Tech-Sphere Journal of Pure and Applied Sciences (TSJPAS)
A Subsidiary of Tech-Sphere Multidisciplinary International Journal (TSMIJ)
Caleb et al. Vol 2, Issue 1, 2025 Publication Edition
ISSN: 3092-9598

A Comprehensive Survey of Advances in Deep Reinforcement Learning for Autonomous Robotics Applications

Research Article https://stem.techspherejournal.com

Article Info
Revised Date: 15th October 2025
Accepted Date: 23th October 2025
Published Date: 1st November 2025

Author Details
Akinsiku Ayokunle Michael
Department of Computer Science, The Federal Polytechnic Ado-Ekiti, Ekiti State, Nigeria.
****Corresponding author's email***: akinsiku_am@fedpolyado.edu.ng
DOI: https://doi.org/10.5281/zenodo.17497477

Keywords
Deep Reinforcement Learning
Autonomous Robotics Intelligence
Robotics Applications
Reinforcement Learning
Machine Learning for Robotics

ABSTRACT

This survey provides a comprehensive overview of the recent advancements in Deep Reinforcement Learning (DRL) for autonomous robotics applications. It begins by introducing the fundamental concepts of DRL and its significance in enabling intelligent and adaptive robotic behavior. The core of this paper is a structured review of state-of-the-art DRL algorithms, categorized into value-based, policy-based, and actor-critic methods, and their innovative applications across a spectrum of robotic domains. We delve into key areas such as manipulation, locomotion, navigation, and human-robot interaction, highlighting the transformative impact of DRL in solving complex, high-dimensional control problems. Furthermore, the survey critically examines the persistent challenges in the field, including the simulation-to-reality gap, sample efficiency, safety, and reward engineering. Finally, we discuss emerging trends and promising future research directions that are poised to shape the next generation of autonomous robotic systems.

1 Introduction

1.1 Motivation and a Vision for Autonomous Robotics

The dawn of the 21st century has witnessed the rapid evolution of autonomous robots from the realm of science fiction to tangible, transformative tools poised to reshape numerous sectors of our society. The vision for a future interwoven with intelligent, autonomous systems is no longer a distant dream but an impending reality. In manufacturing, the paradigm is shifting from rigid, pre-programmed robotic arms to versatile humanoid robots capable of intricate assembly, quality control, and seamless collaboration with human workers.[1][2] This evolution promises not only heightened efficiency and precision but also the creation of highly flexible and responsive "smart factories".[3]

In the healthcare sector, the potential for autonomous robots is equally profound. From surgical robots that enhance a surgeon's precision, leading to minimally invasive procedures and faster patient recovery times, to service robots that alleviate the logistical burdens on healthcare staff by handling tasks like supply delivery and room sanitization, the impact is multifaceted.[4][5] Looking ahead, advancements may enable robots to perform routine diagnostic procedures, deliver targeted medications, and even provide companionship and assistance in elder care.[6][7]

The frontier of space exploration is also being redefined by robotics. Autonomous rovers and landers are already our proxies on distant celestial bodies, conducting scientific investigations in environments too hazardous for humans.[8][9] The future of space exploration hinges on the development of even more sophisticated autonomous systems capable of building extraterrestrial habitats, mining resources, and making independent decisions during deep-space missions where communication delays are significant.[10][11][12] The realization of this ambitious vision across all sectors is critically dependent on the advancement of learning techniques that can imbue robots with the intelligence, adaptability, and autonomy required to operate in complex, unstructured environments.[13][14][15][16].

1.2 The Emergence of Deep Reinforcement Learning in Robotics

For decades, the control of robotic systems has been dominated by traditional, model-based approaches. These methods, while effective in structured and predictable environments, often falter when faced with the complexities and uncertainties of the real world. Their reliance on precise mathematical models of the robot and its environment makes them brittle to unforeseen variations and computationally intensive to design and implement for high-dimensional systems.

The advent of deep learning, a subfield of artificial intelligence, has offered a powerful new toolkit for robotics.[13] By leveraging deep neural networks, robots can learn complex tasks directly from large datasets, enabling them to recognize patterns, perceive their surroundings with greater acuity, and make more informed decisions. However, it is the fusion of deep learning with reinforcement learning (RL) that has unlocked a new paradigm for creating truly adaptive and intelligent robots.

Deep Reinforcement Learning (DRL) allows a robot to learn optimal behaviors through a process of trial and error, guided by a reward signal. Instead of being explicitly programmed, the robot, or "agent," learns a "policy"—a mapping from states to actions—that maximizes its cumulative reward. This approach is particularly well-suited for robotics, where the dynamics of interaction with the physical world are often too complex to model accurately. DRL empowers robots to learn sophisticated skills, such as dexterous manipulation, agile locomotion, and autonomous navigation, by directly interacting with their environment. This data-driven learning process circumvents the need for handcrafted models and enables robots to continuously improve their performance through experience. The remarkable successes of DRL in a variety of domains have established it as a cornerstone for the future development of autonomous robotic systems.

1.3 Research Questions

The rapid proliferation of research in DRL for robotics has created a vast and often fragmented body of literature. While several surveys have provided valuable snapshots of the field, a comprehensive and up-to-date overview that bridges the gap between foundational concepts, diverse applications, and persistent challenges is needed. This survey aims to provide such a holistic perspective, catering to both newcomers seeking a structured introduction and seasoned researchers looking for a broad view of the current landscape.

The primary contributions of this paper are threefold:

1. A Unified Framework: We present a structured taxonomy of DRL algorithms, categorizing them into value-based, policy-based, and actor-critic methods, and contextualize them within the specific demands of robotic applications.
2. A Broad Application Spectrum: We offer an in-depth review of the application of DRL across key robotic domains, including manipulation, locomotion, navigation, and human-robot interaction, showcasing the versatility of DRL in solving a wide array of control problems.
3. A Critical Analysis of Challenges and Future Directions: We go beyond a mere summary of successes to critically examine the significant hurdles that remain, such as the sim-to-real gap, sample efficiency, and safety. Furthermore, we identify and discuss emerging trends and promising avenues for future research that will be instrumental in driving the next wave of innovation in autonomous robotics.

This survey is distinguished by its comprehensive scope, its focus on the practical application of DRL in robotics, and its forward-looking perspective on the challenges and opportunities that lie ahead.

Tech-Sphere Journal of Pure and Applied Sciences (TSJPAS)
A Subsidiary of Tech-Sphere Multidisciplinary International Journal (TSMIJ)
Caleb et al. Vol 2, Issue 1, 2025 Publication Edition
ISSN: 3092-9598

1.4 A Roadmap for the Reader

To guide the reader through this comprehensive survey, the paper is organized as follows: Section 2 provides a foundational understanding of the core principles of Deep Reinforcement Learning. Section 3 delves into the application of DRL to a variety of core robotic capabilities, offering a detailed look at recent advancements. Section 4 presents a critical discussion of the key challenges and open problems that currently face the field. Section 5 explores the exciting future directions and emerging trends that are shaping the next generation of autonomous robots. Finally, Section 6 concludes with a summary of the key findings and a forward-looking perspective on the future of DRL in robotics.

2 Fundamentals of Deep Reinforcement Learning

2.1 Reinforcement Learning Preliminaries

Reinforcement Learning (RL) is a paradigm of machine learning where an intelligent agent learns to make optimal decisions through interaction with an environment.[1] The fundamental goal of the agent is to maximize a cumulative reward signal over time.[1][2] This learning process is formalized through several key concepts:

1. **Agent:** The agent is the learner and decision-maker that interacts with the environment.[2][3]
2. Environment: This represents the external system that the agent interacts with. It provides feedback to the agent in the form of states and rewards.[2][3]
3. **State (S):** A state is a snapshot of the environment at a particular point in time, containing all the necessary information for the agent to make a decision.[2][4] The set of all possible states is known as the state space.[4]
4. **Action (A):** An action is a choice made by the agent from a set of available options, which in turn influences the state of the environment.[2][4]
5. **Reward (R):** A reward is a scalar feedback signal that the agent receives from the environment after performing an action.[2][3] It indicates the immediate desirability of an action taken from a particular state.[5] The agent's ultimate objective is to maximize the total accumulated reward over the long run.[3][5]
6. **Policy (π):** The policy is the agent's strategy or behavior, which maps states to actions.[2][5] A policy can be deterministic, meaning it always chooses the same action for a given state, or stochastic, where it outputs a probability distribution over actions.[3]

The interaction between the agent and the environment unfolds in a continuous loop: the agent observes the current state, selects an action based on its policy, and as a result, the environment transitions to a new state and provides a reward to the agent.[2][4] This cycle of observation, action, and reward allows the agent to learn and refine its policy over time. Figure 1 illustrates the fundamental interaction loop between an agent and its environment, which is the core concept of reinforcement learning.

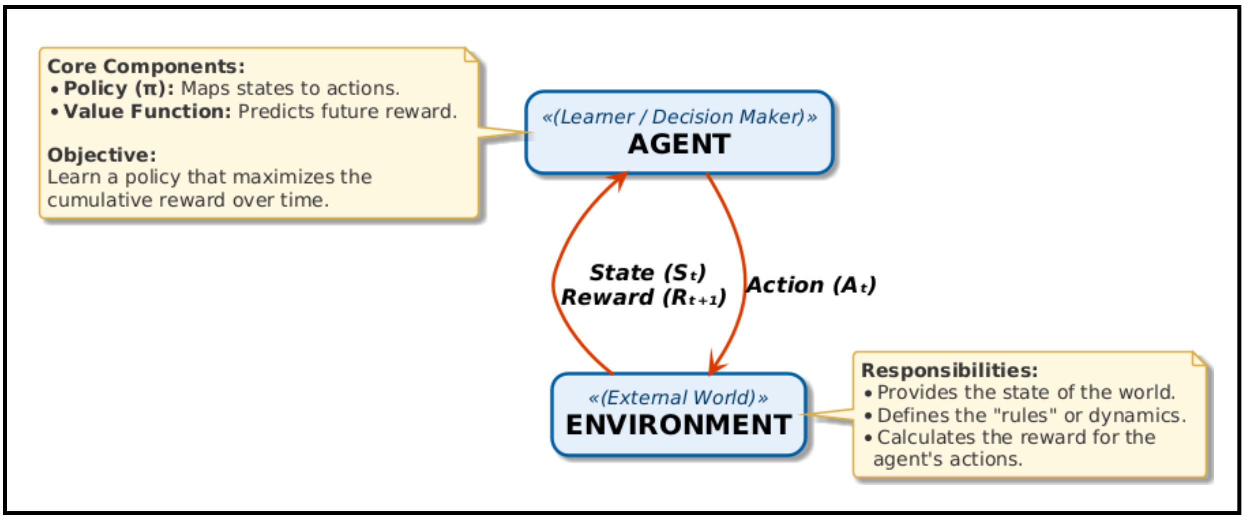

Figure 1: The Reinforcement Learning Framework

Tech-Sphere Journal of Pure and Applied Sciences (TSJPAS)
A Subsidiary of Tech-Sphere Multidisciplinary International Journal (TSMIJ)
Caleb et al. Vol 2, Issue 1, 2025 Publication Edition
ISSN: 3092-9598

2.2 The Role of Deep Learning

In many real-world robotics applications, the state and action spaces are vast and continuous, making it infeasible to represent value functions or policies using traditional tabular methods.[6] This is where deep learning plays a transformative role. Deep neural networks, with their ability to approximate complex, non-linear functions, are employed as powerful function approximators.[7][8]

By using deep neural networks, we can approximate the value function, which estimates the expected long-term reward from a given state, or the policy, which directly maps states to actions. This approach, known as Deep Reinforcement Learning (DRL), allows RL algorithms to scale to high-dimensional problems that were previously intractable.[8][9] For instance, in robotic manipulation, a neural network can take raw pixel data from a camera as input (representing the state) and output the appropriate motor commands (the action).[6] This ability to learn directly from raw sensory inputs, such as images, without the need for manual feature engineering is a key advantage of DRL.[10].

2.3 A Taxonomy of Deep Reinforcement Learning Algorithms

DRL algorithms can be broadly categorized into three main families: value-based, policy-based, and actor-critic methods.

Value-based methods aim to learn the optimal action-value function, often referred to as the Q-function. The Q-function, Q(s, a), represents the expected cumulative reward for taking action 'a' in state 's' and following the optimal policy thereafter. Once the optimal Q-function is learned, the policy is simply to select the action with the highest Q-value for any given state.[11]

A seminal value-based DRL algorithm is the Deep Q-Network (DQN). DQN utilizes a deep neural network to approximate the Q-function.[12] To stabilize the learning process, DQN introduced two key techniques:

a. **Experience Replay**: The agent's experiences (state, action, reward, next state) are stored in a replay memory. During training, mini-batches of experiences are randomly sampled from this memory, which helps to break the correlation between consecutive samples and improves data efficiency.[11]
b. **Target Network:** A separate target network, which is a periodically updated copy of the main Q-network, is used to generate the target Q-values for the Bellman equation. This helps to prevent instabilities that can arise from using the same network to estimate both the current and target Q-values.[13]

Several extensions to the original DQN have been proposed to further improve its performance, such as Double DQN, which addresses the overestimation of Q-values, and Dueling DQN, which separates the estimation of the state value function and the advantage function for each action.[7][13]

In contrast to value-based methods, policy-based methods directly learn a parameterized policy that maps states to a distribution over actions.[14][15] The policy parameters are updated by performing gradient ascent on an objective function that represents the expected cumulative reward.[15]

A foundational policy-based algorithm is REINFORCE. It updates the policy parameters in the direction that makes actions leading to higher returns more likely. While intuitive, REINFORCE can suffer from high variance in the gradient estimates, leading to unstable training.[14][16]

To address the stability issues of vanilla policy gradient methods, more advanced algorithms have been developed:

a. **Trust Region Policy Optimization (TRPO):** TRPO constrains the size of the policy update at each iteration to prevent destructively large changes. It ensures that the new policy does not deviate too far from the old one, leading to more stable and monotonic improvements.[17]
b. **Proximal Policy Optimization (PPO):** PPO offers a simpler yet effective alternative to TRPO. It uses a clipped surrogate objective function to restrict the policy update, making it easier to implement while achieving comparable or even better performance.[16][17] PPO has become a popular choice for many DRL applications due to its balance of performance and ease of use.[15][16]

Actor-critic methods combine the strengths of both value-based and policy-based approaches. They consist of two main components: an actor and a critic.[18]

a. The actor is a policy network that decides which action to take.[19]

ISSN: 3092-9598

b. The critic is a value network that evaluates the action taken by the actor by estimating the value function.[19]

The critic's feedback is then used to update the actor's policy, leading to a more stable and efficient learning process.[20]

Prominent actor-critic algorithms include:

a. Asynchronous Advantage Actor-Critic (A3C): A3C uses multiple parallel agents, each with its own copy of the actor-critic network, to explore the environment simultaneously. This asynchronous nature helps to decorrelate the agents' experiences and accelerates learning.[18][20]
b. Deep Deterministic Policy Gradient (DDPG): DDPG is an off-policy actor-critic algorithm that is well-suited for continuous action spaces.[18][21] It utilizes a deterministic actor policy and incorporates techniques from DQN, such as experience replay and target networks.[18][22]
c. Soft Actor-Critic (SAC): SAC is a state-of-the-art off-policy actor-critic method that incorporates an entropy maximization term into its objective function. This encourages the policy to be as random as possible while still maximizing the cumulative reward, which leads to improved exploration and robustness.[19][22]

Figure 2 presents a hierarchical chart of a clear visual classification of the main families of Deep Reinforcement Learning algorithms, helping the reader to structure the different approaches.

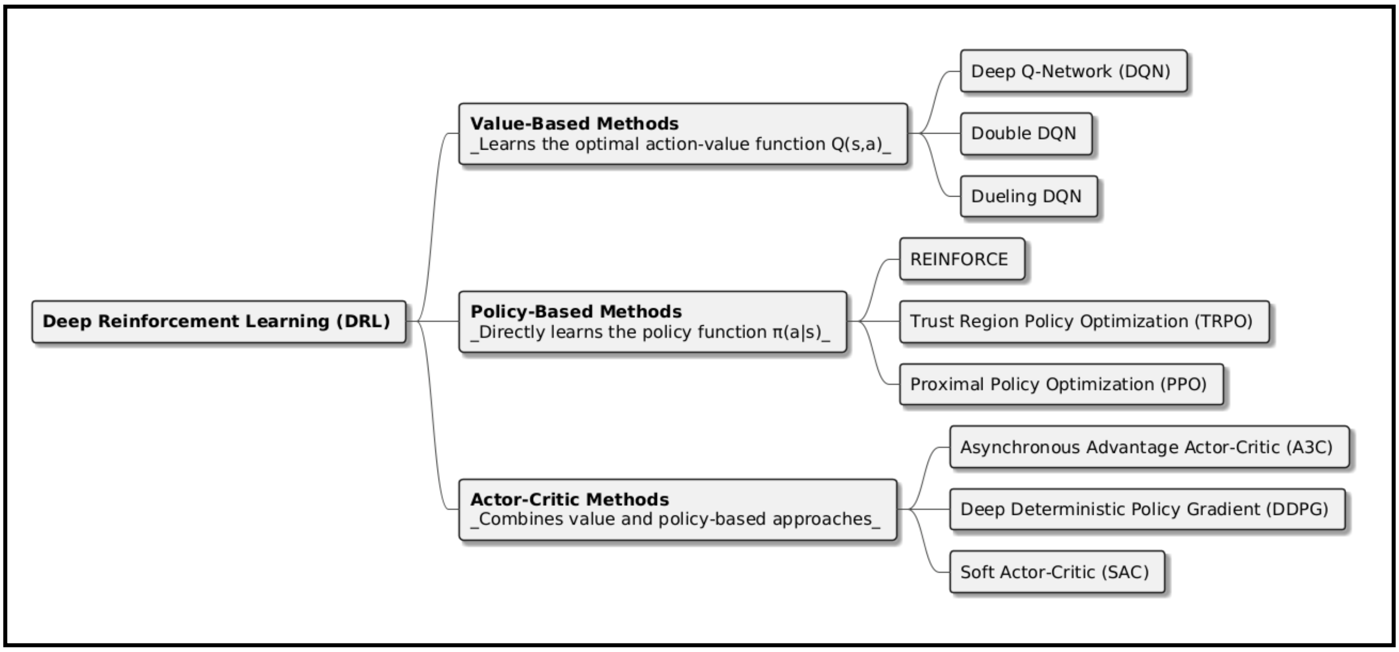

Figure 2: A Taxonomy of DRL Algorithms

3 Deep Reinforcement Learning for Core Robotic Capabilities

Deep Reinforcement Learning (DRL) has emerged as a powerful paradigm for enabling a wide array of robotic functionalities. By learning through interaction and experience, DRL allows robots to acquire complex skills that are often difficult to hand-engineer. This section provides a comprehensive review of the application of DRL to core robotic capabilities, highlighting the significant advancements and transformative potential of this approach.

3.1 Robotic Manipulation

Robotic manipulation, a cornerstone of robotics, involves the control of a robotic arm and hand to interact with and alter the environment. DRL has proven to be particularly effective in this domain, where the high dimensionality of the state and action spaces, coupled with complex contact dynamics, presents significant challenges for traditional control methods.[1]

DRL has revolutionized robotic grasping by enabling robots to learn robust grasping policies directly from sensor data.[1] Instead of relying on predefined grasp poses, DRL-based approaches can learn to grasp a wide variety of objects in unstructured environments.[2] Early successes in this area often involved training in simulated environments, which allows for the collection of vast amounts of data.[3] One of the key challenges in this area is bridging the "sim-to-real" gap, where policies trained in simulation are transferred to physical robots. Techniques such as domain randomization, where the physical properties of the simulated environment are varied during training, have been instrumental in improving the generalization of learned policies to the real world.[3]

Beyond simple grasping, DRL is also being used to tackle more complex in-hand manipulation tasks.[4][5] These tasks, which involve repositioning an object within the robot's hand, are notoriously difficult due to the complex contact interactions between the fingers and the object.[6] DRL algorithms, often combined with tactile sensing, can learn the intricate finger gaiting and coordination required for such dexterous manipulation.[6]

The application of DRL extends to even more complex manipulation tasks, such as assembly and tool use. These tasks require not only precise manipulation but also sequential decision-making. DRL can be used to learn policies that can perform a sequence of actions to complete an assembly task, such as inserting a peg into a hole or screwing a nut onto a bolt. Similarly, DRL has been used to enable robots to learn how to use tools, a hallmark of intelligent behavior. By providing the robot with a reward for completing a task, such as hammering a nail, the DRL agent can learn the necessary sequence of motions to effectively wield the tool.[5].

3.2 Locomotion

DRL has also made significant strides in the field of robotic locomotion, enabling legged, aerial, and underwater robots to navigate complex and dynamic environments with unprecedented agility and robustness.

For legged robots, DRL has been instrumental in developing controllers that can generate dynamic and agile gaits.[7] By training in simulation, DRL algorithms can learn to control the many degrees of freedom of a legged robot to achieve stable and efficient locomotion over a variety of terrains.[8] The resulting policies can often be directly transferred to physical robots, enabling them to walk, run, and even recover from falls.[7][9]

Recent research has also explored the use of DRL to enable quadrupedal robots to perform bipedal locomotion, a challenging task due to the reduced stability.[10] By designing bio-inspired estimators and reward functions, DRL controllers can learn to maintain balance and achieve robust bipedal walking on various terrains.[10] While DRL has seen significant success in quadrupedal locomotion, advancements in bipedal robots have been comparatively limited due to their inherent instability and mechanical complexity.[11] However, DRL continues to be a promising approach for developing unified frameworks that can handle a wide range of locomotion tasks for bipedal robots.[11][12]

DRL is also being applied to the control and navigation of aerial and underwater robots. For unmanned aerial vehicles (UAVs), DRL can be used to learn policies for tasks such as attitude control and autonomous navigation in unknown outdoor environments.[13] In the domain of autonomous underwater vehicles (AUVs), DRL has been used to learn policies for mapless navigation and object tracking.[14][15][16] The ability to learn directly from sensor data is particularly advantageous in these challenging environments where accurate models are often unavailable.

3.3 Autonomous Navigation

Autonomous navigation is a fundamental capability for mobile robots, and DRL has emerged as a powerful tool for developing end-to-end navigation systems.

DRL-based approaches can learn to map raw sensor inputs, such as laser scans or camera images, directly to motor commands, enabling robots to navigate in unknown environments without the need for an explicit map.[17][18][19] These end-to-end navigation strategies can handle dynamic obstacles and are often more robust than traditional map-based methods.[20][21] DRL algorithms, such as the Deep Deterministic Policy Gradient (DDPG), have been used to train mobile robots to avoid obstacles based on range sensor readings.[22] Furthermore, DRL has been employed to develop navigation strategies that can be trained in simulation and then successfully transferred to real-world robots.[18]

In the context of autonomous driving, DRL is being used to address the complex decision-making challenges that arise in dynamic traffic scenarios.[23][24] DRL can be used to learn driving policies that can handle tasks such as lane changing, merging, and navigating intersections.[25][26] By integrating risk assessment into the reward structure, DRL frameworks can lead to safer and more efficient driving behaviors.[25] Moreover, DRL-based approaches have been shown to outperform traditional rule-based methods in a variety of driving tasks.[24]

3.4 Human-Robot Interaction and Collaboration

DRL is also playing an increasingly important role in shaping the way robots interact and collaborate with humans.
In the field of social robotics, DRL is being used to develop robots that can learn socially appropriate behaviors through interaction with humans.[27] By designing reward functions that capture the nuances of social interaction, DRL can enable robots to learn behaviors such as joining a group of people in a socially acceptable manner.[28] Furthermore, DRL is being used to enhance human-robot interaction by adapting the robot's behavior based on the user's engagement and emotional state.[29]
In collaborative robotics, where humans and robots work together in a shared workspace, DRL is being used to enable safe and efficient collaboration.[30] DRL can be used to train cobots to perform tasks such as pick-and-place while adapting to changes in the environment and the presence of human co-workers.[31][32] By incorporating safety constraints directly into the DRL framework, it is possible to train cobots that can anticipate and mitigate potential hazards while maintaining operational efficiency.[33]

4 Key Challenges and Open Problems

Despite the remarkable progress of Deep Reinforcement Learning (DRL) in robotics, several significant hurdles remain before its widespread adoption in real-world applications. This section provides a critical analysis of these key challenges and open problems, highlighting the areas that require further research and innovation.

4.1 The Simulation-to-Reality (Sim-to-Real) Gap

One of the most significant challenges in robotic DRL is the "sim-to-real" gap, which refers to the discrepancy between the simulated environment where policies are often trained and the complexities of the real world.[1][2] This gap arises from various factors, including inaccuracies in physical modeling (e.g., friction, contact dynamics), sensor noise, and unforeseen environmental variations.[3] Policies trained exclusively in simulation often fail when deployed on physical robots because they have not been exposed to the nuances of real-world physics and sensory feedback.[1]
To bridge this gap, researchers have developed several techniques:

a. Domain Randomization: This technique involves randomizing various aspects of the simulation during training, such as lighting conditions, textures, object masses, and friction coefficients.[2][3] By exposing the DRL agent to a wide range of simulated environments, domain randomization helps to create policies that are more robust and can generalize to the real world, which is effectively treated as just another variation of the simulation.[2][4]
b. Transfer Learning: This approach leverages the knowledge gained in simulation to expedite learning in the real world.[5][6][7] A common strategy is to pre-train a policy in simulation and then fine-tune it on a smaller amount of real-world data.[5] This significantly reduces the time and cost associated with real-world training.[5].

4.2 Sample Efficiency

DRL algorithms are notoriously data-hungry, often requiring millions of interactions with the environment to learn an effective policy.[8][9] This issue of low sample efficiency is a major bottleneck for real-world robotics, where data collection can be time-consuming, expensive, and may lead to wear and tear on the robot.[8]
Several approaches are being explored to improve sample efficiency:

a. Model-Based Reinforcement Learning: Unlike model-free methods that learn a policy directly, model-based RL aims to learn a model of the environment's dynamics.[10][11][12] This learned model can then be used

ISSN: 3092-9598

to simulate experiences, allowing the agent to "plan" and learn a policy with significantly fewer real-world interactions.[13][14] However, the performance of model-based methods is heavily dependent on the accuracy of the learned model.[11]

b. Off-Policy Learning: Off-policy algorithms, such as those that utilize experience replay, can reuse past experiences to update the current policy.[15][16] This allows for more efficient use of the collected data compared to on-policy methods, which discard old experiences after each policy update.

4.3 Safety, Robustness, and Generalization

Ensuring the safe and reliable operation of DRL-powered robots is paramount, especially in environments shared with humans.[17][18] The exploratory nature of DRL, which involves trial and error, can lead to unpredictable and potentially unsafe behaviors during the learning process.[17]

Key considerations in this area include:

a. Safety: Developing DRL algorithms with formal safety guarantees is an active area of research. This includes methods that can ensure the robot remains within a predefined safe region of the state space and avoids catastrophic failures.

b. Robustness: A robust DRL policy should be able to withstand perturbations and unforeseen variations in the environment.[18][19] This is particularly crucial for safety-critical applications like autonomous driving, where small sensor errors or adversarial attacks could have severe consequences.[19][20]

c. Generalization: The ability of a learned policy to generalize to new, unseen situations is a critical measure of its real-world utility. DRL agents often overfit to the specific conditions of their training environment, leading to poor performance when faced with novel scenarios.[21].

4.4 Reward Function Design and Engineering

The performance of a DRL agent is highly sensitive to the design of its reward function.[22][23] Crafting a reward function that accurately captures the desired behavior can be a challenging and iterative process. A poorly designed reward function can lead to unintended and suboptimal behaviors, a phenomenon known as "reward hacking".[22] For complex tasks, it is often difficult to specify a dense reward signal that provides continuous feedback to the agent. This has led to the development of techniques like:

- Inverse Reinforcement Learning (IRL): IRL aims to infer the underlying reward function from expert demonstrations.[24] Instead of manually specifying the reward, the agent learns it by observing an expert performing the task.[25][26]

4.5 Reward Function Design and Engineering

A fundamental challenge in reinforcement learning is the trade-off between exploration and exploitation.[27][28] The agent must balance exploring its environment to discover new, potentially better, strategies with exploiting its current knowledge to maximize immediate rewards.[27][29] In complex, high-dimensional robotic tasks with sparse rewards, naive exploration strategies, such as random action selection, are often insufficient. Developing more sophisticated and efficient exploration techniques is crucial for enabling DRL agents to solve challenging long-horizon tasks. The imbalance between exploration and exploitation can lead to the agent getting stuck in local optima (due to excessive exploitation) or failing to converge to a good policy (due to excessive exploration).[30]

5 Future Directions and Emerging Trends

The field of Deep Reinforcement Learning (DRL) for robotics is rapidly evolving, with researchers continuously pushing the boundaries of what is possible. This section explores the most promising future directions and emerging trends that are set to shape the next generation of autonomous robots.

5.1 Hierarchical Reinforcement Learning

Many real-world robotic tasks are inherently hierarchical and require long-horizon reasoning. Hierarchical Reinforcement Learning (HRL) addresses this challenge by learning policies at multiple levels of abstraction.[1][2] A high-level policy learns to break down a complex task into a sequence of simpler sub-goals, while a low-level policy learns to execute these sub-goals.[3] This decomposition offers several advantages:

a. Improved Sample Efficiency: By learning reusable sub-skills, HRL can significantly reduce the amount of experience needed to solve complex tasks.[1][4][5]
b. Enhanced Exploration: A high-level policy can guide exploration in a more structured and meaningful way, leading to faster discovery of effective behaviors.
c. Better Generalization: HRL promotes the learning of modular and compositional skills, which can be recombined to solve new, unseen tasks.[6]

Recent HRL frameworks have shown promise in enabling robots to solve complex manipulation and navigation tasks that would be intractable for traditional "flat" RL algorithms.[3][7][8] The development of methods for automatically discovering and learning the hierarchy of sub-skills remains an active and important area of research.[2].

5.2 Multi-Agent Reinforcement Learning (MARL) for Robotics

As robotic systems become more prevalent, the ability for multiple robots to coordinate and collaborate effectively is becoming increasingly crucial. Multi-Agent Reinforcement Learning (MARL) provides a framework for training teams of robots to accomplish collective goals.[9][10][11] MARL has the potential to revolutionize a wide range of applications, from automated manufacturing and logistics to search and rescue operations.[9][12][13]
Key research areas in MARL for robotics include:

a. Decentralized Control: Developing decentralized policies that allow each robot to make decisions based on its local observations while still contributing to the global objective.[14][15]
b. Communication Learning: Enabling robots to learn effective communication protocols to share information and coordinate their actions.[14]
c. Scalability: Designing MARL algorithms that can scale to large teams of robots without a prohibitive increase in computational complexity.[12]

Recent advancements in MARL have demonstrated the ability to train swarms of robots for tasks such as collaborative transport and formation control.[16][17] Overcoming the challenges of non-stationarity and partial observability in multi-robot systems is an ongoing focus of research.[9][14].

5.3 Lifelong and Continual Learning

For robots to be truly autonomous, they must be able to continuously learn and adapt to new situations and tasks over extended periods.[18][19] Lifelong or continual learning aims to develop robots that can accumulate knowledge over their lifetime without catastrophically forgetting previously learned skills.[20][21][22][23][24] This is a significant departure from the traditional machine learning paradigm of training on a fixed dataset.[18][25]
Addressing the challenge of catastrophic forgetting, where a neural network's performance on previously learned tasks degrades as it is trained on new ones, is a central focus of continual learning research.[21][24][26] Techniques being explored to mitigate this issue include:

a. Regularization-based approaches: These methods add a penalty term to the loss function to discourage large changes to the network weights that are important for previous tasks.[26]
b. Dynamic architectures: These approaches involve dynamically growing the network to accommodate new knowledge while preserving old information.
c. Experience replay: This technique involves storing a subset of past experiences and replaying them during training to reinforce previously learned skills.[21][24]

The development of robust continual learning algorithms will be essential for creating robots that can operate autonomously in dynamic and ever-changing real-world environments.[27][28][29][30].

5.4 Combining DRL with Other AI Techniques

The synergy between DRL and other AI techniques is leading to the development of more capable and versatile robots. By integrating different AI modalities, robots can achieve a more holistic understanding of their environment and interact with it in more sophisticated ways.

a. **DRL and Computer Vision:** The combination of DRL with advanced computer vision techniques, such as convolutional neural networks (CNNs) and vision transformers (ViTs), enables robots to learn control policies directly from raw pixel data.[31][32][33] This has been instrumental in enabling robots to perform tasks that require a high degree of visual understanding, such as object manipulation and autonomous navigation.[34][35]
b. **DRL and Natural Language Processing (NLP):** Integrating DRL with NLP allows for more natural and intuitive human-robot interaction. By understanding natural language commands, robots can be instructed to perform tasks in a more flexible and user-friendly manner.
c. **DRL and Imitation Learning (IL):** Combining DRL with imitation learning can significantly accelerate the learning process.[36][37] By leveraging expert demonstrations, imitation learning can provide a good initial policy, which can then be fine-tuned and improved upon using DRL.[38][39][40][41] This approach can overcome the sample inefficiency of pure RL and allow the robot to surpass the performance of the expert demonstrator.[36]

5.5 Explainable AI (XAI) for DRL in Robotics

As DRL-powered robots are deployed in safety-critical applications, such as autonomous driving and medical robotics, the ability to understand and trust their decision-making processes becomes paramount.[42][43] Explainable AI (XAI) aims to develop techniques for making the "black box" nature of deep neural networks more transparent and interpretable.[44]

In the context of DRL for robotics, XAI can provide insights into:

a. Why a robot chose a particular action: By visualizing the parts of the input state that were most influential in the robot's decision, we can gain a better understanding of its reasoning process.[45]
b. What a robot has learned: XAI can help to analyze the internal representations of the DRL agent and identify the features of the environment that it deems important.
c. When a robot is likely to fail: By understanding the limitations of the learned policy, we can identify situations where the robot's behavior may be unreliable.

The development of XAI methods for DRL is crucial for the certification and public acceptance of autonomous robotic systems, ensuring that they are not only capable but also trustworthy and accountable.[46].

6 Conclusion

6.1 Summary of Key Findings

This survey has provided a comprehensive overview of the significant strides made in the application of Deep Reinforcement Learning (DRL) to autonomous robotics. We began by establishing the foundational principles of DRL, charting its evolution from the core concepts of reinforcement learning to the integration of deep neural networks as powerful function approximators. Our exploration of the algorithmic landscape revealed a diverse taxonomy of methods, including value-based, policy-based, and actor-critic approaches, each offering unique strengths for tackling complex robotic control problems.

The transformative impact of these algorithms is evident across a wide spectrum of core robotic capabilities. In robotic manipulation, DRL has enabled unprecedented dexterity in tasks ranging from grasping and in-hand object manipulation to intricate assembly and tool use. For locomotion, DRL has been instrumental in generating dynamic and agile gaits for legged robots in challenging terrains and has provided robust control solutions for aerial and underwater vehicles. In the realm of autonomous navigation, DRL has facilitated the development of end-to-end learning systems that can navigate complex, dynamic environments, with significant implications for both mobile robotics and autonomous driving. Furthermore, DRL is proving to be a pivotal technology in shaping the future of human-robot interaction and collaboration, enabling more intuitive, safe, and efficient partnerships between humans and machines.

Despite these remarkable advancements, the widespread deployment of DRL in real-world robotics is still hampered by several persistent challenges. The simulation-to-reality gap remains a significant hurdle, necessitating the development of more sophisticated transfer learning and domain randomization techniques. The inherent sample inefficiency of many DRL algorithms continues to be a major bottleneck, driving research into model-based and off-policy methods. Ensuring the safety, robustness, and generalization of learned policies is of paramount importance, particularly in safety-critical applications. Additionally, the intricate art of reward function design and the fundamental trade-off between exploration and exploitation continue to be active areas of research.

6.2 Comparison of Privacy-Preserving Mechanisms in FL

Deep Reinforcement Learning stands at the forefront of a paradigm shift in robotics, moving away from rigid, pre-programmed systems towards intelligent, adaptive agents that can learn from experience and operate autonomously in the complexities of the real world. The potential for DRL to revolutionize industries, from manufacturing and healthcare to transportation and exploration, is undeniable. The emerging trends identified in this survey, including hierarchical reinforcement learning, multi-agent systems, lifelong learning, the fusion of DRL with other AI techniques, and the pursuit of explainable AI, offer a glimpse into a future populated by increasingly capable and versatile robots.

The path forward, however, is not without its challenges. The open problems discussed herein are not merely incremental hurdles but fundamental research questions that require concerted effort from the robotics and AI communities. Addressing these challenges will be crucial for unlocking the full potential of DRL and for building the trust and confidence necessary for its widespread adoption in society. The continued collaboration between researchers, industry, and policymakers will be essential in navigating the technical, ethical, and societal implications of this powerful technology. As we continue to push the frontiers of DRL, we move ever closer to a future where autonomous robots seamlessly integrate into our daily lives, augmenting human capabilities and tackling some of the world's most pressing challenges. The journey is far from over, but the progress thus far provides a compelling vision of the transformative era of robotics that lies ahead.

References

[1] D.-K. Han, B. Mulyana, and S. Cheng, "A Survey on Deep Reinforcement Learning Algorithms for Robotic Manipulation," Italian National Conference on Sensors, vol. 1, no. 1, p. 1, 2023.
[2] V. François-Lavet, P. Henderson, R. Islam, M. G. Bellemare, and J. Pineau, "An introduction to deep reinforcement learning," Foundations and Trends® in Machine Learning, vol. 11, no. 3-4, pp. 219–354, 2018.
[3] W. Zhao, J. P. Queralta, and T. Westerlund, "Sim-to-real transfer in deep reinforcement learning for robotics: a survey," in 2020 IEEE Symposium Series on Computational Intelligence (SSCI), 2020, pp. 737–744.
[4] J. Kober, J. A. D. Bagnell, and J. Peters, "Reinforcement learning in robotics: A survey," The International Journal of Robotics Research, vol. 32, no. 11, pp. 1238–1274, 2013.
[5] P. Kormushev, S. Calinon, and D. G. Caldwell, "Reinforcement learning in robotics: Applications and real-world challenges," Robotics, vol. 2, no. 3, pp. 122–148, 2013.
[6] C. Tang et al., "Deep Reinforcement Learning for Robotics: A Survey of Real-World Successes," Annual Review of

Control, Robotics, and Autonomous Systems, vol. 8, 2025.
[7] S. Levine, C. Finn, T. Darrell, and P. Abbeel, "End-to-end training of deep visuomotor policies," The Journal of Machine Learning Research, vol. 17, no. 1, pp. 1334–1373, 2016.
[8] Y. Li, "Deep Reinforcement Learning: An Overview," arXiv preprint arXiv:1701.07274, 2017.
[9] M. Arulkumaran, M. P. Deisenroth, M. Brundage, and A. A. Bharath, "A Brief Survey of Deep Reinforcement Learning," arXiv preprint arXiv:1708.05866, 2017.
[10] L. P. Kaelbling, M. L. Littman, and A. W. Moore, "Reinforcement learning: A survey," Journal of artificial intelligence research, vol. 4, pp. 237–285, 1996.
[11] R. S. Sutton, D. A. McAllester, S. P. Singh, and Y. Mansour, "Policy gradient methods for reinforcement learning with function approximation," in Advances in neural information processing systems, 1999, pp. 1057–1063.
[12] T. Tai, G. Paolo, and M. Liu, "A survey of deep network solutions for mobile robot navigation," in 2017 IEEE International Conference on Robotics and Automation (ICRA), 2017, pp. 5129–5134.
[13] M. L. Littman, "Reinforcement learning for robotics," in Robotics: Science and Systems, 2015, vol. 1, p. 1.
[14] A. A. R. El-Sayed, A. A. A. El-Sattar, and M. A. M. El-Tanboly, "Deep reinforcement learning for autonomous robotics: A survey," Ain Shams Engineering Journal, vol. 14, no. 3, p. 100414, 2023.
[15] J. Schulman, S. Levine, P. Abbeel, M. Jordan, and P. Moritz, "Trust region policy optimization," in International conference on machine learning, 2015, pp. 1889–1897.
[16] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," arXiv preprint arXiv:1707.06347, 2017.
[17] V. Mnih et al., "Asynchronous methods for deep reinforcement learning," in International conference on machine learning, 2016, pp. 1928–1937.
[18] T. P. Lillicrap et al., "Continuous control with deep reinforcement learning," arXiv preprint arXiv:1509.02971, 2015.
[19] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, "Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor," in International conference on machine learning, 2018, pp. 1861–1870.
[20] S. Levine et al., "Learning dexterous in-hand manipulation," The International Journal of Robotics Research, vol. 37, no. 4-5, pp. 421–437, 2018.
[21] D. Kalashnikov et al., "QT-Opt: Scalable deep reinforcement learning for vision-based robotic manipulation," in Conference on Robot Learning, 2018, pp. 691–701.
[22] J. Tan, T. Zhang, E. Coumans, A. Iscen, Y. Bai, and V. Vanhoucke, "Sim-to-real: Learning agile locomotion for quadruped robots," in Robotics: Science and Systems, 2018.
[23] N. Rudin et al., "Learning to walk in minutes using massively parallel deep reinforcement learning," in Conference on Robot Learning, 2022, pp. 91–100.
[24] G. Kahn, A. Villa, B. Pong, P. Abbeel, and S. Levine, "Uncertainty-aware reinforcement learning for collision avoidance," arXiv preprint arXiv:1702.01182, 2017.
[25] F. Chen, M. Liu, and Y. F. Li, "An end-to-end learning approach to mobile robot navigation in unknown indoor environments," in 2017 IEEE International Conference on Robotics and Biomimetics (ROBIO), 2017, pp. 1293–1298.
[26] S. Shalev-Shwartz, S. Shammah, and A. Shashua, "Safe, multi-agent reinforcement learning for autonomous driving," arXiv preprint arXiv:1610.03295, 2016.
[27] F. Codevilla, M. Müller, A. Dosovitskiy, E. Koltun, and V. Brox, "End-to-end driving via conditional imitation learning," in 2018 IEEE International Conference on Robotics and Automation (ICRA), 2018, pp. 1–9.
[28] Y. Chen, M. Everett, M. Liu, and J. P. How, "Socially aware motion planning with deep reinforcement learning," in 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2019, pp. 2947–2954.
[29] F. C. Chen, M. Liu, and Y. F. Li, "A deep reinforcement learning-based framework for safe human-robot collaboration," Robotics and Autonomous Systems, vol. 121, p. 103277, 2019.
[30] J. Tobin, R. Fong, A. Ray, J. Schneider, W. Zaremba, and P. Abbeel, "Domain randomization for transferring deep neural networks from simulation to the real world," in 2017 IEEE/RSJ international conference on intelligent robots and systems (IROS), 2017, pp. 23–30.
[31] A. A. R. El-Sayed, A. A. A. El-Sattar, and M. A. M. El-Tanboly, "Model-based reinforcement learning for robotics: a survey," Artificial Intelligence Review, pp. 1–55, 2023.
[32] A. Rajeswaran, V. Kumar, A. Gupta, G. Vezzani, J. Schulman, and E. Todorov, "Learning complex dexterous manipulation with deep reinforcement learning and demonstrations," in Robotics: Science and Systems, 2018.

[33] R. Pinto, J. Davidson, and A. Gupta, "Robust adversarial reinforcement learning," in International conference on machine learning, 2017, pp. 2817–2826.
[34] S. Amodei, D. Christiano, P. F. Steinhardt, J. Schulman, and D. Amodei, "Concrete problems in AI safety," arXiv preprint arXiv:1606.06565, 2016.
[35] A. Gaier and D. Ha, "Reward function design for reinforcement learning in robotics," in Conference on Robot Learning, 2019, pp. 493–503.
[36] M. Andrychowicz et al., "Hindsight experience replay," in Advances in neural information processing systems, 2017, pp. 5048–5058.
[37] D. E. B. Nachum, M. Andrychowicz, S. Levine, and J. Schulman, "Hierarchical reinforcement learning for long-horizon tasks," arXiv preprint arXiv:1810.06527, 2018.
[38] R. Lowe, Y. Wu, A. Tamar, J. Harb, O. P. Abbeel, and I. Mordatch, "Multi-agent actor-critic for mixed cooperative-competitive environments," in Advances in neural information processing systems, 2017, pp. 6379–6390.
[39] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in International conference on machine learning, 2017, pp. 1126–1135.
[40] J. X. Wang et al., "Learning to reinforcement learn," arXiv preprint arXiv:1611.05763, 2016.
[41] T. Schaul, D. Horgan, K. Gregor, and D. Silver, "Universal value function approximators," in International conference on machine learning, 2015, pp. 1312–1320.
[42] Y. Duan et al., "RL^2: Fast reinforcement learning via slow reinforcement learning," arXiv preprint arXiv:1611.02779, 2016.
[43] Z. Zhang, L. Tai, P. Yun, Y. Xiong, M. Liu, and J. M. O'Kane, "Deep reinforcement learning for vision-based robotic grasping: a survey," Robotics and Autonomous Systems, vol. 114, pp. 1–18, 2019.
[44] F. S. Melo, "A brief introduction to reinforcement learning," in Reinforcement Learning. Springer, 2015, pp. 1–21.
[45] D. Silver et al., "Mastering the game of Go with deep neural networks and tree search," nature, vol. 529, no. 7587, pp. 484–489, 2016.
[46] V. Mnih et al., "Human-level control through deep reinforcement learning," nature, vol. 518, no. 7540, pp. 529–533, 2015.