Advancing AI Strategies in "I, Zombie"

Exploring RainbowDQN Implemented with MLP and LSTM for A Strategic Puzzle Game

Group Members: Daniel Chen, William Hu, Yunhan Mao, Mengfei Qi, Shuning Zhang

Introduction

Plants vs. Zombies (PvZ) is a popular strategy game where players strategically place plants with various defensive abilities to stop waves of zombies from reaching their house. While many programs have tried to train models to play the traditional PvZ, our project explores advanced machine learning models, specifically Deep Q-Network (DQN) integrated with Multi-Layer Perceptron (MLP) and Long Short-Term Memory (LSTM), to master "I, Zombie" mode where players spend suns to "plant" the zombies and destroy plants, trying to eat all 5 brains at the end of each row. This project aims to train an AI agent to excel in "I, Zombie" focusing on winning and maximizing the remaining suns. Before our project, there has been no learning models developed looking from the adversarial POV of the zombies. Hence, our contributions are threefold. We customized an environment for this game mode, create our own agents and compared them empirically. It is significant to notice that DQN used to play games where the reward is not delayed. However, our project tries to teach agents the importance of long-term strategic planning and decision-making in different ways, such as integrating DQN with Long Short-Term Memory (LSTM) models, so that the agents can realize that their actions have long-term consequences. Our project is important because the optimal strategy learned through the model can be beneficial materials to understand in terms of game theory for future game development and adversary defense strategies in general.

Figure 1: Example of the Start State of "I, Zombie"

Methodology

Environment Customization

We modified an open-source "Plants vs. Zombies" simulator for the "I, Zombie" mini-game to directly obtain state data post-action, eliminating the need for visual state analysis. This streamlined data acquisition was crucial for efficient agent training.

Initial Model: DQN with MLP

Our primary model was a Deep Q-Network (DQN) with a Multilayer Perceptron (MLP), selected for its suitability in environments with defined goals. DQN's strength lies in its ability to make decisions based on the current state. However, its tendency to overestimate Q-values, a known limitation, was a significant consideration for our model's effectiveness (Van Hasselt et al., 2016).

Advancing to Double DQN (DDQN)

To address the overestimation issue in DQN, we incorporated Double DQN (DDQN). DDQN uses a target network alongside the current network, enhancing the accuracy of Q-value estimations and thereby improving the decision-making process (Van Hasselt et al., 2016).

$$Q_{update}(s,a) = r + \gamma \cdot Q_{target}(s', \underset{a'}{argmax}\, Q_{current}(s',a'))$$

Transition to Rainbow DQN

The initial models faced challenges with randomness in starting states and long-term strategic learning. To overcome these, we adopted Rainbow DQN, which integrates DDQN, Prioritized Replay, Dueling Networks, Multi-step learning, and Noisy Nets. This approach helped in handling randomness and learning the long-term consequences of actions more effectively (Hessel et al., 2018).

LSTM Integration

Finally, we experimented with replacing the MLP in our model with LSTM networks to enhance the agent's capability in understanding the long-term implications of its actions, particularly valuable in complex, sequential decision-making scenarios.

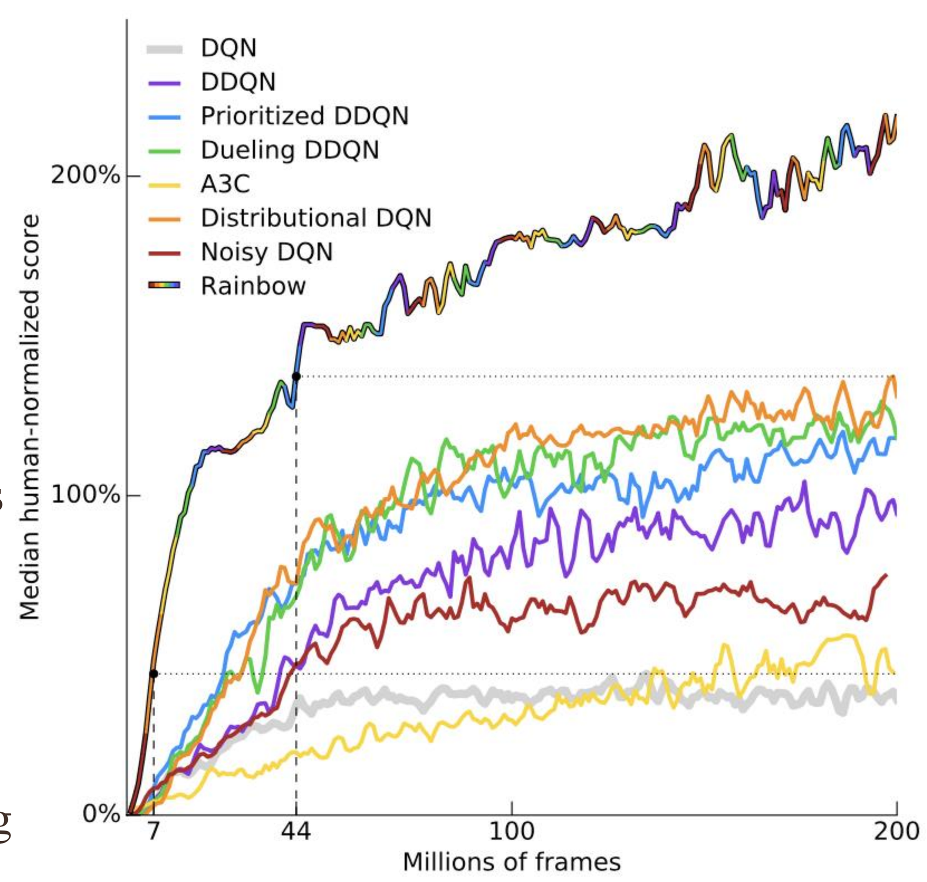

Figure 2: Performance of Rainbow DQN (Hessel et al., 2018).

Result

In our exploration of AI techniques for "I, Zombie" in *Plants vs. Zombies*, we focused on the performance metrics of winning rate and average remaining suns to evaluate the effectiveness of different strategies. The obtained results are visually represented in the graphs below.

The DDQN with a MLP architecture exhibited lower winning rates and average remaining suns throughout the training process, as indicated by the gray line in the graphs. These findings suggest that while DQN and DDQN could navigate the game environment to a certain extent, their ability to optimize gameplay in a manner that maximizes the long-term rewards was limited.

However, the implementation of Rainbow DQN marked a significant improvement in both metrics. The red line in the graphs illustrates a higher level of average remaining suns and a much higher winning rate. This substantial enhancement in performance underlines the capability of Rainbow DQN to effectively manage the game's strategic complexity, balancing immediate actions with their long-term outcomes more proficiently.

Adding to this, the DDQN+LSTM model (depicted with the blue line) achieved a better winning rate than DDQN, but it also showed a lower level of average remaining suns than DDQN. On the other hand, the integration of LSTM with Rainbow DQN, intended to further enhance decision-making through an understanding of temporal sequences, maintained the highest average of remaining suns but paradoxically resulted in the lowest winning rates among the tested models, as seen in the green line of the winning rate graph. These unexpected outcomes suggest that the reward function could potentially be further optimized for the LSTM's pattern recognition capabilities. In consideration of our human baseline for remaining suns that has a mean of 1104.76 and a standard deviation of 107.77—which suggests the human baseline is a stable dataset—the integration of LSTM with Rainbow DQN has surpassed the mean baseline 100% of the time, depicting it having learned the most reliable and effective strategies in comparison to all the other models and average human performances.

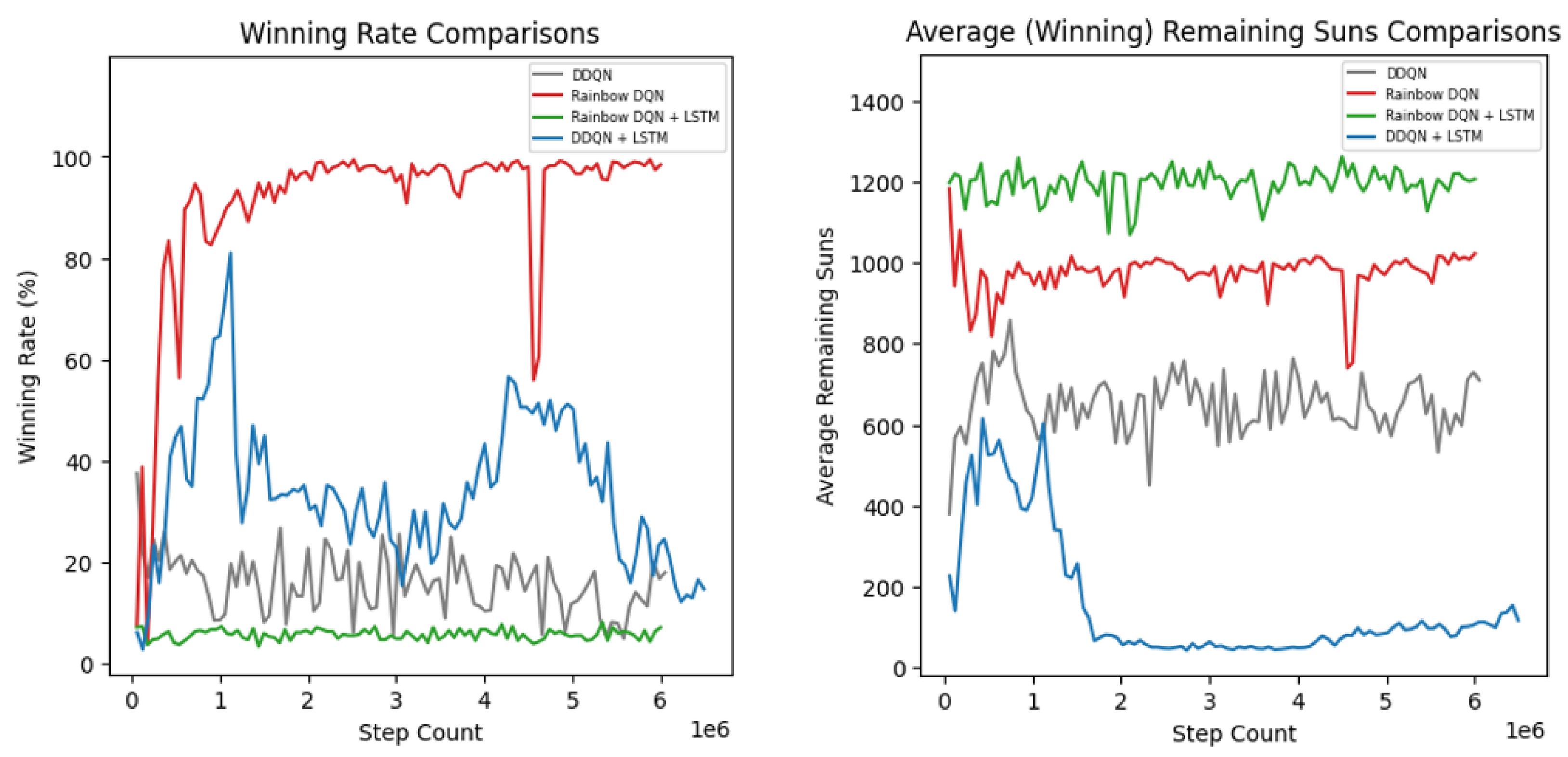

Figure 3: Winning Rates and Remaining Suns Comparison

Conclusion and Discussion

This research demonstrates the differential impact of neural network architectures on AI performance in the "I, Zombie" minigame. The Rainbow DQN with MLP achieved the highest winning rates, highlighting the strength of advanced DQN modifications for decision-making tasks. Conversely, the Rainbow DQN + LSTM, while showing lower victory rates, maintained a high level of remaining suns, pointing to its potential for creating efficient resource management strategies.

The significance of these findings lies in their illustration of the complex trade-offs between achieving immediate objectives and maintaining resource efficiency. The results guide us in choosing the right AI model based on the desired outcome, whether it's maximising success or conserving resources.

Future work will focus on refining the understanding of when and how LSTM can be leveraged to improve performance, possibly by adjusting the model or the reward structure of the learning algorithm. Additionally, investigating hybrid models that can combine the strategic strengths of Rainbow DQN with the resource efficiency of LSTM could yield a model with both high winning rates and suns remaining. Further research could also explore the transferability of these findings to other strategic games or real-world applications, such as resource allocation and logistics, to validate the broader applicability of these AI training methods.

References

1. Van Hasselt, H., Guez, A., & Silver, D. (2016). Deep Reinforcement Learning with Double Q-learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 30(1).
2. 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).