Jurnal Teknik Informatika (JUTIF) Vol. 6, No. 3, Juni 2025, Page. 1189-1202
P-ISSN: 2723-3863 https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871 DOI: https://doi.org/10.52436/1.jutif.2025.6.3.4446

Deep Reinforcement Learning for Autonomous System Optimization in Indonesia: A Systematic Literature Review

Dedi Yusuf*[1], Eko Supraptono[2], Agus Suryanto[3]

[1,2,3]Electrical Engineering Study Program, Faculty of Engineering, Semarang State University, Semarang, Indonesia

Email: [1]dedi11519@students.unnes.ac.id

Received : Feb 23, 2025; Revised : Apr 30, 2025; Accepted : May 3, 2025; Published : Jun 10, 2025

Abstract

Background: The development of artificial intelligence (AI) technology, including Deep Reinforcement Learning (DRL), has brought significant changes in various industrial sectors, especially in autonomous systems. DRL combines the capabilities of Deep Learning (DL) in processing complex data with those of Reinforcement Learning (RL) in making adaptive decisions through interaction with the environment. However, the application of DRL in autonomous systems still faces several challenges, such as training stability, model generalization, and high data and computing resource requirements. **Methods:** This study uses the Systematic Literature Review (SLR) method to identify, evaluate, and analyze the latest developments in DRL for autonomous system optimization. The SLR was conducted by following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework, which consists of four main stages: identification, screening, eligibility, and inclusion of research articles. Data were collected through literature searches in leading scientific journal databases such as IEEE Xplore, MDPI, ACM Digital Library, ScienceDirect (Elsevier), SpringerLink, arXiv, Scopus, and Web of Science. **Results:** This study found that DRL has been widely adopted in various industrial sectors, including transportation, industrial robotics, and traffic management. The integration of DRL with other technologies such as Computer Vision, IoT, and Edge Computing further enhances its capability to handle uncertain and dynamic environments. Therefore, this study is crucial in providing a comprehensive understanding of the potential, challenges, and future directions of DRL development in autonomous systems, in order to foster more adaptive, efficient, and reliable technological innovations.

Keywords: *Autonomous System Optimization, Deep Reinforcement Learning, Systematic Literature Review.*

1. INTRODUCTION

In recent decades, the development of artificial intelligence (AI) technology has brought significant changes in various industrial sectors, including transportation, manufacturing, and robotics. One branch of AI that is increasingly gaining attention is Deep Reinforcement Learning (DRL), which combines the advantages of Deep Learning (DL) in complex data processing with the ability of Reinforcement Learning (RL) in making adaptive decisions through interaction with the environment [1]. Autonomous systems, such as driverless vehicles, industrial robots, and traffic management systems, require adaptive, efficient decision-making mechanisms that are able to handle environmental uncertainty. In this context, DRL is one of the promising approaches to improve the performance and efficiency of autonomous systems.

As the demand for autonomous systems increases, the global autonomous vehicle market is projected to reach USD 556.67 billion by 2026, growing at a compound annual growth rate (CAGR) of 39.47% between 2019 and 2026 [2]. The smart manufacturing sector that relies on DRL-based industrial robots is predicted to experience rapid growth, driven by the need for smarter and more efficient

Vol. 6, No. 3, Juni 2025, Page. 1189-1202
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.3.4446

automation [3]. While DRL offers several advantages in optimizing autonomous systems, its real-world application still faces a number of technical and methodological challenges.

In addition to the urgency at the global level, the development of DRL-based autonomous systems is also a priority in the Indonesian context. The Making Indonesia 4.0 Roadmap launched by the Ministry of Industry targets increasing automation and AI integration in the manufacturing and transportation sectors as one of the main strategies to improve the competitiveness of the national industry [4]. In addition, the 2017–2045 National Research Master Plan (RIRN) emphasizes the importance of developing AI technology and intelligent systems as part of the national digital transformation [5]. This research not only has academic relevance but also supports the direction of national policy in accelerating the application of AI and autonomous systems in various strategic sectors.

One of the main challenges is the stability and efficiency of training. Training DRL models requires a large number of iterations to achieve optimal convergence, which is often very time-consuming and computationally resource-intensive [6]. For example, Deep Q-Networks (DQN)-based models require millions of interactions with the environment before achieving adequate performance [7]. Generalization of DRL models to real environments remains a major problem, as models trained in simulation often experience performance degradation when applied in the real world due to differences in environmental dynamics [8]. Another crucial issue is the need for large amounts of data and computational resources. DRL requires a large amount of exploration to build optimal strategies for decision making, which often results in low data efficiency and slow learning processes [9]. This challenge is further exacerbated by the need for specialized hardware, such as graphics processing units (GPUs) or tensor processing units (TPUs), which adds to the financial burden for developers and researchers in this field.

Various approaches have been proposed to address this issue, including the use of simulation-based models to reduce the reliance on real-world data and transfer learning techniques that allow models trained in simulation to be adapted to real-world environments with less additional data [10]. In addition, integration with unsupervised learning and meta-learning techniques has been a potential strategy to improve the generalization of DRL models [11]. Many previous studies have discussed the application of DRL in autonomous systems, but there are still gaps in terms of optimizing training efficiency, improving model generalization, and adapting to real-world environments. Some previous studies have focused more on the technical aspects of DRL algorithm development, but not many have systematically evaluated the effectiveness of existing approaches and identified key challenges in their implementation in real-world autonomous systems [12]. Therefore, this study contributes by presenting a systematic review that not only examines the latest methods in DRL but also provides perspectives on future research directions and potential strategies to improve the efficiency and effectiveness of DRL-based systems in real-world environments.

Considering the existing challenges and potential solutions that have been proposed in previous studies, a more comprehensive study is needed to summarize the latest developments and design implementation strategies that are more efficient and can be widely adopted across industries. Based on the urgency of the problem, this study aims to conduct a systematic literature review of the latest developments in DRL for autonomous system optimization, identify the methods that have been used, the challenges that are still faced, and the direction of future research. Thus, this study is expected to provide a comprehensive picture of the evolution of DRL in autonomous systems and potential solutions to improve its performance and efficiency in the real world.

2. METHOD

This study adopts the Systematic Literature Review (SLR) method to identify, evaluate, and analyze recent advancements in Deep Reinforcement Learning (DRL) for autonomous system

Vol. 6, No. 3, Juni 2025, Page. 1189-1202
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.3.4446

optimization. SLR is a systematic and transparent approach for collecting and synthesizing relevant research findings, aiming to provide comprehensive research mapping while identifying current trends and research gaps in the field [13]. The implementation of this method follows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework, consisting of four main stages: identification, screening, eligibility, and inclusion of research articles [14].

Data collection was conducted through comprehensive literature searches in reputable scientific journal databases such as IEEE Xplore, MDPI, ACM Digital Library, ScienceDirect (Elsevier), SpringerLink, arXiv, Scopus, and Web of Science. The search strategy involved combinations of keywords relevant to the research topic, including "Deep Reinforcement Learning" AND "Autonomous Systems", "Reinforcement Learning" AND "Optimization" AND "Autonomous Systems", and "Deep Learning" AND "Autonomous Control". The search was limited to publications from 2020 to 2024 to ensure the inclusion of only recent and relevant studies.

Literature selection was based on inclusion and exclusion criteria. The inclusion criteria consisted of articles discussing the application of DRL in autonomous systems (e.g., autonomous vehicles, industrial robotics, traffic management systems), studies focusing on DRL optimization (e.g., training efficiency, transfer learning, real-world adaptation), and publications in reputable journals or conferences (Q1–Q3 according to Scopus rankings). The exclusion criteria included articles that were not available in full-text, studies focusing solely on Reinforcement Learning without Deep Learning components, and non-experimental or opinion-based publications. Articles meeting the inclusion criteria underwent manual screening to ensure their relevance to the research objectives [15].

In the data processing stage, key information from each selected article was extracted using reference management software such as Mendeley. The extracted parameters included the authors' names and publication years, the DRL methods employed (e.g., DQN, PPO, A3C, SAC), application domains within autonomous systems, strengths and limitations of each method, and the main results and findings of the studies. A thematic analysis was then conducted to identify research patterns, trends, and evaluate the performance and effectiveness of various DRL approaches. The results were synthesized and presented in tables, PRISMA flow diagrams, and additional visualizations (where applicable) to offer a comprehensive overview of current developments, challenges, and potential research directions for DRL in autonomous systems. Where relevant, mathematical formulas such as reward function equations, convergence rate calculations, or model accuracy metrics were also incorporated to support the evaluation of DRL performance [16]. The results of this analysis were then synthesized to provide a comprehensive picture of the developments, challenges, and potential for future research in the field of DRL for autonomous systems.

3. RESULT

Recent Developments in DRL Applications for Autonomous Systems.

3.1. Widespread Adoption Of Drl In Various Sectors

Deep Reinforcement Learning (DRL) has seen increasing adoption in various industrial sectors, especially in autonomous systems that require adaptive decision-making capabilities. In the transportation sector, DRL has been used to optimize autonomous vehicle navigation to improve efficiency and road safety. The global autonomous vehicle market is projected to reach USD 556.67 billion by 2026, with a compound annual growth rate (CAGR) of 39.47% between 2019 and 2026 [17]. This growth is largely driven by the increasing adoption of artificial intelligence technologies, including DRL, in vehicle control systems. The industrial robotics sector, DRL is being used to improve the operational efficiency of robots in smart manufacturing environments. DRL-based robots can reduce production time by up to 30% compared to conventional rule-based systems [18].

P-ISSN: 2723-3863 https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871 DOI: https://doi.org/10.52436/1.jutif.2025.6.3.4446

Traffic management system sector, DRL is applied to adaptively optimize traffic light settings based on vehicle density, which can reduce congestion by up to 25% [19]. Meanwhile, in the manufacturing sector, DRL has enabled smarter automation, with studies showing that implementing DRL in manufacturing control systems can increase production throughput by up to 15% compared to traditional heuristic-based methods [20]. With the increasing application of DRL in various industries, this technology has become one of the main solutions in optimizing autonomous systems to improve operational efficiency and effectiveness.

Table 1. DRL Performance in Various Sectors

Sector	Performance Indicator	Reference
Autonomous Vehicles	Increase in object recognition accuracy to 94.3%	[21]
Industrial Robotics	Reduction in production time by 30%	[18]
Manufacturing	Increase in production throughput by 15%	[20]
Traffic Management	Reduction in congestion by 25%	[19]
Predictive Maintenance (IoT)	Reduction in machine downtime by 20%	[22]

Deep reinforcement learning (DRL) significantly improves performance across various sectors. In autonomous vehicles, it increases object recognition accuracy. Industrial robotics see a reduction in production time. Manufacturing experiences increased production throughput, while traffic management sees less congestion. Finally, predictive maintenance using IoT and DRL reduces machine downtime. These results demonstrate DRL's broad applicability and effectiveness in optimizing diverse industrial processes.

3.2. Integration Of Drl With Other Technologies

Along with the development of technology, DRL is increasingly combined with various other technologies such as Computer Vision (CV), Internet of Things (IoT), and Edge Computing to improve operational efficiency in autonomous systems. In the context of autonomous vehicles, the integration of DRL with Computer Vision allows the system to recognize and interpret objects in the environment in real-time, thereby improving navigation accuracy and obstacle detection. DRL model combined with CV can improve object recognition accuracy in autonomous vehicles up to 94.3%, compared to conventional methods which only reach 87.6% [21].

In the industrial sector, the integration of DRL with IoT enables the system to collect and analyze data from various sensors to improve production efficiency and early detection of anomalies. The combination of DRL and IoT in a predictive maintenance system can reduce machine downtime by up to 20% and increase energy efficiency by 12% [22]. In addition, Edge Computing plays a vital role in enabling DRL inference to be performed directly on local devices without having to rely on cloud computing, thereby reducing latency in decision making. The use of Edge Computing in a DRL-based system can reduce latency by up to 40%, making it a highly relevant solution for applications that require real-time response [23].

3.3. Improved Adaptability Of Drl In Dynamic Environments

One of the major developments in DRL is its improved adaptability in dealing with environmental uncertainty. Current DRL models have been designed with more flexible learning mechanisms, allowing the system to dynamically adjust decision-making strategies based on changing conditions in the surrounding environment. For example, the use of Domain Randomization in DRL training has been shown to improve the model's generalization ability to unexpected environmental changes [24]. DRL models trained with this technique can maintain their performance in a variety of real-world scenarios with over 85% accuracy, compared to only 70% for models that do not use the technique [25].

Vol. 6, No. 3, Juni 2025, Page. 1189-1202
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.3.4446

The development of the Meta-Reinforcement Learning method allows the DRL system to learn faster in new scenarios, without the need for extensive retraining. Meta-DRL-based model can adapt to new tasks 4–5 times faster than conventional models, making it a promising solution for autonomous systems operating in ever-changing environments, such as autonomous drones and search and rescue robots [26].

3.4. Application Of Drl In Real-Time Decision Making

The ability of DRL to perform real-time decision making has been significantly improved with the optimization of algorithms and improvements in computational efficiency. In intelligent transportation systems, DRL has been applied to dynamically manage traffic with low latency, allowing traffic regulation based on real-time data collected from road sensors. The use of DRL in adaptive traffic systems can reduce vehicle waiting time at intersections by up to 18% compared to rule-based methods [27].

In autonomous vehicles, DRL algorithms such as Soft Actor-Critic (SAC) and Proximal Policy Optimization (PPO) have been applied to improve the efficiency of decision-making in navigation and speed control. SAC enables more stable decision-making with low latency, reducing the navigation error rate by 15% compared to conventional DRL methods such as Deep Q-Network (DQN) [28].

3.5. Advances In Interpretability And Reliability Of Drl Models

One of the main challenges in the application of DRL in autonomous systems is the limited interpretability of the model, which can hinder its adoption in critical applications such as autonomous vehicles and industrial robotics. To address this issue, recent research has focused on developing methods that increase the transparency of decisions made by DRL models. One of the widely developed approaches is Explainable Reinforcement Learning (XRL), which aims to provide clearer insights into how DRL models make decisions in a given environment.

XRL method can increase user confidence in the DRL model by providing rule-based explanations for each decision taken by the agent. In addition, the application of Saliency Map and Attention Mechanisms techniques has allowed the visualization of the most influential factors in the decision-making process, thereby increasing the transparency of the system [29].

4. DISCUSSIONS

4.1. Methods To Improve DRL Efficiency And Effectiveness In Real World Environments

4.1.1. Transfer Learning And Few-Shot Learning To Reduce Training Data Requirements

One of the main challenges in implementing Deep Reinforcement Learning (DRL) in autonomous systems is the high demand for training data. In real-world environments, collecting high-quality data is very expensive and time-consuming. Therefore, the Transfer Learning approach is used to transfer the knowledge gained from the simulated environment to the real world, thereby reducing the need for training data. A DRL model using Transfer Learning can accelerate the adaptation process in a new environment by up to 40%, compared to a model trained from scratch [30].

Few-Shot Learning allows the DRL system to learn with a limited amount of data through a better generalization mechanism. The use of Few-Shot Learning in DRL can increase training efficiency by 30% while maintaining high accuracy in robotic navigation tasks. With this approach, autonomous systems can adapt more quickly to environmental changes without requiring a large amount of data, which is often a constraint in real-world applications [31].

Model-Agnostic Meta-Learning (MAML), showing that task-to-task transfer significantly reduces data requirements in robotics, emphasizing the relevance of these methods for operational efficiency in real-world AI systems. The urgency of this approach in computer science lies in the increasing demand for adaptive AI systems in data-scarce environments, such as edge computing and sensor-based real-time monitoring.

4.1.2. Meta-Learning And Self-Supervised Learning For Faster Adaptation

Meta-Learning and Self-Supervised Learning approaches have been used to improve the adaptation speed of DRL models to new environmental conditions. Meta-Learning, also known as "learning to learn," allows DRL models to learn faster by adjusting model parameters based on previous experiences. Meta-Learning can improve training efficiency up to 5 times faster compared to traditional learning methods, especially in robotic control-based tasks [32].

On the other hand, Self-Supervised Learning allows systems to learn automatically without the need for explicit labels, thus reducing the reliance on expensive labeled data. Self-Supervised Learning method in DRL can improve learning efficiency by up to 25%, especially in scenarios where the amount of labeled data is very limited. The integration of these two techniques in DRL allows autonomous systems to learn more efficiently and adaptively in various dynamic environmental conditions [33]. Demonstrated that curiosity-driven self-supervision greatly improves exploration in reward-sparse environments, reinforcing the value of combining Meta-Learning and Self-Supervised Learning for robust adaptation in robotics, autonomous vehicles, and intelligent agents.

4.1.3. Using Simulation-Based Models To Improve Training Efficiency

One of the main strategies in improving DRL training efficiency is to use simulation-based models before applying them to the real world. Simulation allows DRL algorithms to learn in a safe and controlled environment, avoiding the risks associated with direct experiments in the physical world. DRL agents trained in simulation can achieve optimal performance up to 50% faster than agents trained only in real-world environments [34].

Some popular simulators used in DRL research include CARLA for autonomous vehicles, MuJoCo for robotics, and AirSim for drones. By using simulation, researchers can test various control and exploration strategies without having to face high cost and time constraints. The Domain Randomization technique in simulation can improve the generalization ability of the model by simulating a variety of diverse environmental conditions [35]. Who showed that domain randomization significantly mitigates the Sim-to-Real Gap, enhancing real-world performance reliability. In the context of computer science, this approach accelerates the prototyping of cost-effective and safe AI systems, crucial in applied informatics and intelligent automation.

4.1.4. Exploration And Exploitation Optimization With PPO And SAC Algorithms

In DRL the balance between exploration (finding new actions) and exploitation (leveraging previous experience) is an important challenge in achieving learning efficiency. Two main algorithms that have been developed to address this issue are Proximal Policy Optimization (PPO) and Soft Actor-Critic (SAC). The PPO algorithm, uses a constraint-based approach in policy updating, thereby improving training stability. PPO can improve learning efficiency by up to 20% compared to previous DRL algorithms such as Trust Region Policy Optimization (TRPO) [36].

The SAC algorithm introduces an entropy-based exploration mechanism, which allows DRL agents to be more efficient in exploring new actions without losing training stability. SAC can improve convergence speed by up to 35% compared to classical methods such as Deep Q-Network (DQN). By adopting more optimal exploration techniques, DRL-based autonomous systems can learn faster with lower risks [37].

Vol. 6, No. 3, Juni 2025, Page. 1189-1202
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.3.4446

4.1.5. Computational Resource Efficiency Through Model And Hardware Optimization

One of the main obstacles in the implementation of DRL is the high computational requirements, which can lead to high power consumption and long training times. To overcome this problem, several strategies have been developed, including the use of lightweight architecture-based models and processing optimization with more efficient hardware such as GPUs (Graphics Processing Units) and TPUs (Tensor Processing Units).

The use of lightweight neural network-based models can reduce power consumption by up to 30% without sacrificing accuracy in robotic navigation tasks. In addition, by utilizing hardware acceleration such as GPUs and TPUs, DRL model training can be accelerated up to 10 times compared to conventional CPUs [38].

Hardware optimization, approaches such as quantization and pruning have also been used to reduce model complexity without sacrificing performance. The pruning method can reduce model size by up to 90%, which significantly improves execution efficiency in resource-constrained autonomous systems [39].

4.1.6. Hybrid Approaches: Combining DRL With Other AI Techniques

To improve the stability and generalization of DRL models, Hybrid Approaches that combine DRL with other artificial intelligence techniques have been developed. One method that is widely applied is the combination of DRL with Supervised Learning and Unsupervised Learning to improve learning efficiency and model stability.

The combination of DRL with Supervised Learning in the AlphaZero model allows agents to learn faster by utilizing experience from historical data. As a result, this model is able to outperform traditional reinforcement learning-based systems in games such as chess and Go with an increase in learning efficiency of up to 50% [40].

The combination of DRL with Unsupervised Learning, such as in the Curiosity-Driven Learning method, allows agents to explore the environment more efficiently without having to rely on explicit feedback. This method can increase exploration efficiency by up to 35%, especially in environments with sparsity or unclear rewards [41].

4.2. Key Challenges In DRL Implementation In Autonomous Systems

4.2.1. Training Stability In DRL

One of the main challenges in applying Deep Reinforcement Learning (DRL) to autonomous systems is training stability. DRL algorithms often experience slow and unstable convergence, especially in dynamic and complex environments. Unlike traditional machine learning methods, DRL relies on exploration and interaction with the environment to update the policy, which can cause large fluctuations in training performance. DRL models require up to 10 million iterations to achieve near-optimal performance in robotic navigation tasks, compared to supervised learning which can achieve similar results with only hundreds of thousands of iterations [42].

Sensitivity to hyperparameter selection is also a major challenge. Parameters such as learning rate, discount factor (γ), and exploration-exploitation balance have a significant impact on model performance. Mistakes in hyperparameter selection can lead to training instability, where the model fails to find the optimal strategy or gets stuck in a local optimum. That small changes in learning rate (e.g. from 0.001 to 0.0005) can cause model performance to drop by up to 35% in some DRL-based control tasks [43].

The balance between exploration and exploitation is a fundamental aspect of DRL training. If an agent exploits a learned strategy too much, it may miss opportunities to find better solutions. Conversely, excessive exploration without a clear strategy can lead to inefficient and time-consuming training.

Jurnal Teknik Informatika (JUTIF) Vol. 6, No. 3, Juni 2025, Page. 1189-1202
P-ISSN: 2723-3863 https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871 DOI: https://doi.org/10.52436/1.jutif.2025.6.3.4446

Algorithms such as ε-greedy, Upper Confidence Bound (UCB), and Thompson Sampling have been developed to address this challenge, but still face obstacles in highly complex and noisy environments [44].

Another challenge that complicates training stability is the reliance on simulation. Most DRL models are trained in simulated environments before being deployed in the real world. Although simulation allows for safer and more efficient experimentation, the difference between the simulated world and the real world (Sim-to-Real Gap) often causes a performance drop of up to 40% when the model is deployed in real conditions [45]. Transfer learning or domain adaptation methods are needed to address this gap.

4.2.2. Challenges Of DRL Model Generalization

Generalizability or the ability of DRL models to adapt to various environments is a major challenge in autonomous systems. One of the main problems in generalization is the Sim-to-Real Gap, where models trained in simulation often experience performance degradation in the real world. This is caused by differences in physical dynamics, such as sensor imperfections, friction, or environmental disturbances that cannot be fully simulated. DRL models applied directly from simulation to physical robots experienced a decrease in accuracy of up to 30% in navigation tasks due to these factors [46].

DRL models are often less robust to noise and disturbances. Inaccurate sensors, unexpected environmental disturbances, or changes in operational conditions can cause the model to lose its effectiveness. That small changes in sensory input (e.g., 5% noise) can decrease model performance by up to 20% in object recognition tasks. Therefore, strategies such as domain randomization and adversarial training are used to improve model robustness to environmental variations [47].

Another challenge is overfitting to the training environment, where the DRL model is too well-suited to a particular scenario and is less able to adapt when applied to a new environment. This phenomenon often occurs when the model is trained in an environment that is less varied or has high reward sparsity. DRL models that are only trained in one specific scenario experience a performance decrease of up to 50% when applied to an environment with little structural change. One solution to overcome this problem is unsupervised environment augmentation, where the model is trained with a variety of environments to improve generalization [48].

4.2.3. Data And Computational Resource Efficiency

DRL is known to require very large data to learn effectively. Unlike supervised learning that can learn from existing datasets, DRL requires direct interaction with the environment, which often results in inefficiencies in data collection. The Deep Q-Network (DQN) model requires more than 200 million interactions with the environment to achieve human performance in Atari games. This shows that without a good data management strategy, DRL training can be very expensive and inefficient. In addition to the large data requirements, DRL also has high computational costs. Training DRL models requires high-performance hardware such as GPUs or TPUs, which can be a barrier for many researchers and industries [49]. That training large-scale DRL models on Google TPUs can consume up to 2.5 kWh of energy per hour, which is equivalent to the power consumption of a small household. Optimizing model architecture, using distributed computing, and utilizing model compression techniques (quantization and pruning) are important steps in reducing the need for computing resources [47].

The problem of reward sparsity and credit assignment is also a challenge in many DRL applications. In many cases, agents only receive rewards after achieving a certain goal, making it difficult to determine which actions contributed most to the success. In a robot control task, an agent may only receive rewards after successfully completing a mission, but there is no direct feedback on which actions moved it closer or further away from the goal. To address this problem, several approaches

Vol. 6, No. 3, Juni 2025, Page. 1189-1202
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.3.4446

have been developed, such as reward shaping, hierarchical reinforcement learning (HRL), and intrinsic motivation that encourages agents to explore the environment more efficiently [50].

4.3. Strategies To Overcome Challenges In DRL For Autonomous Systems

4.3.1. Improving Training Stability

Training stability is a critical aspect in the development of Deep Reinforcement Learning (DRL) for autonomous systems. One of the main approaches used to improve stability is the implementation of more stable algorithms, such as Trust Region Policy Optimization (TRPO), Proximal Policy Optimization (PPO), and Soft Actor-Critic (SAC). TRPO controls policy changes by limiting the Kullback-Leibler (KL) divergence, thus preventing unwanted fluctuations in learning [51]. TRPO was shown to improve training stability by up to 35% compared to the vanilla policy gradient method [52].

PPO as an improvement of TRPO, optimizes the policy with a simpler and more efficient approach. PPO reduces variance by up to 20% compared to TRPO, while maintaining learning stability. Meanwhile, SAC introduces regulation entropy, which allows agents to continue exploring the environment without sacrificing exploitation performance [53].

Regularization and experience normalization strategies are also applied to reduce instability during training. Techniques such as batch normalization help reduce fluctuations in the reward distribution, while reward clipping is used to limit extreme reward values that can lead to explosive gradients. The use of reward clipping in robotics simulations can increase the convergence rate by 25% [54].

A more efficient exploration strategy is also a key factor in improving the stability of DRL. Techniques such as curiosity-driven exploration and intrinsic motivation are used to overcome suboptimal exploration. Curiosity-driven learning-based models, such as the Intrinsic Curiosity Module (ICM) allow agents to explore the environment more actively, with a 30% increase in exploration performance in DRL-based video games [50].

4.3.2. Improve Model Generalization

Generalizability is a major challenge in implementing DRL in the real world. One solution that has proven effective is domain randomization, which involves creating variations in the simulation environment to improve the model's resilience to changes in the real environment. Implementing domain randomization in robotics simulations, the success rate of model transfer from simulation to the real world increased by up to 40%, compared to models trained without environmental variations [24].

Meta-learning allows models to learn common patterns from various tasks, thereby accelerating adaptation to new environments. Algorithms such as Model-Agnostic Meta-Learning (MAML) have been shown to reduce model adaptation time by up to 50% in robotic navigation tasks. Continual learning, on the other hand, allows models to learn from experience incrementally without losing previous information [55]. Continual learning strategies can reduce catastrophic forgetting by up to 60% in multi-task learning scenarios [55].

To improve the model's resilience to environmental disturbances, adversarial training is applied in DRL training. This technique involves injecting noise or artificial disturbances into the training environment to improve the model's robustness to extreme conditions. Adversarial training in robotics increases resistance to sensory noise by up to 35%, which has a direct impact on improving the reliability of autonomous systems in the real world [50].

4.3.3. Data And Computational Efficiency Improvement

Data and computational resource efficiency are critical aspects in DRL implementation, given the very high training costs. One of the main strategies in improving data efficiency is the application of transfer learning and few-shot learning. By using a pre-trained model, the number of direct interactions with the environment can be significantly reduced. The application of transfer learning in the game of Go reduces the need for training data by up to 70%, without sacrificing model performance [53].

Model-based reinforcement learning (MBRL) is used to reduce the dependence on direct interactions with the environment. In this approach, predictive models are used to simulate experiences, thereby reducing the amount of real data required. Experimental results that the application of MBRL in the Deep Planning Network (PlaNet) is able to reduce the number of environmental interactions by up to 80%, while maintaining competitive performance [52].

Another approach used to improve efficiency is Hierarchical Reinforcement Learning (HRL), which divides complex tasks into smaller, easier-to-learn subtasks. By dividing the problem into multiple hierarchical levels, HRL can speed up the learning process and reduce data consumption. HRL can increase the convergence speed by up to 50% in robot navigation tasks [42].

To overcome the challenges related to computational costs, the use of parallel computing and distributed reinforcement learning is an effective solution. DRL running in a distributed system can utilize multiple GPUs or TPUs, which allows the model to learn faster by dividing the workload. a distributed RL framework, shows that with parallel computing, the training speed increases by up to 30 times compared to standard methods [55].

5. CONCLUSION

The application of Deep Reinforcement Learning (DRL) in autonomous systems improves operational efficiency in various sectors, including transportation, industrial robotics, and traffic management. Challenges such as training stability, model generalization, and the need for high data and computing resources still remain, strategies such as transfer learning, meta learning, and the use of model-based simulations have proven effective in addressing these issues. The integration of DRL with other technologies such as Computer Vision, IoT, and Edge Computing further strengthens its ability to deal with uncertain dynamic environments. DRL is expected to provide better and more reliable solutions for autonomous system applications in the future.

ACKNOWLEDGEMENT

The author would like to thank the Ministry of Education, Culture, Research, and Technology of the Republic of Indonesia for the financial support provided for this research. Furthermore, high appreciation is also conveyed to the Faculty of Engineering, Semarang State University, and all parties who have become research subjects in this research. The support and contributions provided are very valuable in completing the literature review on Deep Reinforcement Learning For Optimization Of Autonomous Systems: A Systematic Literature Review.

REFERENCES

[1] M. Panzer and B. Bender, "Deep reinforcement learning in production systems: a systematic literature review," *Int. J. Prod. Res.*, vol. 60, no. 13, pp. 4316–4341, 2022, doi: 10.1080/00207543.2021.1973138.
[2] S. Bauskar, "Leveraging AI for Intelligent Data Management in Multi-Cloud Database Architectures," 2025.
[3] Z. Tang, T. Li, D. Wu, J. Liu, and Z. Yang, "A Systematic Literature Review of Reinforcement Learning-based Knowledge Graph Research," *Expert Syst. Appl.*, vol. 238, no. March 2023, 2024, doi: 10.1016/j.eswa.2023.121880.
[4] A. Farooq and K. Iqbal, "A Survey of Reinforcement Learning for Optimization in Automation," *IEEE Int. Conf. Autom. Sci. Eng.*, pp. 2487–2494, 2024, doi: 10.1109/CASE59546.2024.10711718.
[5] G. N. P. Pratama, I. Hidayatulloh, H. D. Surjono, and T. Sukardiyono, "Enhance Deep

Reinforcement Learning with Denoising Autoencoder for Self-Driving Mobile Robot," *J. Robot. Control*, vol. 5, no. 3, pp. 667–676, 2024, doi: 10.18196/jrc.v5i3.21713.

[6] M. Khalid, L. Wang, K. Wang, N. Aslam, C. Pan, and Y. Cao, "Deep reinforcement learning-based long-range autonomous valet parking for smart cities," *Sustain. Cities Soc.*, vol. 89, no. November 2022, p. 104311, 2023, doi: 10.1016/j.scs.2022.104311.

[7] Y. Yi, "A Deep Reinforcement Learning-Based Speed Optimization System to Reduce Fuel Consumption and Emissions for Smart Cities," 2025.

[8] Y. Chen, C. Ji, Y. Cai, T. Yan, and B. Su, "Deep Reinforcement Learning in Autonomous Car Path Planning and Control: A Survey," pp. 1–29, 2024, [Online]. Available: http://arxiv.org/abs/2404.00340

[9] M. Panzer and B. Bender, "Deep reinforcement learning in production systems: a systematic literature review," *Int. J. Prod. Res.*, vol. 60, no. 13, pp. 4316–4341, 2022, doi: 10.1080/00207543.2021.1973138.

[10] M. Panzer, B. Bender, and N. Gronau, "Deep Reinforcement Learning In Production Planning And Control: A Systematic Literature Review," *Proc. Conf. Prod. Syst. Logist.*, pp. 535–345, 2021, doi: 10.15488/11238.

[11] K. Sivamayil, E. Rajasekar, B. Aljafari, S. Nikolovski, S. Vairavasundaram, and I. Vairavasundaram, "A Systematic Study on Reinforcement Learning Based Applications," *Energies*, vol. 16, no. 3, pp. 1–23, 2023, doi: 10.3390/en16031512.

[12] G. Icarte-Ahumada, J. Montoya, and Z. He, "Learning in multi-agent systems to solve scheduling problems: a systematic literature review," *Ingeniare*, vol. 32, pp. 1–14, 2024, doi: 10.4067/s0718-33052024000100214.

[13] I. Karisma and Y. Arifatul Fatimah, "Literature Review : Teknik Perancangan Tata Letak Fasilitas Gudang Pada Perushaan Manufaktur Yang Efisien," *Borobudur Eng. Rev.*, vol. 2, no. 1, pp. 12–22, 2022, doi: 10.31603/benr.6300.

[14] L. Fadilah, A. Dwi Wanti, S. Z. Latifah, and M. A. Yaqin, "Fadilah, Lailatul," *JACIS J. Autom. Comput. Inf. Syst.*, vol. 2, no. 02, pp. 116–129, 2022.

[15] A. Siski Rohana, L. Rakhmawati, L. Endah Cahya Ningrum, and Y. Sutoto Nugroho, "Studi Literatur: Pembelajaran Menggunakan Media Augmented Reality Di Sekolah Menengah Kejuruan," *J. Pendidik. Tek. Elektro*, vol. 11, no. 03, pp. 479–490, 2022.

[16] A. Ladha and D. Mohan, "Energy optimization in autonomous driving using deep reinforced learning and stochastic system modelling," vol. 6, no. 9, pp. 132–140, 2021, [Online]. Available: https://www.semanticscholar.org/paper/c5d3ccac1f00e9eece693f0abff1fc0b34e49a8a

[17] Z. Yu, W. Zheng, K. Zeng, R. Zhao, Y. Zhang, and M. Zeng, "Energy optimization management of microgrid using improved soft actor-critic algorithm," *Int. J. Renew. Energy Dev.*, vol. 13, no. 2, pp. 329–339, 2024, doi: 10.61435/ijred.2024.59988.

[18] S.-F. Wen and B. Katt, *Articial intelligence for system security assurance: A systematic literature review*, vol. 24, no. 1. Springer Berlin Heidelberg, 2024. doi: 10.1007/s10207-024-00959-0.

[19] Y. Bekkemoen, *Explainable reinforcement learning (XRL): a systematic literature review and taxonomy*, vol. 113, no. 1. Springer US, 2024. doi: 10.1007/s10994-023-06479-7.

[20] O. Gheibi, D. Weyns, and F. Quin, "Applying Machine Learning in Self-adaptive Systems," *ACM Trans. Auton. Adapt. Syst.*, vol. 15, no. 3, 2021, doi: 10.1145/3469440.

[21] *et al.*, "Machine Learning Methods for (Dis-)Assembly Sequence Planning - A Systematic Literature Review," *Int. J. Adv. Prod. Res.*, vol. 1, no. 1, pp. 83–98, 2024, doi: 10.62743/uad.8279.

[22] I. N. Yazid and E. Rachmawati, "Autonomous driving system using proximal policy optimization in deep reinforcement learning," *IAES Int. J. Artif. Intell.*, vol. 12, no. 1, pp. 422–431, 2023, doi: 10.11591/ijai.v12.i1.pp422-431.

[23] E. H. Sumiea *et al.*, "Deep deterministic policy gradient algorithm: A systematic review," *Heliyon*, vol. 10, no. 9, p. e30697, 2024, doi: 10.1016/j.heliyon.2024.e30697.

[24] T. Zhang and H. Mo, "Reinforcement learning for robot research: A comprehensive review and open issues," *Int. J. Adv. Robot. Syst.*, vol. 18, no. 3, pp. 1–22, 2021, doi: 10.1177/17298814211007305.

[25] P. L. Donti and J. Z. Kolter, "Machine Learning for Sustainable Energy Systems," *Annu. Rev.*

Environ. Resour., vol. 46, pp. 719–747, 2021, doi: 10.1146/annurev-environ-020220-061831.
[26] J. Zhang and J. Zhang, "Artificial Intelligence Applied on Traffic Planning and Management for Rail Transport: A Review and Perspective," *Discret. Dyn. Nat. Soc.*, vol. 2023, 2023, doi: 10.1155/2023/1832501.
[27] I. Miadowicz, D. M. Quinto, and M. Felderer, "Self-X Characterization of Autonomous Systems : A Systematic Literature Review," vol. 1, no. 1, 2024.
[28] F. Den Hengst, E. M. Grua, A. El Hassouni, and M. Hoogendoorn, *Reinforcement learning for personalization: A systematic literature review*, vol. 3, no. 2. 2020. doi: 10.3233/DS-200028.
[29] N. M. Ashraf, R. R. Mostafa, R. H. Sakr, and M. Z. Rashad, "Optimizing hyperparameters of deep reinforcement learning for autonomous driving based on whale optimization algorithm," *PLoS One*, vol. 16, no. 6 June, pp. 1–24, 2021, doi: 10.1371/journal.pone.0252754.
[30] C. Ma, A. Li, Y. Du, H. Dong, and Y. Yang, "Efficient and scalable reinforcement learning for large-scale network control," *Nat. Mach. Intell.*, vol. 6, no. September, 2024, doi: 10.1038/s42256-024-00879-7.
[31] S. Shisiali, E. Amuti, O. Calvins, and O. James, "A Systematic Literature Review of Path-Planning Strategies for Robot A Systematic Literature Review of Path-Planning Strategies for Robot Navigation in Unknown Environment," vol. 4523, no. September, pp. 154–177, 2023.
[32] M. Noaeen *et al.*, "Reinforcement learning in urban network traffic signal control: A systematic literature review," *Expert Syst. Appl.*, vol. 199, no. March 2021, p. 116830, 2022, doi: 10.1016/j.eswa.2022.116830.
[33] S. Krishnamoorthi and G. K. Shyam, "Design of Recommendendation Systems Using Deep Reinforcement Learning – Recent Advancements and Applications," *J. Theor. Appl. Inf. Technol.*, vol. 102, no. 7, pp. 2908–2923, 2024.
[34] F. AlMahamid and K. Grolinger, "Autonomous Unmanned Aerial Vehicle navigation using Reinforcement Learning: A systematic review," *Eng. Appl. Artif. Intell.*, vol. 115, pp. 0–2, 2022, doi: 10.1016/j.engappai.2022.105321.
[35] M. Hu, J. Zhang, L. Matkovic, T. Liu, and X. Yang, "Reinforcement learning in medical image analysis: Concepts, applications, challenges, and future directions," *J. Appl. Clin. Med. Phys.*, vol. 24, no. 2, pp. 1–21, 2023, doi: 10.1002/acm2.13898.
[36] H. Zhao *et al.*, "Combination optimization method of grid sections based on deep reinforcement learning with accelerated convergence speed," *Front. Energy Res.*, vol. 11, no. October, pp. 1–9, 2023, doi: 10.3389/fenrg.2023.1269854.
[37] A. A. Volk *et al.*, "AlphaFlow: autonomous discovery and optimization of multi-step chemistry using a self-driven fluidic lab guided by reinforcement learning," *Nat. Commun.*, vol. 14, no. 1, pp. 1–16, 2023, doi: 10.1038/s41467-023-37139-y.
[38] K. Mitsopoulos, S. Somers, J. Schooler, C. Lebiere, P. Pirolli, and R. Thomson, "Toward a Psychology of Deep Reinforcement Learning Agents Using a Cognitive Architecture," *Top. Cogn. Sci.*, vol. 14, no. 4, pp. 756–779, 2022, doi: 10.1111/tops.12573.
[39] K. Athania Purba and T. Dewayanto, "Penerapan Artificial Intelligence, Machine Learning Dan Deep Learning Pada Kurikulum Akuntansi-a Systematic Literature Review," *Diponegoro J. Account.*, vol. 12, no. 3, pp. 1–15, 2023, [Online]. Available: http://ejournal-s1.undip.ac.id/index.php/accounting
[40] H. Hoppe, T. Enders, Q. Cappart, and M. Schiffer, "Global Rewards in Multi-Agent Deep Reinforcement Learning for Autonomous Mobility on Demand Systems," *Proc. Mach. Learn. Res.*, vol. 242, pp. 260–272, 2024.
[41] P. Macaluso, "Deep Reinforcement Learning for Autonomous Systems," *Master Thesis, Dep. Control Comput. Eng. Univ. Politec. DI TORINO*, vol. 44, no. 5, pp. 4125–4133, 2020.
[42] F. Jerald, S. Murugaanandam, S. K. Thillaigovindan, P. Prasanna, and M. Vimala, "Future Trends of AI-Based Modelling in the Context of Smart Devices and Smart Systems," *Nanotechnol. Perceptions*, vol. 20, no. S7, pp. 14–26, 2024, doi: 10.62441/nano-ntp.v20iS7.2.
[43] H. Zhang and C. Zhang, "Low-carbon power demand forecasting models for the performance optimization of new energy robotics systems," no. December 2024, pp. 341–352, 2025.
[44] A. Tammewar *et al.*, "Improving the Performance of Autonomous Driving through Deep Reinforcement Learning," *Sustain.*, vol. 15, no. 18, 2023, doi: 10.3390/su151813799.

Vol. 6, No. 3, Juni 2025, Page. 1189-1202
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.3.4446

[45] R. Tang *et al.*, "A literature review of Artificial Intelligence applications in railway systems," *Transp. Res. Part C Emerg. Technol.*, vol. 140, no. April, 2022, doi: 10.1016/j.trc.2022.103679.
[46] M. Aria, "Algoritma Perencanaan Jalur Kendaraan Otonom di Lingkungan Perkotaan dari Sudut Pandang Filosofi Kuhn dan Filosofi Popper," *Telekontran J. Ilm. Telekomun. Kendali dan Elektron. Terap.*, vol. 7, no. 2, 2020, doi: 10.34010/telekontran.v7i2.2627.
[47] D. Arya and S. Pratama, "Pengembangan Kontrol Adaptif untuk Kendaraan Otonom dengan Studi Kasus pada Mobil Elektrik Berbasis Deep Reinforcement Learning," no. 2019, pp. 234–239, 2024.
[48] S. Abdallaoui, H. Ikaouassen, A. Kribèche, A. Chaibet, and E. Aglzim, "Advancing autonomous vehicle control systems: An in-depth overview of decision-making and manoeuvre execution state of the art," *J. Eng.*, vol. 2023, no. 11, pp. 1–16, 2023, doi: 10.1049/tje2.12333.
[49] A. M. Geske, D. M. Herold, and S. Kummer, "Artificial intelligence as a driver of efficiency in air passenger transport: A systematic literature review and future research avenues," *J. Air Transp. Res. Soc.*, vol. 3, no. June, p. 100030, 2024, doi: 10.1016/j.jatrs.2024.100030.
[50] D. D. Cahyo, "Simulasi Self-Driving Car dengan Reinforcement Learning dan NeuroEvolution of Augmenting Topologies (NEAT)," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 3, pp. 1752–1761, 2022, doi: 10.35957/jatisi.v9i3.2154.
[51] D. Xiang, "Reinforcement learning in autonomous driving," *Appl. Comput. Eng.*, vol. 48, no. 1, pp. 17–23, 2024, doi: 10.54254/2755-2721/48/20241072.
[52] D. Pembimbing and A. W. S. T, *SIMULASI DEEP LEARNING PADA AUTONOMOUS VEHICLE MENGGUNAKAN NVIDIA JETBOT UNTUK*. 2022.
[53] S. H. Ashwin and R. Naveen Raj, "Deep reinforcement learning for autonomous vehicles: lane keep and overtaking scenarios with collision avoidance," *Int. J. Inf. Technol.*, vol. 15, no. 7, pp. 3541–3553, 2023, doi: 10.1007/s41870-023-01412-6.
[54] A. Gupta, A. Anpalagan, L. Guan, and A. S. Khwaja, "Deep learning for object detection and scene perception in self-driving cars: Survey, challenges, and open issues," *Array*, vol. 10, no. December 2020, p. 100057, 2021, doi: 10.1016/j.array.2021.100057.
[55] L. Ge, X. Zhou, Y. Li, and Y. Wang, "Deep reinforcement learning navigation via decision transformer in autonomous driving," *Front. Neurorobot.*, vol. 18, 2024, doi: 10.3389/fnbot.2024.1338189.