Action priors significantly increase the learning performance and eventually solved any fruits sequence tasks; For the block stacking experiments, it executed two experiments in the domain, one focused on model-free learning with action priors accompanied and compared with DQN RS and DQN HS respectively, and the other one focused on solely exploration with action priors. The testing results shows that our approach is able to solve the problem in the experiments and is of higher success rate comparing to the traditional approaches.

## VII. Limitation and drawbacks

The paper does not clarify the difference between expert supervision learning in conventional model training and the imitation learning adopted in this novel approach, as the highlight of the approach is the claimed learning without expensive expert guidance, so it is crucial to elaborate that the imitation learning is not an expert supervision learning.

## VIII. Future Work

The paper indicates the an important part of the future work is to enable the learning agent to learn progressively online during training on a new task, in this way, the agent become stronger and smarter during exploration overtime without additional dedicated training tasks.


[1] https://www.khoury.northeastern.edu/people/ondrej-biza/
[2] https://www.khoury.northeastern.edu/people/robert-platt/
[3] https://www.khoury.northeastern.edu/people/dian-wang/
[4] https://www.khoury.northeastern.edu/people/jan-willem-van-de-meent/
[5] https://www.khoury.northeastern.edu/people/lawson-wong/
[6] https://github.com/ondrejba/action_priors
[7] https://pointw.github.io/asrse3-page/


Commented [DD1]: