Paper:1

1. Title: “Good Robot!”: Efficient Reinforcement Learning for Multi-Step Visual Tasks with Sim to Real Transfer
2. Authors: Andrew Hundt, Benjamin Killeen, Nicholas Greene, Hongtao Wu, Heeyeon Kwon, Chris Paxton, and Gregory D. Hager
3. Affiliation: Andrew Hundt, Benjamin Killeen, Nicholas Greene, Hongtao Wu, Heeyeon Kwon, and Gregory D. Hager are with The Johns Hopkins University, Baltimore, MD 21218 USA; Chris Paxton is with NVIDIA, Seattle, WA, 98105 USA
4. Keywords: Computer vision for other robotic applications, deep learning in grasping and manipulation, reinforcement learning
5. Urls: Paper: https://ieeexplore.ieee.org/document/9161711, Github code: https://github.com/jhulcsr/good_robot
6. Summary:

- (1): The research background of this paper is that current Reinforcement Learning (RL) algorithms have difficulty with long-horizon tasks where time can be wasted exploring dead ends and task progress may be easily reversed.
- (2): Past methods have struggled with incorporating common sense constraints, leading to inefficient learning. The authors of this paper propose the Schedule for Positive Task (SPOT) framework, which incorporates these constraints to accelerate learning and task efficiency. The approach is well motivated by the observation that RL wastes significant time exploring unproductive actions.
- (3): The research methodology proposed in this paper is the SPOT framework, which explores within action safety zones, learns about unsafe regions without exploring them, and prioritizes experiences that reverse earlier progress to learn with remarkable efficiency.
- (4): The SPOT framework successfully completes simulated trials of a variety of tasks, improving a baseline trial success rate from 13% to 100% when stacking 4 cubes, from 13% to 99% when creating rows of 4 cubes, and from 84% to 95% when clearing toys arranged in adversarial patterns. Efficiency with respect to actions per trial typically improves by 30% or more, while training takes just 1-20 k actions, depending on the task. Furthermore, the authors demonstrate direct sim to real transfer, creating real stacks in 100% of trials with 61% efficiency and real rows in 100% of trials with 59% efficiency. This is the first instance of RL with successful sim to real transfer applied to long-term multi-step tasks such as block-stacking and row-making with consideration of progress reversal. The achieved performance supports the authors’ goals of efficiently completing long-term multi-step tasks with RL.

7. Methods:

- (1): The methodological idea of this article is to propose the Schedule for Positive Task (SPOT) framework for efficient reinforcement learning (RL) for long-horizon tasks with sparse and approximate notion of task progress. SPOT framework incorporates common sense constraints to accelerate learning and task efficiency, and it explores within action safety zones, learns about unsafe regions without exploring them, and prioritizes experiences that reverse earlier progress to improve efficiency.

- (2): The authors frame the problem as a Markov Decision Process (S, A, P, R), with a simplifying assumption equating sensor observations and state. They use Q-learning to produce a deterministic policy for choosing actions that estimates the expected reward of an action from a given state. The authors also introduce reward shaping techniques to optimize R, including baseline rewards, an approximate task progress function, and Situation Removal (SR) and SPOT Trial Reward functions to propagate rewards effectively across the whole trial.

- (3): The authors leverage a priori knowledge about the environment to reduce unproductive attempts and accelerate training by assuming the existence of an oracle, M(st, a) $\rightarrow \{0, 1\}$, to predict certain action failures, and they introduce SPOT-Q Learning that replaces the standard discounted reward with a new target value function that masks out failed actions and prioritizes exploration of unmasked actions with high Q-values. The authors demonstrate the effectiveness of the SPOT framework in simulated trials of a variety of tasks, improving a baseline trial success rate and efficiency by 30% or more.

8. Conclusion:

- (1): The significance of this work lies in the proposal of the Schedule for Positive Task (SPOT) framework, which incorporates common sense constraints to efficiently complete long-horizon tasks with sparse and approximate notion of task progress using reinforcement learning (RL). Additionally, the authors demonstrate successful sim to real transfer of multi-step visual tasks such as block-stacking and row-making, which is the first instance of RL achieving this in these types of tasks with consideration of progress reversal.
- (2): Innovation point: The SPOT framework proposes a new approach to incorporating common sense constraints in RL for long-horizon tasks, which significantly improves efficiency and task completion rate. Performance: The SPOT framework improves baseline trial success rate and efficiency by 30% or more in simulated trials of a variety of tasks, and achieves 100% completion rate in real block-stacking and row-making tasks with successful sim to real transfer. Workload: The training takes just 1-20 k actions, depending on the task, which is a relatively low workload compared to other RL algorithms. However, the limited applicability of the framework to tasks that require intermediate rewards may be a weakness that needs to be addressed in future research. Additionally, the manual design of action space mask M may limit its scalability to more complex tasks.