PLAYING ATARI WITH DEEP REINFORCEMENT LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

In this paper, we propose a novel deep reinforcement learning (DRL) algorithm for achieving human-level performance in playing Atari games. The algorithm integrates deep neural networks with reinforcement learning techniques, enabling the agent to learn control policies directly from high-dimensional sensory inputs. We address the limitations of existing methods, such as overestimations in Q-values and the reliance on off-policy learning, by incorporating innovative techniques and strategies. Furthermore, we explore the use of alternative neural network architectures, such as recurrent and attention-based models, to better capture the temporal dependencies and spatial relationships in Atari game environments. The proposed algorithm is evaluated on a range of Atari games and compared with state-of-the-art methods, demonstrating its potential to advance the field of DRL and contribute to the development of AI systems capable of tackling complex decision-making problems in various domains.

1 INTRODUCTION

Reinforcement learning (RL) has emerged as a powerful technique for solving complex decision-making problems in various domains, including robotics, finance, and healthcare. In recent years, deep reinforcement learning (DRL) has gained significant attention in the artificial intelligence (AI) community due to its ability to learn high-level abstractions from raw sensory inputs and achieve human-level performance in various tasks (Mnih et al., 2013; Berner et al., 2019). One of the most popular benchmarks for evaluating DRL algorithms is the Atari 2600 game suite, which offers a diverse set of challenges that require agents to learn sophisticated strategies and adapt to different environments (Mnih et al., 2013).

The problem addressed in this paper is to develop a DRL algorithm capable of achieving human-level performance in playing Atari games. The proposed solution is based on the integration of deep neural networks with reinforcement learning techniques, which allows the agent to learn control policies directly from high-dimensional sensory inputs. The specific research objectives include the design and implementation of an efficient DRL algorithm, the evaluation of its performance on a range of Atari games, and the comparison of its performance with existing state-of-the-art methods.

Several key works have contributed to the development of DRL algorithms for Atari games. Mnih et al. (2013) introduced the Deep Q-Network (DQN) algorithm, which combines Q-learning with deep convolutional neural networks to learn control policies from raw pixel inputs. Wang et al. (2015) proposed the dueling network architecture, which improves policy evaluation in the presence of many similar-valued actions. Hessel et al. (2017) combined six extensions to the DQN algorithm to achieve state-of-the-art performance on the Atari 2600 benchmark in terms of data efficiency and final performance.

The main differences between the proposed work and existing approaches lie in the design of the DRL algorithm, the choice of neural network architecture, and the optimization techniques employed. The proposed algorithm aims to address the limitations of current methods, such as overestimations in Q-values (Hasselt et al., 2015) and the reliance on off-policy learning (Fujimoto et al., 2018), by incorporating novel techniques and strategies. Furthermore, the proposed work will explore the use of alternative neural network architectures, such as recurrent and attention-based models, to better capture the temporal dependencies and spatial relationships in Atari game environments.

In conclusion, this paper presents a novel DRL algorithm for achieving human-level performance in playing Atari games. By addressing the limitations of existing methods and exploring alternative neural network architectures, the proposed work aims to advance the state of the art in DRL and contribute to the ongoing development of AI systems capable of tackling complex decision-making problems in various domains. The evaluation of the proposed algorithm on the Atari 2600 benchmark will provide valuable insights into its performance and potential for further improvement.

2 RELATED WORKS

Deep Reinforcement Learning: Deep reinforcement learning (DRL) has gained significant attention in recent years due to its success in various applications. Mnih et al. (2013) introduced the first DRL model that learned control policies directly from high-dimensional sensory input, achieving impressive results on Atari games. However, it struggled with games that required long-term planning. Hasselt et al. (2015) proposed an adaptation to the DQN algorithm that reduced overestimations and improved performance on several games. Wang et al. (2015) introduced a new neural network architecture for model-free reinforcement learning, leading to better policy evaluation and state-of-the-art performance on the Atari 2600 domain. Hessel et al. (2017) combined six extensions to the DQN algorithm, achieving state-of-the-art performance in terms of data efficiency and final performance. Mnih et al. (2016) presented a lightweight framework for DRL that used asynchronous gradient descent for optimization, showing success on a wide variety of continuous motor control problems.

Continuous Control with Deep Reinforcement Learning: Continuous control tasks have been a major focus in DRL research. Lillicrap et al. (2015) presented an actor-critic, model-free algorithm based on the deterministic policy gradient that operates over continuous action spaces and learns policies end-to-end from raw pixel inputs. Haarnoja et al. (2018) proposed soft actor-critic, an off-policy actor-critic DRL algorithm based on the maximum entropy reinforcement learning framework, achieving state-of-the-art performance on various continuous control benchmark tasks. Fujimoto et al. (2018) introduced a novel class of off-policy algorithms called batch-constrained reinforcement learning, which restricts the action space to force the agent towards behaving close to on-policy with respect to a subset of the given data.

Visual Reinforcement Learning: Visual reinforcement learning focuses on learning policies from raw visual input. Oh et al. (2015) was the first to make and evaluate long-term predictions on high-dimensional video conditioned by control inputs, proposing two deep neural network architectures based on convolutional neural networks and recurrent neural networks. Ye et al. (2021) proposed a sample-efficient model-based visual RL algorithm built on MuZero, achieving super-human performance on Atari games with minimal data. Mengistu et al. (2022) presented a solution that learns state representations from sparsely sampled or randomly shuffled observations, marginally enhancing the representation powers of encoders to capture high-level latent factors of the agents' observations.

Deep Reinforcement Learning in Complex Environments: DRL has been applied to complex environments such as Dota 2 and unmanned ship path planning. Berner et al. (2019) demonstrated that self-play reinforcement learning can achieve superhuman performance on a difficult task by defeating the Dota 2 world champion (Team OG). Wu et al. (2022) used DRL to solve the optimization problem in the path planning and management of unmanned ships, minimizing the total travel time of unmanned ships passing through the path.

Safety and Evaluation in Deep Reinforcement Learning: Ensuring safety and reliable evaluation in DRL is crucial for real-world applications. Agarwal et al. (2021) argued that reliable evaluation in the few-run deep RL regime cannot ignore the uncertainty in results and advocated for reporting interval estimates of aggregate performance and proposing performance profiles to account for variability. Giacobbe et al. (2021) presented the first exact method for analyzing and ensuring the safety of DRL agents for Atari games, improving the safety of all agents over multiple properties.

3 BACKGROUNDS

Deep reinforcement learning (DRL) has emerged as a powerful technique for solving complex control problems, including playing Atari games (Mnih et al., 2013). The central problem in this field is to learn control policies directly from high-dimensional sensory input, such as raw pixel data, to achieve human-level performance in various tasks.

One of the foundational concepts in DRL is the use of deep neural networks (DNNs) as function approximators to estimate the value function or the Q-function. The Q-function, denoted as $Q(s, a)$, represents the expected cumulative reward when taking action a in state s and following a specific policy thereafter. The goal of DRL is to learn an optimal policy $\pi^*(s)$ that maximizes the expected cumulative reward. The Q-function obeys the Bellman equation:

$$Q(s, a) = \mathbb{E}_{s', r \sim p(.|s,a)}[r + \gamma \max_{a'} Q(s', a')], \tag{1}$$

where p is the transition probability, r is the immediate reward, and γ is the discount factor.

A popular algorithm for solving DRL problems is the Deep Q-Network (DQN) (Mnih et al., 2013), which combines Q-learning with DNNs. DQN uses a neural network $Q(s, a; \theta)$ parameterized by θ to approximate the Q-function. The loss function for updating the network parameters is given by:

$$L(\theta) = \mathbb{E}_{(s,a,r,s') \sim \mathcal{D}} \left[\left(r + \gamma \max_{a'} Q(s', a'; \theta^-) - Q(s, a; \theta) \right)^2 \right], \tag{2}$$

where $\mathcal{D}$ is a replay buffer containing past experiences, and θ^- are the parameters of a target network that is periodically updated to stabilize learning.

Several extensions to the DQN algorithm have been proposed to improve its performance, such as Double DQN (Hasselt et al., 2015), Dueling DQN (Wang et al., 2015), and Rainbow DQN (Hessel et al., 2017). These extensions address various challenges, such as overestimation of Q-values and handling many similar-valued actions.

Another important concept in DRL is the actor-critic framework, which combines value-based and policy-based methods. Soft Actor-Critic (SAC) (Haarnoja et al., 2018) is an off-policy actor-critic algorithm based on the maximum entropy reinforcement learning framework. SAC introduces an entropy term to the objective function, encouraging exploration and robustness. The policy optimization in SAC is performed by minimizing the following objective:

$$J(\pi) = \mathbb{E}_{s \sim \mathcal{D}, a \sim \pi} \left[\alpha \log \pi(a|s) - Q(s, a) \right], \tag{3}$$

where α is the temperature parameter controlling the trade-off between exploration and exploitation.

In summary, DRL has shown great success in solving complex control problems, such as playing Atari games. The field relies on foundational concepts like Q-learning, deep neural networks, and the actor-critic framework. Various algorithms, including DQN and its extensions, as well as SAC, have been developed to address the challenges of learning control policies directly from high-dimensional sensory input.

REFERENCES

Rishabh Agarwal, Max Schwarzer, P. S. Castro, Aaron C. Courville, and Marc G. Bellemare. Deep reinforcement learning at the edge of the statistical precipice. *Neural Information Processing Systems*, 2021. URL dblp.org/rec/journals/corr/abs-2108-13264.

Christopher Berner, Greg Brockman, Brooke Chan, Vicki Cheung, Przemyslaw Debiak, Christy Dennison, David Farhi, Quirin Fischer, Shariq Hashme, Christopher Hesse, R. Jzefowicz, Scott Gray, Catherine Olsson, J. Pachocki, Michael Petrov, Henrique Pond de Oliveira Pinto, Jonathan Raiman, Tim Salimans, Jeremy Schlatter, Jonas Schneider, Szymon Sidor, Ilya Sutskever, Jie Tang, F. Wolski, and Susan Zhang. Dota 2 with large scale deep reinforcement learning. *arXiv.org*, 2019. URL dblp.org/rec/journals/corr/abs-1912-06680.

Scott Fujimoto, D. Meger, and Doina Precup. Off-policy deep reinforcement learning without exploration. *International Conference on Machine Learning*, 2018. URL dblp.org/rec/journals/corr/abs-1812-02900.

Mirco Giacobbe, Mohammadhosein Hasanbeig, D. Kroening, and H. Wijk. Shielding atari games with bounded prescience. *Adaptive Agents and Multi-Agent Systems*, 2021. URL dblp.org/rec/journals/corr/abs-2101-08153.

Tuomas Haarnoja, Aurick Zhou, P. Abbeel, and S. Levine. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. *International Conference on Machine Learning*, 2018. URL dblp.org/rec/conf/icml/HaarnojaZAL18.

H. V. Hasselt, A. Guez, and David Silver. Deep reinforcement learning with double q-learning. *AAAI Conference on Artificial Intelligence*, 2015. URL dblp.org/rec/journals/corr/HasseltGS15.

Matteo Hessel, Joseph Modayil, H. V. Hasselt, T. Schaul, Georg Ostrovski, Will Dabney, Dan Horgan, Bilal Piot, M. G. Azar, and David Silver. Rainbow: Combining improvements in deep reinforcement learning. *AAAI Conference on Artificial Intelligence*, 2017. URL dblp.org/rec/journals/corr/abs-1710-02298.

T. Lillicrap, Jonathan J. Hunt, A. Pritzel, N. Heess, T. Erez, Yuval Tassa, David Silver, and Daan Wierstra. Continuous control with deep reinforcement learning. *International Conference on Learning Representations*, 2015. URL dblp.org/rec/journals/corr/LillicrapHPHETS15.

Menore Tekeba Mengistu, G. Alemu, Pierre Chevaillier, and P. D. Loor. Unsupervised learning of state representation using balanced view spatial deep infomax: Evaluation on atari games. *International Conference on Agents and Artificial Intelligence*, 2022. URL dblp.org/rec/conf/icaart/MengistuACL22.

Volodymyr Mnih, K. Kavukcuoglu, David Silver, A. Graves, Ioannis Antonoglou, Daan Wierstra, and Martin A. Riedmiller. Playing atari with deep reinforcement learning. *arXiv.org*, 2013. URL dblp.org/rec/journals/corr/MnihKSGAWR13.

Volodymyr Mnih, Adri Puigdomnech Badia, Mehdi Mirza, A. Graves, T. Lillicrap, Tim Harley, David Silver, and K. Kavukcuoglu. Asynchronous methods for deep reinforcement learning. *International Conference on Machine Learning*, 2016. URL dblp.org/rec/journals/corr/MnihBMGLHSK16.

Junhyuk Oh, Xiaoxiao Guo, Honglak Lee, Richard L. Lewis, and Satinder Singh. Action-conditional video prediction using deep networks in atari games. *NIPS*, 2015. URL dblp.org/rec/journals/corr/OhGLLS15.

Ziyun Wang, T. Schaul, Matteo Hessel, H. V. Hasselt, Marc Lanctot, and N. D. Freitas. Dueling network architectures for deep reinforcement learning. *International Conference on Machine Learning*, 2015. URL dblp.org/rec/journals/corr/WangFL15.

Dingwei Wu, Yin Lei, Maoen He, Chunjiong Zhang, and Lili Ji. Deep reinforcement learning-based path control and optimization for unmanned ships. *Wireless Communications and Mobile Computing*, 2022.

Weirui Ye, Shao-Wei Liu, Thanard Kurutach, P. Abbeel, and Yang Gao. Mastering atari games with limited data. *Neural Information Processing Systems*, 2021. URL dblp.org/rec/journals/corr/abs-2111-00210.