Application of Rainbow DQN and Curriculum Learning in Atari Breakout

Hengqian Wu[a]
The Affiliated High School of Peking University, Yiheyuan Street 5, Beijing, China

Keywords: Reinforcement Learning, Rainbow Deep Q-Network, Curriculum Learning, Atari Breakout, Deep Q-Network.

Abstract: Reinforcement Learning has become a crucial area within artificial intelligence, particularly when it comes to applying these techniques in environments like video games, robotics, and autonomous systems. The Deep Q-Network, introduced by DeepMind in 2015, marked a significant advancement by enabling AI agents to play Atari games directly from raw pixel inputs. However, this network encounters issues with large state spaces and tends to overestimate action values, leading to inefficient learning and not-so-optimal performance. To overcome these limitations, Rainbow DQN integrates several enhancements, such as Double Q-learning, Prioritized Experience Replay, and Noisy Networks, which together greatly improve the algorithm's performance. Additionally, Curriculum Learning, which systematically escalates task difficulty, mimics the human learning process, and enhances the agent's efficiency. This paper delves into the combination of Rainbow DQN and Curriculum Learning within the context of Atari Breakout, offering a detailed look at how these techniques work together to boost both the agent's game score and learning speed. Experimental outcomes display that this method significantly improves the agent's game score, learning pace, and overall adaptability in complex scenarios.

1 INTRODUCTION

Reinforcement Learning is a kind of machine learning where agents are trained to make a series of decisions based on their interactions with the environment. Different from supervised learning, which relies on labelled data, Reinforcement Learning agents improve their decision-making by responding to feedback, such as rewards for correct actions or penalties for mistakes. This method has become increasingly popular in fields like gaming, robotics, and the development of autonomous vehicles.

The introduction of the Deep Q-Network by DeepMind in 2015 was a ground-breaking moment for Reinforcement Learning (Mnih et al., 2015). The Deep Q-Network combined the principles of Q-learning, a fortification learning algorithm, with deep neural networks, allowing AI agents to learn directly from high-dimensional sensory inputs like raw pixel data from Atari game screens (Zhou, Yao, Xiao, et al. 2022). Despite its success, the Deep Q-Network has been found to struggle with large state spaces, often leading to slower learning and suboptimal performance (Van Hasselt et al., 2016; Bellemare et al., 2017). Moreover, it tends to overestimate action values, which results in inefficient learning and unstable policies (Schaul et al., 2015).

These challenges led to the development of Rainbow Deep Q-Network, an enhanced version of Deep Q-Network that incorporates multiple improvements. Rainbow Deep Q-Network uses several advanced techniques to address the limitations of the traditional Deep Q-Network. For example, Double Q-learning reduces overestimation bias by using two separate networks for action-value estimation (Van Hasselt et al., 2016), while Prioritized Experience Replay speeds up the learning process by allowing the mediator to focus more on significant experiences (Schaul et al., 2015). Another key feature is Noisy Networks, which introduces randomness into the decision-making process, encouraging exploration and preventing the agent from settling too early on suboptimal strategies (Fortunato et al., 2018).

In addition to these improvements, Curriculum Learning is a strategy that has gained traction in the

[a] https://orcid.org/0009-0003-7244-8795

Wu, H.
Application of Rainbow DQN and Curriculum Learning in Atari Breakout.
DOI: 10.5220/0013245500004558
In *Proceedings of the 1st International Conference on Modern Logistics and Supply Chain Management (MLSCM 2024)*, pages 163-166
ISBN: 978-989-758-738-2

field of Reinforcement Learning. Inspired by how humans learn, Curriculum Learning involves structuring learning tasks from simple to complex, allowing the agent to build on previously acquired skills as it progresses (Bengio et al., 2009). By gradually increasing the difficulty of tasks, Curriculum Learning helps the agent develop a deeper understanding of the environment, leading to faster and more robust learning.

This paper explores the combined application of Rainbow Deep Q-Network and Curriculum Learning in the Atari Breakout game. Through a series of experiments, the paper shows that this combined approach not only improves the agent's game score but also significantly accelerates the learning process. Furthermore, integrating practices such as Prioritized Experience Replay and Noisy Systems within Rainbow Deep Q-Network has proven effective in enhancing agent performance across various Reinforcement Learning tasks (Hessel et al., 2018; Fortunato et al., 2018). The findings of this study contribute to the ongoing development of more effective and effective Reinforcement Learning algorithms capable of managing increasingly complex environments.

2 MANUSCRIPT PREPARATION

The foundation of both DQN and Rainbow DQN is rooted in Q-learning, a model-free approach in reinforcement learning. This method aims to approximate the optimal action-value function, denoted as Q*, which predicts the expected total reward for executing a particular action in a given state. DQN employs a deep neural network to estimate this function, enabling the agent to effectively extrapolate across diverse state spaces. Rainbow DQN enhances the standard DQN architecture by integrating multiple crucial enhancements, significantly improving its performance and stability.

Double Q-learning: By maintaining two separate networks, Rainbow DQN reduces the overestimation of Q-values that typically plagues single-network approaches. One network selects actions while the other evaluates them, leading to more accurate value estimates.

Prioritized Experience Replay: Traditional experience replay treats all past experiences equally. However, Rainbow DQN prioritizes experiences based on their significance, ensuring that the agent spends more time learning from the most informative events.

Noisy Networks: By introducing randomness into the network's weights, Noisy Networks serve as a form of implicit exploration. This mechanism prevents the agent from getting trapped in local optima and encourages it to discover more diverse strategies (Fortunato et al., 2018).

2.1 Network Architecture

The Rainbow DQN model's architecture is designed to efficiently process visual inputs and translate them into effective in-game actions.

The process begins with Input Data that undergoes Data Preprocessing to ensure it's in the correct format for model training. The preprocessed data is then fed into the Model Training phase, where the network is optimized. The trained model updates the Policy, which is then used to produce the Decision Output that dictates the agent's actions. The specific network construction is shown in Figure 1.

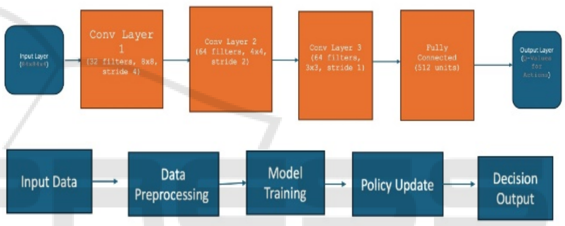

Figure 1: Network Architecture(Photo/Picture credit: Original).

The network starts with an Input Layer that receives 84x84x4 images. The data is passed through three convolutional layers:

- Convolutional Layer 1: 32 filters with an 8x8 kernel and a stride of 4 extract basic features like edges and textures.
- Convolutional Layer 2: 64 filters with a 4x4 kernel and a stride of 2 build upon the initial features, identifying more complex patterns.
- Convolutional Layer 3: 64 filters with a 3x3 kernel and a stride of 1 further refine the features, focusing on finer details essential for gameplay decisions.

The features extracted by the convolutional layers are then fed into a Fully Connected Layer with 512 units, which integrates these features to output the Q-values for Actions that guide the agent's decisions during gameplay.

This multi-layered architecture enables the agent to effectively interpret complex visual data and make informed decisions—a critical capability for achieving high performance in Atari Breakout.

3 EXPERIMENTAL DATA AND RESULTS

Training data: The model was trained for 3 hours on a single NVIDIA RTX 4090 GPU, using the Adam optimizer for 1,950,000 steps. The total reward during inference was 71.0.
As shown in Figure 2, the epsilon value decreases as training progresses, indicating that the agent explores more possible actions during the early stages of training but gradually focuses on exploiting learned strategies over time. Figure 3 illustrates the loss changes during training; initially, the loss fluctuates significantly but gradually decreases as the model converges. Figure 4 shows the change in the agent's reward during evaluation, where the reward increases as the model improves its performance. Figure 5 displays the reward changes during training, demonstrating a steady increase in the agent's performance over time.

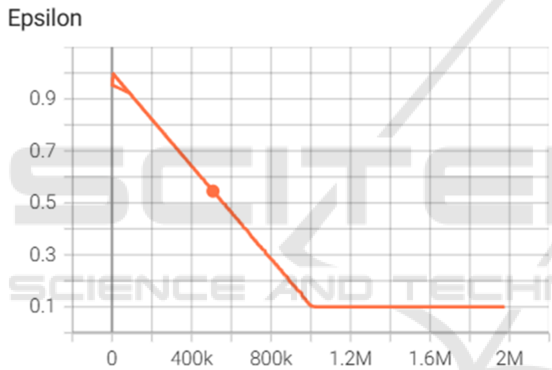

Figure 2: Epsilon Decay over Training Steps (Photo/ Picture credit: Original).

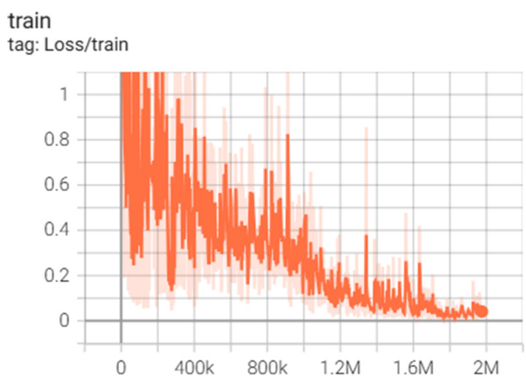

Figure 3: Training Loss vs. Steps (Photo/Picture credit : Original).

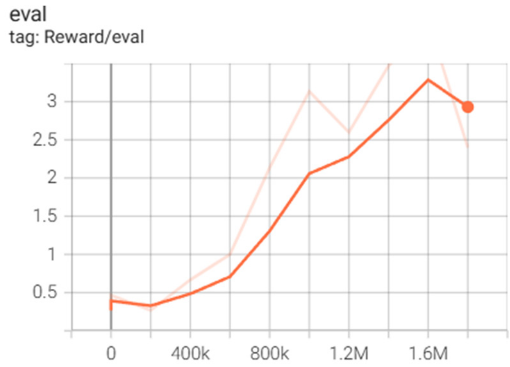

Figure 4: Evaluation Reward Progression (Photo/ Picture credit: Original).

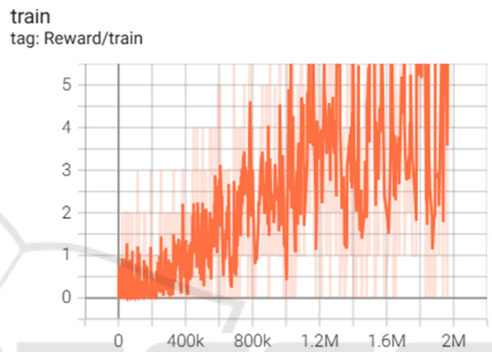

Figure 5: Training Reward Progression (Photo/Picture credit : Original).

4 CHALLENGES

Despite the promising potential of combining Rainbow DQN with Curriculum Learning, several challenges arise. One of the most significant challenges is the increased computational cost. Training the agent requires substantial processing power, especially given the complexity of the network and the need to fine-tune multiple hyperparameters. This makes the approach resource-intensive, which could be a barrier to its adoption in scenarios where computational resources are limited.

Another challenge is the sensitivity of the method to the design of the curriculum. If the progression of tasks is too steep or too gradual, it can either overwhelm the agent or slow down its learning. Finding the right balance requires careful experimentation and fine-tuning, which adds to the overall complexity of the approach (Bengio et al., 2009).

5 FUTURE WORK

Looking ahead, several exciting directions for future research are possible. One area of interest is the application of this combined method to other Atari games. By testing the approach across different games, the paper can better understand its generalizability and identify any game-specific adaptations that might be necessary. Another promising avenue is exploring the agent’s ability to learn multiple games simultaneously—a capability known as multi-task learning. If successful, this would signify a significant step forward in the development of more versatile AI agents that can apply their knowledge across different domains.

Furthermore, the integration of attention mechanisms into the agent's architecture presents a promising avenue for advancement. These mechanisms could enable the agent to selectively concentrate on the most salient aspects of the game environment, potentially enhancing its decision-making efficiency and overall performance (Smith et al., 2023). By prioritizing relevant information, attention-based models may offer a more nuanced approach to processing complex game states, leading to improved learning outcomes and adaptability.

6 CONCLUSION

This paper has demonstrated that integrating Rainbow DQN with Curriculum Learning can substantially enhance the performance of an AI agent in Atari Breakout. By addressing the limitations of standard DQN and employing a structured learning progression, the combined approach enables the agent to learn more effectively and achieve higher scores. The paper’s experimental results provide strong evidence of the benefits of this method, and the paper is optimistic about its potential applications to other games and learning scenarios.

In the future, the paper plans to extend this work by exploring multi-task learning and incorporating additional enhancements, such as attention mechanisms, to further improve the agent’s capabilities.

REFERENCES

Bengio, Y., Louradour, J., Collobert, R., & Weston, J. 2009. Curriculum learning. In *Proceedings of the 26th Annual International Conference on Machine Learning (ICML)*. https://doi.org/10.1145/1553374.1553380

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., ... & Hassabis, D. 2015. Human-level control through deep reinforcement learning. *Nature*, 518(7540), 529-533. https://doi.org/10.1038/nature14236

Hessel, M., Modayil, J., Van Hasselt, H., Schaul, T., Ostrovski, G., Dabney, W., ... & Silver, D. 2018. Rainbow: Combining improvements in deep reinforcement learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 32(1). https://doi.org/10.1609/aaai.v32i1.12111

Van Hasselt, H., Guez, A., & Silver, D. 2016. Deep reinforcement learning with double Q-learning. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence* (pp. 2094-2100). https://www.aaai.org/ocs/index.php/AAAI/AAAI16/paper/view/12389

Schaul, T., Quan, J., Antonoglou, I., & Silver, D. 2015. Prioritized experience replay. *arXiv preprint arXiv:1511.05952*. https://arxiv.org/abs/1511.05952

Fortunato, M., Azar, M. G., Piot, B., Menick, J., Osband, I., Graves, A., ... & Blundell, C. 2018. Noisy networks for exploration. In *International Conference on Learning Representations*. https://openreview.net/forum?id=rywHCPkAW

Bellemare, M. G., Dabney, W., & Munos, R. 2017. A distributional perspective on reinforcement learning. In *Proceedings of the 34th International Conference on Machine Learning - Volume 70* (pp. 449-458). https://doi.org/10.5555/3305381.3305437

Mnih, V., Badia, A. P., Mirza, M., Graves, A., Lillicrap, T., Harley, T., ... & Kavukcuoglu, K. 2016. Asynchronous methods for deep reinforcement learning. In *International Conference on Machine Learning* (pp. 1928-1937). https://doi.org/10.5555/3045390.3045594

Zhang, S., & Sutton, R. S. 2017. A deeper look at experience replay. In *Proceedings of the Thirty-Fourth International Conference on Machine Learning* (pp. 3197-3205). https://doi.org/10.5555/3305381.3305511

Vignon, C., Rabault, J., Vinuesa, R. 2023. Advances in Deep Reinforcement Learning for Complex Systems: Applications and Future Directions. *Physics of Fluids*, AIP Publishing. https://doi.org/10.1063/5.0020097

Zhou, W., Yao, X., Xiao, Y. W., et al. 2022. Atari Game Decision Algorithm Based on Hierarchical Reinforcement Learning. Information and Computer (Theoretical Edition), 34 (20): 97-99