Formatting Instructions For NeurIPS 2025

Anonymous Author(s)
Affiliation
Address
`email`

Abstract

The abstract paragraph should be indented ½ inch (3 picas) on both the left- and right-hand margins. Use 10 point type, with a vertical spacing (leading) of 11 points. The word **Abstract** must be centered, bold, and in point size 12. Two line spaces precede the abstract. The abstract must be limited to one paragraph.

1 Introduction

Recent research on game strategy agents has flourished in response to the growing demand for intelligent systems capable of playing strategic games either alongside or against human players. Reinforcement learning (RL) has established itself as a powerful paradigm for training such agents, enabling them to acquire optimal behaviors through interaction with an environment to maximize cumulative rewards over time. A variety of algorithmic advancements have been proposed for a comprehensive generalist mastering diverse games within a unified framework. **Deep Q-Network (DQN)** [1] pioneered the use of deep neural networks to approximate value functions, enabling end-to-end learning from raw pixel inputs. **Proximal Policy Optimization (PPO)** [3] introduced a more stable and sample-efficient policy gradient method, supporting large-scale training through self-play and mini-batch updates from stored trajectories.

Building on these foundations, **AlphaZero** [4] demonstrated the power of combining deep learning with tree-based planning in a model-based fashion, but it relied on known environment dynamicsmost notably, the rules of the game. **MuZero** [2] advanced this line of work by learning not only the policy and value functions but also a latent, implicit model of the environments dynamics. This abstraction allowed **MuZero** to extend the planning-based advantages of **AlphaZero** to previously unknown or complex domains, where explicit modeling is impractical. MuZero represents a paradigm shift in model-based RL by learning not only the value and policy functions, but also the dynamics of the environment without access to the actual game rules. Its success has been demonstrated across a diverse range of domains, including board games like Go, Chess, and Shogi, as well as in more complex and dynamic environments such as Atari. MuZeros ability to integrate planning with learned models and to perform well in both deterministic and stochastic environments positions it as one of the most general and powerful agents to date.

As part of our investigation into MuZero's adaptability to varied strategic domains, we selected *Connect6*, a relatively underexplored game known for its balanced mechanics and heightened complexity relative to classic games like *Gomoku*. *Connect6* addresses the first-player advantage problem inherent in many connection games by introducing the continuous dual-step rule. This subtle change significantly increases the strategic depth and makes the game an ideal candidate for advanced AI research. Though prior works leverage **Monte Carlo Tree Search** [6] or **AlphaZero** [5], most research relies on hand-crafted heuristics and incorporates explicit rule features to structure the training framework. In this context, we conduct the experiments with respect to application of **MuZero** to *Connect6*, aiming to evaluate its performance and adaptability in a novel and partially observable environment. This research sheds new light on **MuZero**s scalability and progressively helps advance in game-playing intelligent agents.

Submitted to 39th Conference on Neural Information Processing Systems (NeurIPS 2025). Do not distribute.

References

[1] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin Riedmiller. Playing atari with deep reinforcement learning, 2013.

[2] Julian Schrittwieser, Ioannis Antonoglou, Thomas Hubert, Karen Simonyan, Laurent Sifre, Simon Schmitt, Arthur Guez, Edward Lockhart, Demis Hassabis, Thore Graepel, Timothy Lillicrap, and David Silver. Mastering atari, go, chess and shogi by planning with a learned model. *Nature*, 588(7839):604–609, Dec 2020.

[3] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms, 2017.

[4] David Silver, Thomas Hubert, Julian Schrittwieser, Ioannis Antonoglou, Matthew Lai, Arthur Guez, Marc Lanctot, Laurent Sifre, Dharshan Kumaran, Thore Graepel, Timothy Lillicrap, Karen Simonyan, and Demis Hassabis. Mastering chess and shogi by self-play with a general reinforcement learning algorithm, 2017.

[5] Tzu-Yee Yang. *An AlphaZero-Based Design for Connect6*. PhD thesis, Taiwan R.O.C., 2020. - Database copyright ProQuest LLC; ProQuest does not claim copyright in the individual underlying works.

[6] Shi-Jim Yen and Jung-Kuei Yang. Two-stage monte carlo tree search for connect6. *IEEE Transactions on Computational Intelligence and AI in Games*, 3(2):100–118, 2011.

A Technical Appendices and Supplementary Material

Technical appendices with additional results, figures, graphs and proofs may be submitted with the paper submission before the full submission deadline (see above), or as a separate PDF in the ZIP file below before the supplementary material deadline. There is no page limit for the technical appendices.