Journal of Physics:
Conference Series

PAPER • **OPEN ACCESS**

Optimizing ride comfort through deep reinforcement learning for autonomous vehicle control

To cite this article: N A Abdellah *et al* 2025 *J. Phys.: Conf. Ser.* **3058** 012003

Optimizing ride comfort through deep reinforcement learning for autonomous vehicle control

N A Abdellah[1*], T S Ghoniemy[2], and M M Hegaze[1]

[1]Mechatronics Engineering Department, Arab Academy for Science, Technology & Maritime Transport, Smart Village, Egypt
[2]Computer & Artificial Intelligence Deprartment, Military Technical Collge, Cairo, Egypt

Email: N.Abdella33547@student.aast.edu

Abstract. Autonomous vehicles have been a research trend over the past two decades, and both industrial and academic institutions have exerted considerable effort to achieve the highest level of autonomy. All these efforts have resulted in the use of deep reinforcement learning for autonomous driving and autonomous vehicles, as it provides great flexibility in achieving autonomy, especially in end-to-end control. One challenge when using deep reinforcement learning for end-to-end control of autonomous vehicles is the reward function design, which is the basis on which the behaviour of the vehicle is designed. Many efforts have been made by researchers and engineers to achieve an ideal reward function design, but to the best of our knowledge, this has still not been achieved. Reward function design has many challenges, one of which is the missing attributes that pose a great deal for end users, such as comfort. This study presents a novel reward function specifically designed to enhance ride comfort in autonomous vehicles. The proposed design process surpasses other methods by prioritizing passenger comfort as a core objective. The efficacy of the proposed reward function is demonstrated by the increased total accumulated rewards per episode and the acceleration profiles proved by a 44.34% reduction compared to the baseline model.

1. Introduction

Research has been dedicated to autonomous systems over the past decade owing to the evolution of science and technology. An autonomous system is an intelligent system that incorporates vision and planning to achieve its goals. Decision-making for autonomous systems has become an increasingly interesting field of research, and in this study, planning and decision-making are used interchangeably. According to Dimensions.ai [1], the number of publications on the topic up to this year increased by four times the number of publications in 2013. Thus, the planning and control of autonomous vehicles have been thoroughly researched. Two main approaches are studied: sequential decision making and end-to-end decision making. Because end-to-end decision-making is favored in literature, several methodologies have been proposed, one of which is Deep Reinforcement Learning (DRL). Recently, DRL is used in end-to-end learning because it offers greater flexibility and autonomy. However, challenges arise when utilizing DRL for autonomous driving, such as reward function design, which can be regarded as a feedback signal for the system. Thus, numerous aspects must be considered. In the literature, there are many challenges in reward function design, such as incomplete problem description, missing attributes, redundant attributes, and inefficient reward shaping [2]. The remainder of this paper explains the use of DRL for autonomous vehicle control. A novel reward function design with a complete problem description that addresses missing attributes such as ride comfort is proposed. The remainder of this paper is organized as follows: Section 2 presents the related research works, Section 3 discusses the methodology, Section 4 shows the results' analysis and Section 5 concludes the presented research.

2. Related Works

This section discusses the literature background, starting with autonomous vehicles and ride comfort, then DRL for autonomous vehicles as per recent research on decision making for autonomous systems, and challenges in DRL for autonomous vehicles.

2.1. Autonomous Vehicles and Ride Comfort

An autonomous vehicle (AV) is an intelligent system that can perceive its environment and observe, decide, and take actions based on feedback from the environment. There are two main software architectures for AVs: sequential and end-to-end (E2E). The authors of [3] stated the present approaches for planning and decision-making for autonomous vehicles and classified these approaches as sequential planning, behavior-aware planning, and end-to-end planning, behavior-aware planning can also be considered a form end-to-end planning, the authors claimed that integrating the perception and planning systems of an autonomous vehicle achieves better performance and accuracy. In [4], the authors stated that there are several methodologies for sequential planning but only learning-based approaches are suitable for end-to-end planning to introduce adaptability. Moreover, [5] argued that integrating perception data with planning and control to create an end-to-end planning architecture offers more simplicity in the system design and adds flexibility and adaptability to the AV. Sensor data collection, perception, planning, and control were implemented based on the software architecture of the AV. Autonomous systems have four main modules: sensing, perception, planning, decision-making, and control. As shown in Figure 1, the AV collects sensor data and observations from the environment, then processes this information and forms an understanding of the environment and then makes decisions based on its understanding of the environment and sends signals to the control module to act. Figure 2 shows the differences between sequential and end-to-end planning.

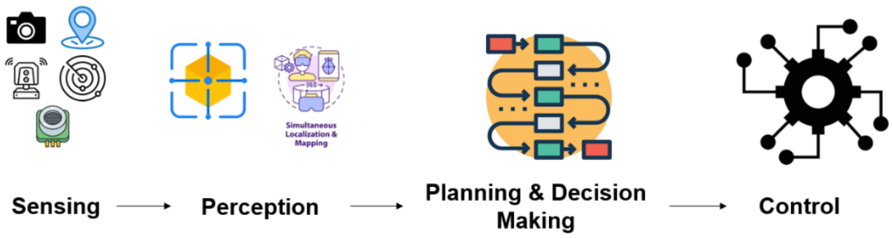

Figure 1. Autonomous Vehicles Flow

Upon surveying the literature on end-to-end methods, it was found that learning-based methods were the most common in realizing this approach; deep learning, imitation learning, and Reinforcement Learning (RL) are the most common. Each method has its advantages and disadvantages however this paper tackles the DRL approaches. [6] proposed an RL model with a generative adversarial network (GAN) to convert non-realistic virtual input images to a realistic one with a similar scene structure to bridge the gap between simulation and reality because an RL model cannot be trained in real life. Furthermore, [7] proposed a novel framework of RL with an image semantic segmentation network to narrow the gap between simulation and reality. In [8], new reward and learning strategies were presented and together they resulted in faster convergence using only RGB image from a front facing camera. The authors in [9] used the A3C framework proposed to learn the car control in a stochastic rally game. Finally, the authors in [10] presented an end-to-end model with a hybrid of imitation-reinforcement learning architecture.

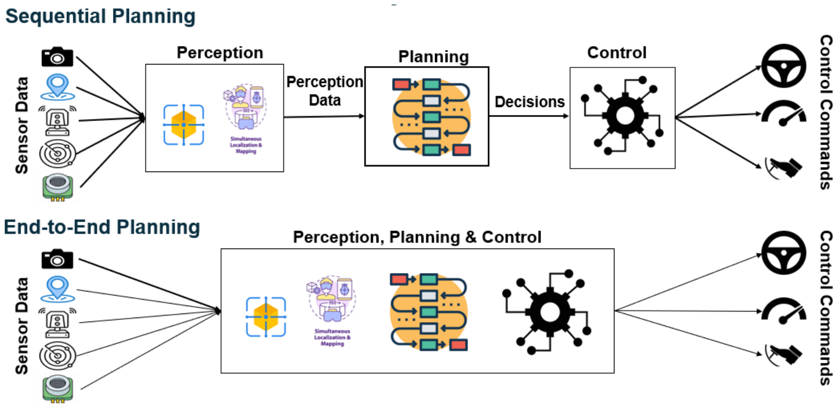

Figure 2. Sequential vs. E2E Planning Approaches

Despite the evolution of technical methodologies in the automotive domain, such as ADAS and other technologies, passenger experience has not been considered thoroughly, which has immense potential in introducing AVs to the market. Research on AVs has been increasing, but still with no regard to passenger ride comfort. Ride comfort can describe a several aspects of a ride, and according to [11], there are distinct factors affecting passenger experience and ride comfort: controllability factors, robotic control factors and environmental factors. Controllability factors include motion sickness, naturality of paths, and appearance of surfaces, robotic control factors include apparent safety, disturbances, and vibrations while environmental factors include air quality, temperature, sound and noise, road condition and light condition. Furthermore, [12] illustrates the factors affecting human comfort in AVs and categorizes the factors as the factors affecting ride comfort in conventional vehicles such as air quality, sound and noise, temperature, and vibrations, and factors affecting AVs such as naturality, disturbances, apparent safety, and motion sickness. Disturbances are not specified to a certain source but are thought to be caused by jerks, rough maneuvers, high acceleration or deceleration, or high steering angles. Multiple studies have discussed motion sickness factor such as [11,12], others discussed the design of the suspension system to optimize ride comfort such as [13]. Another study discussed the motion optimization for comfort that utilized DRL for suspension control at planned speeds[14]. In addition, [15] proposed an RL based vibrations control for the semi-active suspension system.

All the above-mentioned studies pointed out the mechanical vibration level as a factor affecting ride comfort, but little research has been conducted on this topic. Hence, the study of ride comfort is a new field of research. To study the ride comfort of a passenger, one must first know how to evaluate the comfort level and then start working with the data. According to [16] there are three types of assessments of ride comfort: subjective assessment, objective assessment, and mathematical models. Subjective assessment relies on the verbal feedback of passengers during a ride, and objective assessment uses sensors or devices to quantify comfort, while mathematical models use physiological and biomechanical parameters to predict the level of comfort based on the data. The need for comfort level evaluation to be simulation-based or objective was stated by [17], in which the number of physical prototypes is reduced significantly if virtualization is achieved for key design parameters and full vehicle configurations, even in the early stages of development which is originally stated in [18], and ride comfort is for sure one of the key design parameters as is it a foundational part of the passenger experience in AVs. Most ride comfort studies are based on public transportation [19], the authors classified sixteen papers, seven of which were AVs. Thus, the paper's main interest was only in AVs, and only one of these seven utilized machine learning. All the methods introduced in these papers only predict the vibration level, but do not prevent or reduce it. Studies were conducted to find the factors affecting ride discomfort, one of which is [20], the authors concluded their study with 3 key observations: Discomfort is directly proportional to the acceleration

Journal of Physics: Conference Series **3058** (2025) 012003 doi:10.1088/1742-6596/3058/1/012003

magnitude which makes it the most important factor in predicting discomfort, The effect of jerks seems to be negative which means that higher jerks are associated with lower levels of discomfort and direction affects discomfort, which means that forward motion comes first in the levels of comfort, backward motion comes second and lastly lateral motion which is the most uncomfortable. Authors in [16] introduced a novel model that represents a vehicle passenger by including all body parts which were not given enough attention in previous research. In their design of the model, they stated that vertical acceleration is the most important factor in ride discomfort, as it directly affects the human spine.

2.2. Deep Reinforcement Learning for Autonomous Vehicles

2.2.1. Theoretical Foundations

Over the last decade, RL has become a key technology for improving the autonomy of autonomous systems, including robots and autonomous vehicles. RL is learning by experience in which an agent learns by interacting with an environment and receiving feedback in the form of rewards, so it surpasses the performance of supervised learning in autonomous driving or robotics because supervised learning does not learn the dynamics of the environment or the agent [21] [22]. To better understand RL, the key components of an RL model are first explained.

1. **Agent:** also known in the engineering community as the controller, an agent in RL is the decision maker, and it is the software embedded in the system.
2. **Environment:** also known in the engineering community as the controlled system, is everything that surrounds the agent and everything that the agent interacts with.
3. **Reward Function,** also known as the control signal in the engineering community, can be considered as the feedback signal that the agent receives after taking a certain action to evaluate its performance or the goal of the reinforcement agent.
4. **Policy:** This is the behaviour that the agent is trying to learn to achieve optimal performance or the mapping from states to actions.
5. **Value:** accumulated rewards over time.

Deep Learning (DL) is utilized for end-to-end learning, planning tasks, and perception. Given the inputs and outputs, DL learns to predict the mapping between both and then generalizes to unseen data. In autonomous driving, one of the DL limitations is that it requires labeled data that may not always be available. This limitation can be overcome by DRL, which is discussed in the next section.

2.2.2. Deep Reinforcement Learning

DRL is a combination of DL and RL, using a neural network as a approximation function for the RL model, which then creates the working principle of DRL, as shown in Figure 3 below:

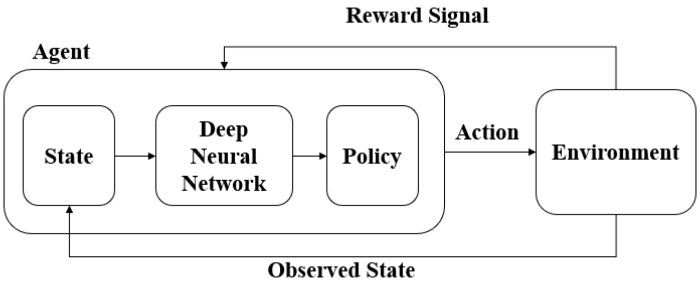

Figure 3. Deep Reinforcement Learning Cycle

DRL algorithms are classified into value-based, policy-based and actor-critics [22], value-based works by estimating the values of each state or of each action without having a defined policy, while policy-based works by learning the policy directly as it learns state-action pairs. Actor critics combine both value-based and policy-based methods in which the agent (the actor or policy-based) learns a policy and the critic (value-based) evaluates it by giving back feedback to the actor in the form of values. This is the most complex type of DRL algorithm and is well suited for complex environments and complex tasks. There are variations of this algorithm, some of the most recent actor-critical algorithms are Deep Deterministic Policy Gradient (DDPG) [23], Policy Proximal Optimization (PPO) [24], Soft Actor-Critic (SAC) [25], Asynchronous Advantage Actor-Critic algorithm (A3C) [9] and Synchronous Advantage Actor-Critic algorithm (A2C) [9]. DRL algorithms can be further classified based on environment type. In [26], the authors classified the DRL algorithms based on the environment type, first they classified the environment into continuous and discrete, in the discrete environment discrete actions can be taken, and in the continuous environment, both discrete and continuous actions can be taken. The estimation of continuous actions does not depend on the action's probability, but on the action's statistics. Therefore, actions must be sampled from a continuous range, which may be a multi-variate or uni-variate Gaussian distribution. Value-based can only work in discrete environments, policy-based can work in continuous environments, and actor-critic can work in both discrete and continuous environments.

2.3. Challenges in Deep Reinforcement Learning for Autonomous Vehicles

Reviewing the research on decision making for autonomous systems, the challenges that still stand after all the proposed technologies, experiments, and research can be pointed out. In this section, the challenges faced when implementing a decision-making system for autonomous vehicles using DRL are discussed.

2.3.1. Simulator-to-Reality Gap: Considering the nature of RL, learning by trial and error, it can be extremely expensive and dangerous to train an agent in real life because it can expose the humans involved and other vehicles/agents to great dangers. Consequently, the performance of an RL agent in real life varies drastically from its performance in a simulator. Many researchers mentioned this challenge in their papers, including [27] and [21]. Approaches to overcome this challenge will be discussed in the next section.

2.3.2. Generalization: Each RL agent can be defined for a single objective or multiple objectives, as discussed in this chapter. This makes the ability of an agent to generalize to other environments non-existent. In other words, an agent trained for a specific objective or set of objectives cannot maintain the same performance in other tasks. Which makes it very computationally expensive to start from scratch and train an agent from scratch each time for a new objective or task [3] [5]

2.3.3. Verification: As mentioned above, it is impractical to train an RL agent in real life because of its nature; therefore, its verification and validation have become a challenge as well. How can an RL agent be verified if it cannot be tested in real life? This question remains a challenge [9], [27]

2.3.4. Safety: It is extremely important for decision making systems to provide safe driving [3] [5] [27], [21] but there is no safety standard in the literature for researchers to abide by [2,28]

2.3.5. Reward Function Design: Reward function is a crucial element in RL and shaping it can be an indicator of whether the agent will learn properly. Reward shaping can make learning faster because it gives hints to the agent about the desired state that will lead to the actual reward, which are pseudo-rewards that the agent is rewarded when it makes any progress.

2.3.6. Multi-Agent Reinforcement Learning: Multiple RL-based agents communicate with each other to obtain an optimal policy for each.

2.3.7. End-to-End Learning: End-to-end learning models are a challenge in the decision-making and control of autonomous systems.

2.4. Reward Function Design

Reward function design is a challenging area of research, as there are no standards or rules to ensure an effective and reliable design. However, researchers and engineers worked on this topic to facilitate the design process through conducting research on the topic. In [28] the authors reviewed all aspects of designing a reward function for autonomous vehicles and autonomous driving. In [2] the authors discussed the mistakes performed in the literature that affects the design and the resulted behaviour of the vehicle. For that matter, they proposed eight sanity checks to be considered when designing a reward function for autonomous vehicles or autonomous driving. Their sanity checks were used to evaluate the proposed design process. Reward function in autonomous driving should address several key aspects of driving like safety, comfort, progress to destination, fuel consumption, time spent driving, distance covered, …, etc. However, some attributes are not addressed enough as mentioned in [2] such as comfort. It was noted that many studies disregarded comfort in their work such as [6,7,9,29–32]. As far as our knowledge, no research addressed comfort completely until now. Comfort in literature is all passenger focused but since the number of passengers can differ throughout a single ride, the best way to evaluate comfort according to the literature is acceleration and its derivative, jerk [28]. Comfort or passenger experience can be improved using numerous ways, one of which is steering smoothness as it is used in[33–36]. Some studies [34,35] addresses steering smoothness by penalizing high steering angles. Ride comfort falls under the missing attributes problem of the reward function design, so comfort is taken into consideration by improving both acceleration and steering smoothness.

3. Methodology

Ride comfort optimization using DRL is based on the reward function design. As discussed in the previous section, reward function design is crucial for an agent's success, but it has many challenges to be designed properly. One of which is the problem description and the goal behind the design. However, autonomous driving being a multi-objective problem, certain goals cannot be specified or quantified. Only general goals can be to achieve and ensure safe and efficient driving. The first step in designing the reward function is to have a complete and clear problem description, but since the problem is autonomous driving, the agent cannot be expected to learn to drive as a human does. Thus, the expected behaviour is specified for a self-driving car set by end-users/consumers as stakeholders for which comfort is a priority and in this is study, comfort is the main parameter to be studied. The process can be explained using the pseudo code below:

Algorithm (1): Autonomous Vehicle Agent

```
Initialize environment, global actor, and critic networks weights θ and θ_v
Define number of maximum episodes (E) and number of workers (W)
Initialize local actor and critic networks weights θ' and θ'_v
Synchronise local actor and critic weights with global actor and critic θ = θ' and θ_v = θ'_v
For x in number of workers (W):
        While #episode < max episodes:
                Get Initial State (s_t)
                Get mean (μ) & Standard Deviation (σ) from Local Actor Network
                Sample action using mean (μ) & Standard Deviation (σ) ⟹ a_t
                Take action a_t
                For y number of states:
                        Local Critic Network Value Prediction V(s_t)
                        Calculate TD Targets G_t = r_t + γr_{t+1} + γ^2 r_{t+2} + ⋯ + γ^n V(s_{t+n})
                        Calculate the reward R and new state s_{t+1}
```

Calculate Advantages $A(s_t, a_t) = G_t - V(s_t)$
Update Global actor $d\theta + \nabla_{\theta'} \log \pi(a_t|s_t; \theta')(R - V(s_t; \theta'_v))$
Update Global Critic $\frac{d\theta_v + \partial\left(R - V(s_t; \theta'_v)\right)^2}{\partial\theta'_v}$
Perform asynchronous update of global networks weights with local networks: θ using $d\theta$ and of θ_v using $d\theta_v$

3.1. Environmental Setup
Setting up the environment for training the agent requires first selecting the simulator as this study is simulation-based. When designing an agent for autonomous driving, CARLA [37] is selected as the simulator and then came the selection of the Town in which the vehicle will drive. The authors of [37] used Town 01 for their evaluation of the RL algorithms, this is considered the benchmark for evaluating RL algorithms in CARLA, thus it was used. The simulation is done on CARLA simulator using Town 01 which is characterized by being a small, simple town with rivers and bridges. Next came the sensors setup, the input to the algorithm is an RGB image so an RGB camera was installed on the vehicle, an IMU sensor was installed to measure the acceleration in all directions and a collision sensor was installed to detect collisions.

3.2. Algorithm Selection and Neural Networks Architecture
As per the problem specification, the environment is continuous, stochastic, sequential, multi-agent, dynamic, and partially observable. The problem is multi-objective, driving in general is a multi-objective problem, but since the study is concerned with ride comfort, based on the survey done in Section 2.2 the algorithms to select from were PPO, DDPG and A3C, since the problem does not depend mainly on the algorithm, all three were good to use but A3C offers faster training and convergence in comparison to the other two in addition to its being the algorithm used by the authors of [37], so A3C is selected, Since continuous control is opted for so the actor network had to output both the mean and standard deviation. A3C works by creating parallel actors at the same time to replace having a memory like other algorithms, each actor updates the critic asynchronously which makes the agent have access to more than one state at a time. These asynchronous update works collectively to decrease the total training losses.

3.3. Reward Function Design
The reward function design needed for the proposed goals presented many challenges since the problem is multi-objective function for comfort optimization. The numerous factors affecting the passenger experience in the literature are studied first, and they were the acceleration, the steering smoothness, and jerks. The ISO2631-1 [38] standard is selected for the vertical acceleration to be within the comfortable and the slightly uncomfortable values stated in [38]. The proposed reward function aims to keep a smooth ride and ensures comfort, noting that incorporating multiple factors for ride comfort does not fall under the problem of redundant attributes mentioned in [2] as each factors contributes to the overall passenger experience, such that steering smoothness contributes to the lateral movement and maneuver, acceleration contributes to the longitudinal movement, jerks contributes to both lateral and longitudinal movements, and vertical acceleration influences the vibrations affecting the human spine which directly affects the human experience. Furthermore, applying arbitrary weights to the reward terms can introduce bias without a justification; hence, it was not used as this approach ensures that all factors are of equal importance and will contribute equally to a balanced optimization process and a step further to insist that each factor plays a distinct role in optimizing the ride comfort. Thus, the final reward term for ride comfort optimization is as follows:

$$R = -(a_{vertical} + J_{longitudinal} + J_{lateral} + S_{steering}) \quad (1)$$

Where: $a_{vertical}$ is the vertical acceleration penalty, $J_{longitudinal}$ is the longitudinal jerk penalty, $J_{lateral}$ is the lateral jerk penalty, and $S_{steering}$ is the steering smoothness penalty. Other rewards are shown in the table:

Table 1. Proposed Reward Function

Ref	Reward
Reward (1) [9]	$v\ \cos\alpha$
Reward (2) [6,7]	$v\ \cos\alpha - d$
Ours	$-(a_{vertical} + J_{longitudinal} + J_{lateral} + S_{steering})$

3.4. Model Training

The A3C framework allows us to create multiple instances of the proposed agent. By creating multiple instances, the number of training episodes for the agents can be reduced, 20 actors were used, for 3000 episodes and a 100-state update interval, The actor network is a Single Input Multiple Output (SIMO) Neural Network where the RGB image is passed as an input to the network, and it outputs a mean and a standard deviation. While the critic network is a Single Input Multiple Output (SISO) Neural network where it received the RGB image as an input and outputs the predicted value. Table 2 shows the problem setup. In Figure 4, n-Actors work together to asynchronously update the whole model.

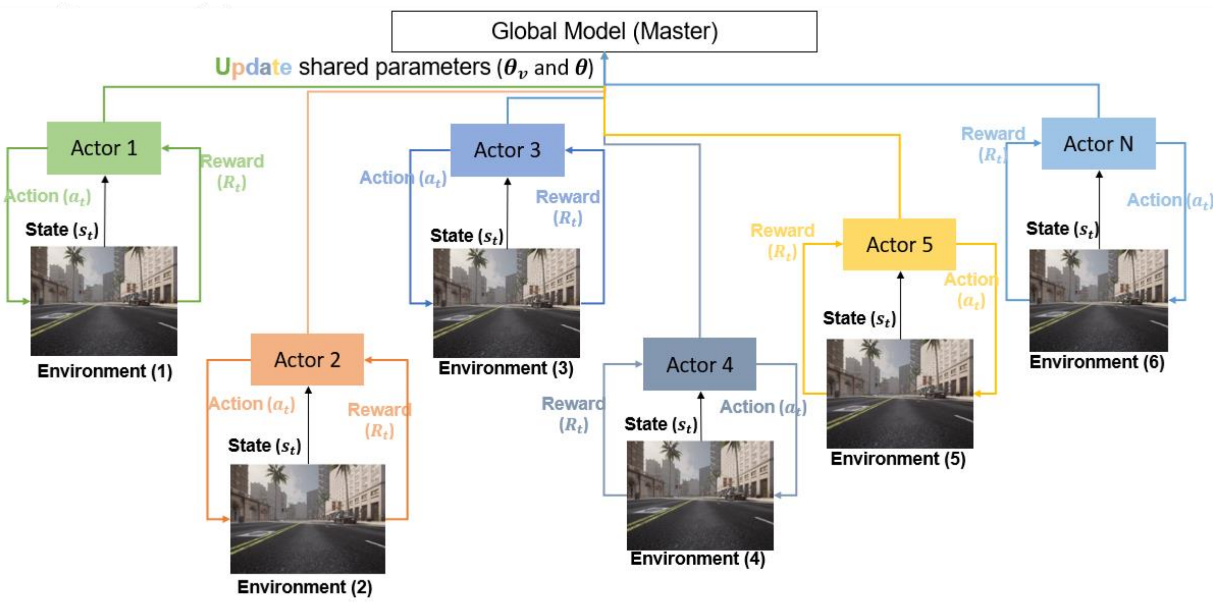

Figure 4. The Proposed Model Setup

Each actor in the shown model setup (Figure 4) works as shown in figure 5. After each reaches terminal state or the update interval ends, the actor updates the global model by equation (2) and global critic is updated by equation (3). Global model is locked during updates to prevent two actors from updating at the same time.

$$d\theta + \nabla_{\theta'} \log \pi(a_t|s_t;\ \theta')(R - V(s_t;\theta_v')) \qquad (2)$$

$$\frac{d\theta_v + \partial\left(R - V\left(s_t;\theta_v'\right)\right)^2}{\partial\theta_v'} \qquad (3)$$

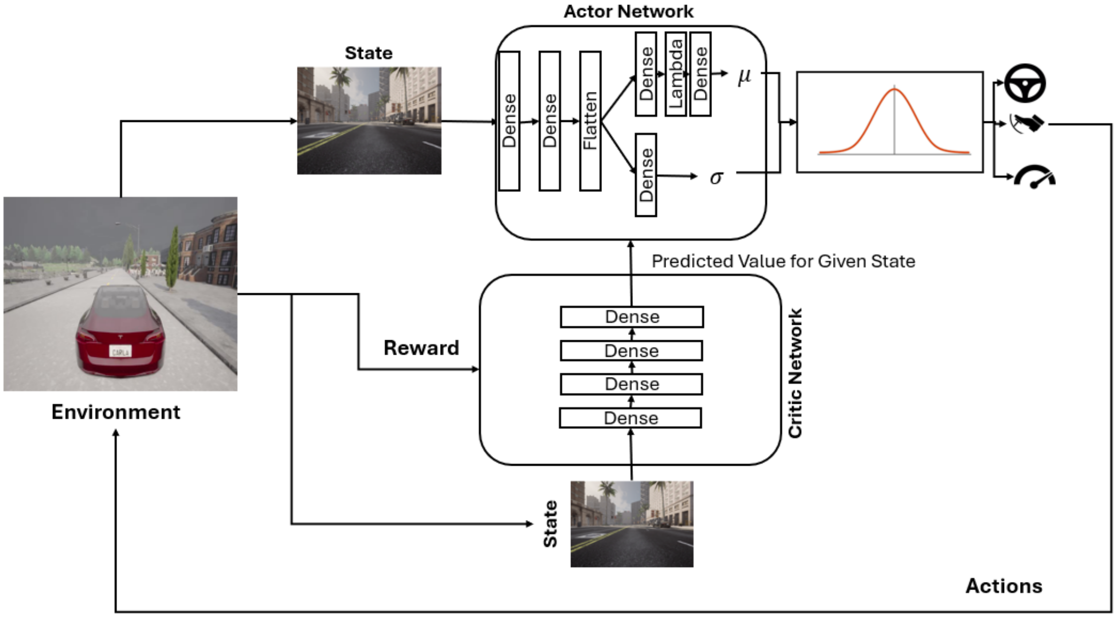

Figure 5. Proposed Model Training Algorithm

This makes our problem setup as follows:

Table 2. Problem Setup

Observations (States)	RGB Image
Environment	CARLA Simulator (Town 01)
Actions	Continuous Action Space - Steering $\in [-1, 1]$ - Braking $\in [0, 1]$ - Throttling $\in [0, 1]$
Rewards	$R = -(a_{vertical} + J_{longitudinal} + J_{lateral} + S_{steering})$

4. Results and Discussion

The training of the A3C algorithm was done for 300k simulation steps but for the sake of visualization it was sampled, the results showed that the proposed reward function surpasses the other rewards, reward 1 is the original paper [9] and reward 2 is [6,7]. In figure 7, The graph compares the rewards per episode, while R_1 and R_2 stabilize early around a reward value of 10, the proposed reward function achieves significantly higher rewards, starting near 10^1 and gradually decreasing but remaining in the 10 range even after 3000 episodes. This suggests that the proposed approach learns a more effective policy that consistently outperforms the baselines by several orders of magnitude. Additionally, the smooth and stable trend of the proposed approach indicates robustness, while R_1 and R_2 show little improvement over time. The logarithmic scale further emphasizes the vast gap in performance, demonstrating that the proposed approach is far superior in maximizing rewards. Overall, the consistency of the proposed reward function indicates a better performance than the other two reward function, but an interpretation in terms of acceleration is still needed to show how the ride comfort was optimized.

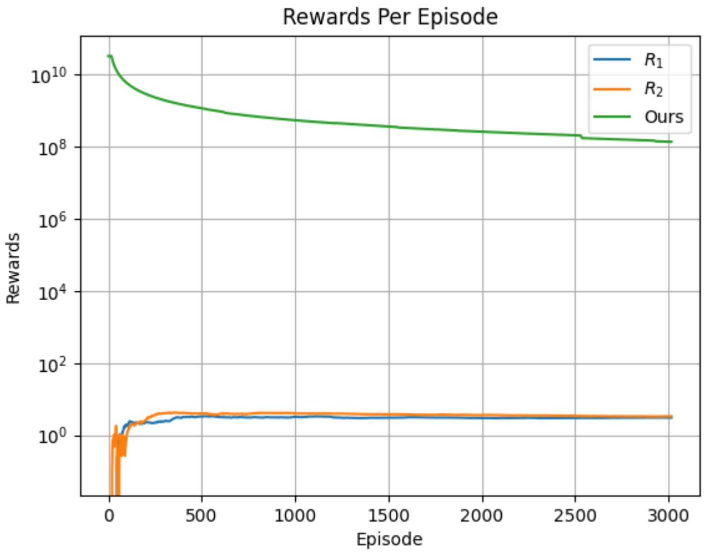

Figure 6. Reward Per Episode

Figures 8 (a) and (b) compare the proposed reward function against R_1 and R_2, clearly demonstrating that the proposed approach leads to significantly lower and more stable accelerations over the episodes. Both R_1 and R_2 exhibit high fluctuations, with sharp spikes indicating abrupt changes in acceleration. In contrast, the proposed approach consistently maintains lower and more controlled acceleration values, with significantly fewer peaks. This suggests that the proposed approach achieves a smoother driving profile, minimizing rapid acceleration variations that could negatively impact passenger comfort. The reduced variance in acceleration further highlights the robustness of the proposed approach, ensuring a more predictable and controlled driving experience. Since high accelerations contribute to discomfort, the proposed approach effectively optimizes ride quality by prioritizing stability while still achieving high rewards. This balance between maximizing performance and maintaining smooth acceleration makes the proposed approach a superior solution for autonomous vehicle control, significantly improving passenger experience. This shows that by sustaining low and steady acceleration levels with noticeably fewer and smaller spikes, the proposed approach performs better than both R_1 and R_2, achieving a 44.34% reduction and more consistent vertical acceleration.

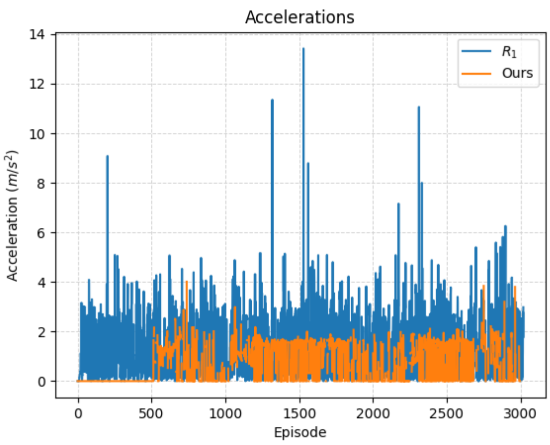

Figure 8 (a). Acceleration Data (Ours vs R_1)

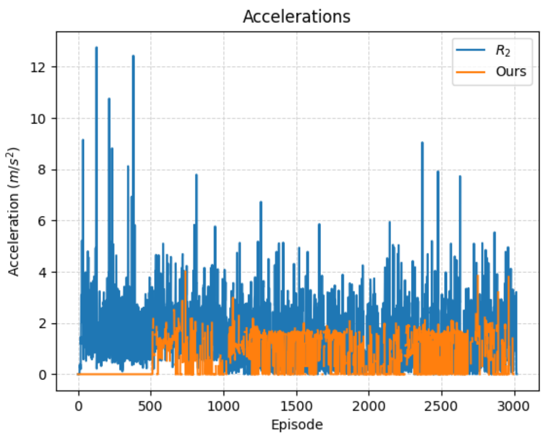

Figure 8 (b). Acceleration Data (Ours vs R_2)

The two graphs in Figures 9 (a) and (b) compare vertical acceleration across episodes. The vertical acceleration metric is crucial in evaluating ride comfort, as lower and more stable values indicate a smoother driving experience. In both graphs, R_1 and R_2 exhibit highly fluctuating and often high vertical accelerations, which suggests significant oscillations and discomfort for passengers. In contrast, the proposed approach maintains significantly lower and more stable vertical acceleration levels throughout all episodes. The controlled acceleration profile of the proposed approach indicates better optimization in minimizing vibrations, which directly translates to a smoother and more comfortable ride. Furthermore, the reduction in extreme acceleration spikes demonstrates the effectiveness of the proposed approach in mitigating sudden disturbances. The consistency and stability seen in the proposed approach reinforce its superiority in optimizing ride comfort, making it a more effective solution for real-world autonomous driving applications.

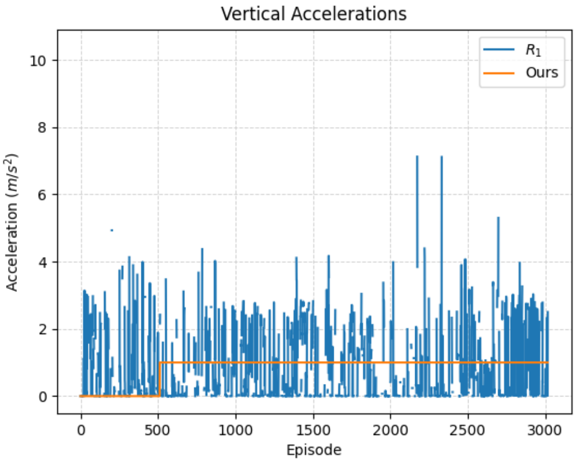

Figure 9 (a). Vertical Acceleration Data (Ours vs R_1)

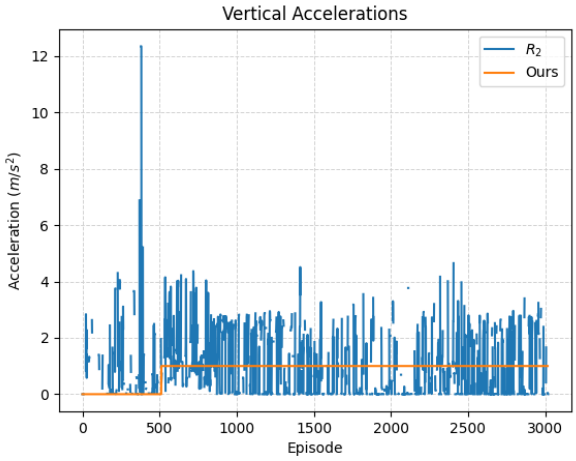

Figure 9 (b). Vetircal Acceleration Data (Ours vs R_2)

The Fast Fourier Transform (FFT) is utilized to analyze vertical acceleration and gain deeper insight into the vehicle's behavior, further validating the results. As shown in Figure 10, the amplitude gradually decreases as frequency increases, indicating a well-controlled acceleration profile. The critical 4–8 Hz range, marked by dashed lines, represents the frequency band where human sensitivity to vibrations is highest. The results show that while there is some signal energy within this range, amplitudes steadily decay, suggesting effective vibration damping and minimizing discomfort. The absence of sharp peaks indicates that no dominant frequency is excessively amplified, preventing resonance effects that could worsen ride quality. Furthermore, the higher-frequency components above 8 Hz are significantly suppressed, reducing sudden, high-frequency jolts that might otherwise lead to discomfort. This suppression of high-frequency vibrations contributes to a smoother and more stable ride, as abrupt acceleration changes are minimized. The FFT analysis supports comfort optimization by showing that acceleration energy is primarily distributed in a controlled manner, avoiding excessive oscillations and ensuring a more predictable and comfortable ride experience.

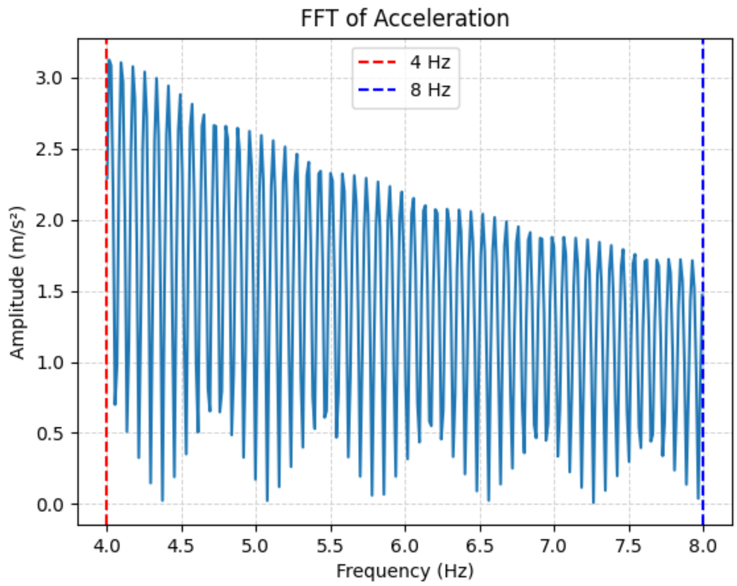

Figure 10. FFT Analysis

5. Conclusion

This paper proposes an end-to-end DRL model with a novel reward function specifically designed to optimize ride comfort by addressing key factors that uniquely impact the passenger experience. By carefully considering acceleration constraints and leveraging RL, the model effectively minimizes discomfort-inducing forces, ensuring a smoother ride. The experimental results demonstrate a significant improvement in overall acceleration, achieving a 44.34% reduction compared to the baseline model. Additionally, the proposed approach consistently maintains more stable vertical acceleration values across episodes, as evidenced by the comparative analysis against alternative reward designs. The reduced occurrence of high-magnitude acceleration spikes indicates improved control over vehicle dynamics, directly translating to enhanced ride quality. This stability is crucial where passenger comfort is a key factor in the adoption and trust of AVs. By systematically optimizing acceleration profiles, the proposed model contributes to a more seamless and reliable driving experience, highlighting the potential of DRL in refining autonomous vehicle control strategies. These findings highlight the effectiveness of the proposed reward function and set the foundation for future research into more advanced comfort-aware RL frameworks for intelligent transportation systems.

References

[1] Hook D W, Porter S J and Herzog C 2018 Dimensions: Building Context for Search and Evaluation *Front Res Metr Anal* **3** 23

[2] Knox W B, Allievi A, Banzhaf H, Schmitt F and Stone P 2023 Reward (Mis)design for autonomous driving *Artif Intell* **316** 103829

[3] Schwarting W, Alonso-Mora J and Rus D 2018 Planning and Decision-Making for Autonomous Vehicles *Annu. Rev. Control Robot. Auton. Syst*

[4] Garrido F and Resende P 2022 Review of Decision-Making and Planning Approaches in Automated Driving *IEEE Access* **10** 100348–66

[5] Liu Q, Li X, Yuan S and Li Z 2021 Decision-Making Technology for Autonomous Vehicles: Learning-Based Methods, Applications and Future Outlook *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)* 30–7

[6] Pan X, You Y, Wang Z and Lu C 2017 Virtual to Real Reinforcement Learning for Autonomous Driving *Proceedings of the British Machine Vision Conference*

[7] Xu N, Tan B and Kong B 2019 Autonomous Driving in Reality with Reinforcement Learning and Image Translation

[8] Jaritz M, de Charette R, Toromanoff M, Perot E and Nashashibi F 2018 End-to-End Race Driving with Deep Reinforcement Learning **2**
[9] Mnih V, Badia A P, Mirza M, Graves A, Lillicrap T P, Harley T, Silver D and Kavukcuoglu K 2016 Asynchronous Methods for Deep Reinforcement Learning *International Conference on Machine Learning*
[10] Samak C V, Samak T V and Kandhasamy S 2022 Autonomous Racing using a Hybrid Imitation-Reinforcement Learning Architecture
[11] Aledhari M, Rahouti M, Qadir J, Qolomany B, Guizani M and Al-Fuqaha A 2023 Motion Comfort Optimization for Autonomous Vehicles: Concepts, Methods, and Techniques *IEEE Internet Things J* **11** 378–402
[12] Elbanhawi M, Simic M and Jazar R 2015 In the Passenger Seat: Investigating Ride Comfort Measures in Autonomous Cars *IEEE Intelligent Transportation Systems Magazine* **7** 4–17
[13] Wang C, Cui X, Zhao S, Zhou X, Song Y, Wang Y and Guo K 2024 Enhancing vehicle ride comfort through deep reinforcement learning with expert-guided soft-hard constraints and system characteristic considerations *Advanced Engineering Informatics* **59**
[14] Du Y, Chen J, Zhao C, Liao F and Zhu M 2023 A hierarchical framework for improving ride comfort of autonomous vehicles via deep reinforcement learning with external knowledge *Computer-Aided Civil and Infrastructure Engineering* **38** 1059–78
[15] Han S Y and Liang T 2022 Reinforcement-Learning-Based Vibration Control for a Vehicle Semi-Active Suspension System via the PPO Approach *Applied Sciences (Switzerland)* **12**
[16] Guruguntla V, Lal M, Ghantasala G S P, Vidyullatha P, Alqahtani M S, Alsubaie N, Abbas M and Soufiene B O 2023 Ride comfort and segmental vibration transmissibility analysis of an automobile passenger model under whole body vibration *Scientific Reports 2023 13:1* **13** 1–13
[17] Deubel C, Ernst S and Prokop G 2023 Objective evaluation methods of vehicle ride comfort—A literature review *J Sound Vib* **548**
[18] Huber W 2016 Industrie 4.0 in der Automobilproduktion *Industrie 4.0 in der Automobilproduktion*
[19] Genser A, Spielhofer R, Nitsche P and Kouvelas A 2022 Ride comfort assessment for automated vehicles utilizing a road surface model and Monte Carlo simulations *Computer-Aided Civil and Infrastructure Engineering* **37** 1316–34
[20] de Winkel K N, Irmak T, Happee R and Shyrokau B 2023 Standards for passenger comfort in automated vehicles: Acceleration and jerk *Appl Ergon* **106**
[21] Talpaert V, Sobh I, Ravi Kiran B, Mannion P, Yogamani S, El-Sallab A and Perez P 2019 Exploring Applications of Deep Reinforcement Learning for Real-world Autonomous Driving Systems *VISIGRAPP 2019 - Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications* vol 5 (SciTePress) pp 564–72
[22] Sutton R S and Barto A 2018 *Reinforcement Learning: An Introduction*
[23] Lillicrap T P, Hunt J J, Pritzel A, Heess N, Erez T, Tassa Y, Silver D and Wierstra D 2015 Continuous control with deep reinforcement learning
[24] Schulman J, Wolski F, Dhariwal P, Radford A and Openai O K 2017 Proximal Policy Optimization Algorithms
[25] Haarnoja T, Zhou A, Abbeel P and Levine S 2018 Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor *35th International Conference on Machine Learning, ICML 2018* **5** 2976–89

Journal of Physics: Conference Series **3058** (2025) 012003 doi:10.1088/1742-6596/3058/1/012003

[26] AlMahamid F and Grolinger K 2022 Reinforcement Learning Algorithms: An Overview and Classification *IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*
[27] Kiran B R, Sobh I, Talpaert V, Mannion P, Sallab A A A, Yogamani S and Perez P 2022 Deep Reinforcement Learning for Autonomous Driving: A Survey *IEEE Transactions on Intelligent Transportation Systems* **23** 4909–26
[28] Abouelazm A, Michel J and Zoellner J M 2024 A Review of Reward Functions for Reinforcement Learning in the context of Autonomous Driving
[29] Wu L-C, Zhang Z, Haesaert S, Ma Z and Sun Z 2023 Risk-Aware Reward Shaping of Reinforcement Learning Agents for Autonomous Driving
[30] Sharif A and Marijan D 2021 Evaluating the Robustness of Deep Reinforcement Learning for Autonomous Policies in a Multi-agent Urban Driving Environment
[31] Wang L, Liu J, Shao H, Wang W, Chen R, Liu Y and Waslander S L 2023 Efficient Reinforcement Learning for Autonomous Driving with Parameterized Skills and Priors
[32] Palanisamy P 2019 Multi-Agent Connected Autonomous Driving using Deep Reinforcement Learning *Proceedings of the International Joint Conference on Neural Networks*
[33] Liang X, Wang T, Yang L and Xing E 2018 CIRL: Controllable Imitative Reinforcement Learning for Vision-based Self-driving
[34] Chen J, Yuan B and Tomizuka M 2019 Model-free Deep Reinforcement Learning for Urban Autonomous Driving *2019 IEEE Intelligent Transportation Systems Conference (ITSC)* pp 2765–71
[35] Chen J, Li S E and Tomizuka M 2020 Interpretable End-to-end Urban Autonomous Driving with Latent Deep Reinforcement Learning *IEEE Transactions on Intelligent Transportation Systems* **23** 5068–78
[36] Wolf P, Hubschneider C, Weber M, Bauer A, Härtl J, Dürr F and Zöllner J M 2017 Learning how to drive in a real world simulation with deep Q-Networks *2017 IEEE Intelligent Vehicles Symposium (IV)* pp 244–50
[37] Dosovitskiy A, Ros G, Codevilla F, Lopez A and Koltun V 2017 CARLA: An Open Urban Driving Simulator
[38] Anon ISO 2631-1:1997(en), Mechanical vibration and shock — Evaluation of human exposure to whole-body vibration — Part 1: General requirements