COMPARATIVE STUDY OF WORLD MODELS, NVAE-BASED HIERARCHICAL MODELS, AND NOISYNET-AUGMENTED MODELS IN CARRACING-V2

Vidyavarshini Holenarasipur Jayashankar, Banafsheh Rekabdar
Department of Computer Science, Portland State University
{vidyav2, rekabdar}@pdx.edu

ABSTRACT

In the case of OpenAI's CarRacing-V2, Reinforcement Learning (RL) needs to solve both the problem of world modeling and exploration. This work primarily focuses at solving the issues of efficient world modeling and exploration strategies in RL for continuous control tasks by comparing different approaches for improving the performance. It exhibits an experimental evaluation of three approaches: (i) standard World Models, (ii) NVAE-based hierarchical World Models, and (iii) NoisyNet-augmented World Models. We compare these methods based on cumulative reward performance, training stability, and computational efficiency. The comparison of the cumulative rewards and training stability in the experiments showed that the NVAE-based models improve the feature representation and the generalization of the models while the NoisyNet augmentation improves the adaptive exploration. The work also shows trade-offs, for instance, the computational cost versus the reward performance among these approaches. It also proposes that a future model-based RL for autonomous driving should incorporate NVAE for feature extraction and NoisyNet for exploration as they could yield the best results. The results show that standard World Models have the highest cumulative reward, whereas the NoisyNet-augmented models have similar performance with fewer rollouts, thus indicating better exploration efficiency.

1 INTRODUCTION

The RL agents have recently shown great improvement in control-based tasks including autonomous driving, robotics and other fields of studies including (Mnih et al., 2015; Lillicrap et al., 2015). Exploration-exploitation trade-offs and data efficiency are quite important challenges to the present day. World Models (Ha & Schmidhuber, 2018) address these issues by the development of compact representations of the environment, reducing sample complexity and supporting more stable policy learning.

Hierarchical representations further improve model performance by introducing structure in the learned latent space. NVAE-based World Models (Vahdat & Kautz, 2020; Ayyalasomayajula et al., 2023) employ deep hierarchical latent variable modeling to capture complex dependencies in high-dimensional sensory inputs. In contrast, NoisyNet (Fortunato et al., 2018) augments policy networks with trainable noise to improve exploration's robustness.

This paper exhibits experimental comparison of these three approaches to evaluate their strengths, weaknesses, and applicability to continuous control tasks. Specifically, we aim to answer the following questions:

- How do NVAE-based hierarchical models compare to standard World Models in terms of cumulative reward and training stability?
- Does NoisyNet augmentation improve exploration efficiency and performance in model-based RL?
- What are the computational trade-offs between these approaches?

2 Background and Related Work

2.1 World Models

The World Models, as introduced by Ha & Schmidhuber (2018), uses a Variational Autoencoder (VAE) for efficient feature extraction, alongside a Mixture Density Recurrent Neural Network (MDRNN) to accurately model environmental dynamics. To optimize the control mechanism, they employ evolutionary strategies. These models offer a compact and efficient means to represent complex environments in decision-making tasks relevant to fields like robotics and gaming. By simulating the dynamics of the real world within a computational framework, World Models facilitate predictive and strategic planning. This approach reduces computational demands and enhances algorithm performance by abstracting and focusing on the essential features of environments.

2.2 NVAE-Based Hierarchical World Models

NVAE-Based Hierarchical World Models, leveraging the advancements in NVAE as described by Vahdat & Kautz (2020), extend traditional VAE architectures by using hierarchical latent variables. This design enhances the capability of VAEs to capture complex dependencies within high-dimensional inputs, greatly improving feature retention and generalization capabilities across unseen scenarios, as supported by foundational works (Kingma & Welling, 2014; Rezende et al., 2014). By integrating these hierarchical structures into world models, the approach efficiently models multi-scale environmental dynamics. This hierarchical organization allows the model to operate at varying levels of abstraction—from broad, general patterns to specific, detailed phenomena—making it suitable for complex tasks like predicting future states in dynamic systems.

2.3 NoisyNet-Augmented World Models

Noisynet-Augmented World Models use the ideas outlined in Fortunato et al. (2018), which introduce parameter noise into policy networks to foster efficient exploration. This innovation minimizes the dependence on conventional heuristic exploration strategies such as epsilon-greedy, enhancing the stability and robustness of RL outcomes (Bellemare et al., 2016). By integrating noisy nets into world models, this approach amplifies exploration capabilities within learning algorithms. The deliberate addition of noise to neural network parameters broadens the exploration of potential actions and strategies, essential in complex and uncertain environments. This strategic variability not only helps avoid local optima but also results in deeper understanding of environmental dynamics, thereby improving decision-making efficacy.

3 Proposed Methodology

Our study comprises of a tendered pipeline that includes the training and fine tuning of different model configurations and evaluation of these models in a controlled environment. The methodology consists of several important phases that play a crucial role in comparing the performance of the models in question.

Pre-trained Controller Initialization: The first stage of the process includes the initialization of a pre-trained controller which has been obtained from the CarRacing-v0 environment. This stage includes the process of de-serializing the JSON configuration file to precisely set and adjust policy network weights. The controller is a fully connected neural network that has been trained with the help of evolutionary strategies to optimize its decision-making abilities based on previous performance data.

Rollout Generation: We then start the CarRacing-V2 environment with a 96x96 pixel observation space. We use the pre-trained controller to perform multiple runs of 10,000 rollouts in order to generate adequate action-state sequences. Each rollout produces a sequence of images, actions, and the corresponding rewards at regular time step intervals which are used to create a rich data source for analysis.

Data Collection and Preprocessing: The obtained rollouts are utilized to gather sequences of states, actions and rewards which are then saved for training purposes. The dataset is then prepared for model training through a preprocessing step where the raw images are transformed into a suitable format using a CNN-based encoder for the latent state representations. The dataset is then divided into 80% training set and 20% validation set for the model training phase.

VAE Retraining with Noisy Layer: In the next phase, the VAE is retrained to enhance its latent representation capabilities. More rollout data is employed to perform fine tuning of the VAE in order to achieve a more robust encoding of the latent space.

MDRNN Retraining: We retrain the Mixture Density Recurrent Neural Network (MDRNN) with LSTM-based architectures to enhance the sequence prediction accuracy of the MDRNN. The retraining is conducted with the help of both KL-divergence loss and mean squared error metrics in order to maintain the stability of transition predictions. Also, we applied weight perturbations based on NoisyNet to enhance the system's uncertainty modeling.

Controller Optimization with CMA-ES: The Covariance Matrix Adaptation Evolution Strategy (CMA-ES) is used to optimize the controller. This strategy entails that perturbations are sampled. It chooses the configurations that produce the best results, with a special emphasis on maximizing cumulative reward metrics in simulated environments.

Model Deployment and Testing: As a final step, the optimized controller is deployed to evaluate its performance on new rollouts. The testing phase compares NoisyNet-augmented World Models against baseline approaches and performs ablation studies to determine the effects of particular components within the experimental pipeline.

4 Results and Discussion

To evaluate model performance, we performed our experiments under different conditions including, altering the number of rollouts, weight-sigma values, and input-output feature mappings. Key performance indicators included cumulative rewards, training stability, and computational efficiency.

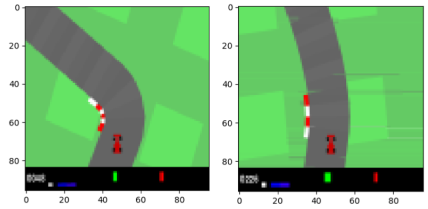

Figure 1: **NoisyNet-Augmented Trajectory Following.** CarRacing-V2 agent following the track. The agent demonstrates improved exploration and stability in trajectory following after NoisyNet augmentation, enabling more efficient navigation of the track.

Table 1: Performance Comparison of Different Approaches

Model	**Rollouts**	**Cumulative Reward**	**Training Time (hrs)**
World Models	10,000	906 ± 21	25
NVAE-Based Models	10,000	800 ± 18	22
NoisyNet-Augmented	5,000	801 ± 156	21

Table 2: Comparison of NoisyNet-Augmented Models across Rollouts

Rollouts	**Weight Sigma**	**Cumulative Reward**	**Time (hrs)**
2,000	0.017	512 ± 117	7
3,000	0.050	631 ± 247	11.9
4,500	0.050	709 ± 194	14.27
5,000	0.050	801 ± 156	20.75

The comparative analysis given above indicates that while standard World Models achieve the highest cumulative reward, they significantly require a higher count of rollouts. The NVAE-based models generalize well but need more structured latent spaces. The NoisyNet-augmented models accelerate exploration efficiency, achieving near-optimal results with fewer rollouts.

The results presented in Table 1 indicate that the standard World Models achieved the highest cumulative reward, albeit with a longer training time. The NVAE-based models, while slightly less performative, showed more stable training with lower variance in rewards. The NoisyNet-augmented models, despite being trained with fewer rollouts, demonstrated competitive performance, suggesting that the introduction of noise layers can lead to more efficient exploration and potentially faster convergence.

5 Conclusion

In this study, we explored the strengths and limitations of World Models (Ha & Schmidhuber, 2018), NVAE-based hierarchical models (Vahdat & Kautz, 2020), and NoisyNet-augmented approaches (Fortunato et al., 2018) in the CarRacing-V2 environment. Our findings highlights some of the key trade-offs in reward performance, training stability, and computational cost among these architectures. While World Models excel in achieving high rewards, which demands a large number of rollouts to be generated, making them computationally expensive (Ha & Schmidhuber, 2018). The NVAE-based models improves generalization through hierarchical feature representations, but their increased complexity can make training and inference more challenging (Vahdat & Kautz, 2020; Kingma & Welling, 2014). On the other hand, the NoisyNet-augmented models introduced controlled randomness to the network parameters, improving exploration efficiency while reducing sample requirements (Fortunato et al., 2018; Bellemare et al., 2016).

Our results suggest that a hybrid approach—combining the structured feature learning of hierarchical models with adaptive exploration techniques like NoisyNet—could lead to more efficient and scalable RL architectures (Hafner et al., 2023). Moving forward, future research could focus on developing hybrid frameworks (Oord et al., 2018) and testing these models in real-world applications, such as robotics, healthcare and autonomous driving, to assess their adaptability beyond simulated environments (Bojarski et al., 2016; Codevilla et al., 2019). By bridging the gap between theoretical advancements and practical deployment, RL models can become more robust, efficient, and capable of handling real-world complexities.

6 Future Work and Scope

This study opens new directions for enhancing world models in RL. One key extension is adaptive latent exploration, where NoisyNet is integrated into hierarchical latent spaces to dynamically regulate exploration based on uncertainty levels. Unlike existing NoisyNet applications, which focus on policy networks, applying it within structured latent representations could improve sample efficiency in complex, partially observable environments (Fortunato et al., 2018; Vahdat & Kautz, 2020).

Another promising direction is multi-scale representation learning, where NVAE's hierarchical encoding is further refined using self-supervised contrastive learning. This approach could help the model distinguish between fine-grained temporal dependencies and high-level scene abstractions, leading to more stable long-term planning (Oord et al., 2018; Hafner et al., 2023).

Finally, real-world validation remains as a critical step. While most benchmarks rely on simulation, (Bojarski et al., 2016; Codevilla et al., 2019). By bridging the structured world modeling with adaptive exploration, this work aims to push RL towards more generalizable and sample-efficient decision-making systems.

Acknowledgments

In this paper, we build upon the concepts presented in all the research works cited throughout the comparative study. The research conducted by these authors has been instrumental in guiding our experiments.

We would like to extend our special thanks to Arulkumaran et al. (2017) for surveying deep RL, Sutton & Barto (2018) for reinforcing learning principles in detail and Finn et al. (2017) for model-agnostic meta-learning insights and Parisi et al. (2019) for detailing continual lifelong learning techniques.

REFERENCES

Kai Arulkumaran, Marc Peter Deisenroth, Miles Brundage, and Anil Anthony Bharath. Deep reinforcement learning: A brief survey. *IEEE Signal Processing Magazine*, 2017.

Sriharshitha Ayyalasomayajula, Banafsheh Rekabdar, and Christos Mousas. Deep hierarchical variational autoencoders for world models in reinforcement learning. In *TransAI*, pp. 128–134, 2023.

Marc G. Bellemare, Will Dabney, and Rémi Munos. Unifying count-based exploration and intrinsic motivation. *arXiv preprint arXiv:1606.01868*, 2016.

Mariusz Bojarski, Davide Testa, Daniel Dworakowski, Bernhard Firner, Beat Flepp, Prasoon Goyal, Larry D. Jackel, Mathew Monfort, Urs Muller, Jiakai Zhang, Xin Zhang, Jake Zhao, and Karol Zieba. End to end learning for self-driving cars. *arXiv preprint arXiv:1604.07316*, 2016.

Felipe Codevilla, Matheus Miiller, Alexey Dosovitskiy, Antonio López, and Vladlen Koltun. Exploring the limitations of behavior cloning for autonomous driving. *Proceedings of the IEEE International Conference on Computer Vision*, 2019.

Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. *Proceedings of the 34th International Conference on Machine Learning*, 2017.

Meire Fortunato, Mohammad Gheshlaghi Azar, Bilal Piot, Jacob Menick, Ian Osband, Alex Graves, Vlad Mnih, Rémi Munos, Demis Hassabis, et al. Noisy networks for exploration. *arXiv preprint arXiv:1706.10295*, 2018.

David Ha and Jürgen Schmidhuber. World models. *arXiv preprint arXiv:1803.10122*, 2018.

Danijar Hafner, Timothy Lillicrap, Jimmy Ba, and Mohammad Norouzi. Mastering diverse domains through world models. *arXiv preprint arXiv:2301.04104*, 2023.

Diederik P. Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2014.

Timothy P. Lillicrap, Jonathan J. Hunt, Alexander Pritzel, Nicolas Heess, Tom Erez, Yuval Tassa, David Silver, and Daan Wierstra. Continuous control with deep reinforcement learning. *arXiv preprint arXiv:1509.02971*, 2015.

Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A. Rusu, Joel Veness, Marc G. Bellemare, Alex Graves, Martin Riedmiller, Andreas K. Fidjeland, Georg Ostrovski, Stig Petersen, Charles Beattie, Amir Sadik, Ioannis Antonoglou, Helen King, Dharshan Kumaran, Daan Wierstra, Shane Legg, and Demis Hassabis. Human-level control through deep reinforcement learning. *Nature*, 518:529–533, 2015.

Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*, 2018.

German I. Parisi, Ronald Kemker, Jose L. Part, Christopher Kanan, and Stefan Wermter. Continual lifelong learning with neural networks: A review. *Neural Networks*, 2019.

Danilo Jimenez Rezende, Shakir Mohamed, and Daan Wierstra. Stochastic backpropagation and approximate inference in deep generative models. *arXiv preprint arXiv:1401.4082*, 2014.

Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*. MIT Press, 2 edition, 2018.

Arash Vahdat and Jan Kautz. Nvae: A deep hierarchical variational autoencoder. *arXiv preprint arXiv:2007.03898*, 2020.