

Exploration Techniques in Reinforcement Learning for Autonomous Vehicles [†]

Ammar Khaleel ^{1,2,*}  and Áron Ballagi ²

¹ Doctoral School of Multidisciplinary Engineering Sciences, Széchenyi István University, Egyetem tér 1, H-9026 Győr, Hungary

² Department of Automation, Széchenyi István University, Egyetem tér 1, H-9026 Győr, Hungary; ballagi@ga.sze

* Correspondence: ammarkhaleel2022@gmail.com

[†] Presented at the Sustainable Mobility and Transportation Symposium 2024, Győr, Hungary, 14–16 October 2024.

Abstract: Autonomous vehicles (AVs) have the potential to revolutionize the transportation system by enhancing road safety, reducing traffic congestion, and freeing drivers from monotonous tasks. Effective exploration is essential for AVs to navigate safely and adapt to dynamic environments. Reinforcement learning (RL) enables AVs to learn optimal behaviors through continuous interaction with their environment. This paper reviews recent RL research on designing exploration strategies for single- and multi-agent AV systems. It categorizes exploration methods based on underlying principles and addresses the challenges. It analyzes key RL algorithms' strengths, limitations, and empirical performance. By compiling and analyzing the current state of research, this paper aims to facilitate future advancements in AV exploration using RL, offering insights into current trends and future directions in this evolving field.

Keywords: autonomous vehicles exploration; exploration vs. exploitation; exploration strategies



Citation: Khaleel, A.; Ballagi, Á. Exploration Techniques in Reinforcement Learning for Autonomous Vehicles. *Eng. Proc.* **2024**, *79*, 24. <https://doi.org/10.3390/engproc2024079024>

Academic Editors: András Lajos Nagy, Boglárka Eisinger Balassa, László Lendvai and Szabolcs Kocsis-Szürke

Published: 4 November 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The advent of AVs marks a significant milestone in technological innovation, promising to revolutionize the transportation system. These self-driving systems, which rely on advanced sensors, cameras, and software algorithms, have evolved substantially since their inception in the mid-20th century. Automated guided vehicles (AGVs) operate on fixed paths within controlled environments. In contrast, today's AVs navigate dynamic and unpredictable surroundings, requiring advanced real-time decision making capabilities.

Exploration problems focus on a vehicle's ability to navigate and gather information about its surroundings efficiently and safely. Effective exploration is essential for creating accurate maps, localizing the vehicle within these maps, and planning paths to desired destinations. AV exploration involves traversing unknown environments while collecting the necessary data to achieve specific objectives. Various exploration strategies have been proposed. These strategies are classified into reactive, deliberative, and hybrid paradigms [1]. RL has become a powerful tool for AV exploration; it enables the vehicle to learn optimal behaviors through interactions with its environment, making it well-suited to tasks involving uncertain or dynamic conditions. RL allows AVs to autonomously acquire skills through trial and error, thus adapting to new and complex scenarios.

This article surveys recent advances in AV exploration using RL while highlighting the methodologies and strategies used in single-agent and multi-agent systems. The goal is to showcase the potential of RL in enhancing AV capabilities and to offer insights into current research trends and future directions in this rapidly evolving field.

2. Exploration Role in AVs

Autonomous exploration refers to the capability of a vehicle to navigate through unknown environments while collecting essential data to accomplish specific objectives. Ideally, an autonomous vehicle (AV) should be able to explore its environment while simultaneously gathering data, constructing precise maps, and determining its position within these maps. The core navigation tasks—mapping, localization, and path planning—form the foundation of methodologies like Simultaneous Localization and Mapping (SLAM), active localization, and integrated navigation [2]. Figure 1a illustrates the fundamental elements of autonomous navigation [3]. Key challenges include map acquisition and target identification. Effective, application-specific exploration strategies balance map quality and coverage time, thus eliminating the need for human intervention [4].

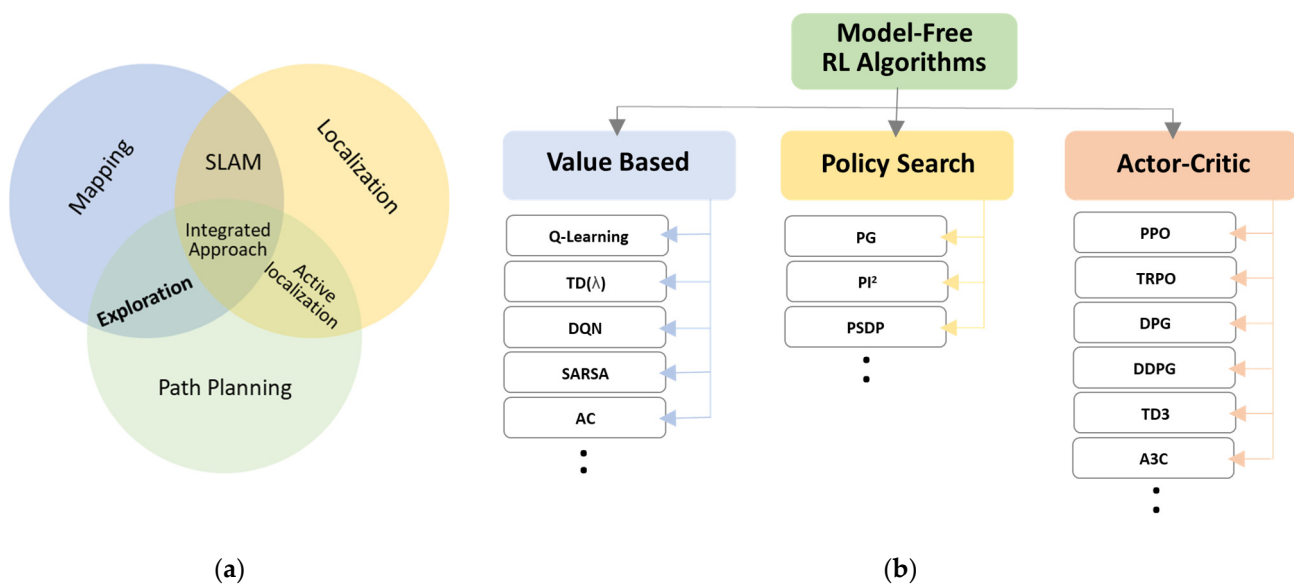


Figure 1. (a) Fundamental elements of autonomous navigation. (b) Model-free RL algorithms.

Exploration strategies can be classified into three main paradigms: reactive, deliberative (or hierarchical), and hybrid. Reactive strategies involve the AV responding to events without a planning phase [5]. Although easy to implement, reactive strategies lack learning, making them inefficient for complex tasks. In contrast, deliberative strategies involve perceiving the environment, planning, and executing actions in sequence. This approach improves decision making, but it can be challenging in dynamic conditions due to high computational demands [6]. However, hybrid strategies address this issue by combining planning with adaptability to unpredictable conditions. Techniques like the generalized Voronoi graph [7] and the frontier-based approach [8] segment environments, making logical decision making for autonomous navigation more effective. A critical aspect is selecting the next optimal destination, which often involves using interest attributes or cost utility models, such as market-based or next-best-view selection algorithms [9]. Furthermore, for path planning, graph search algorithms like RRT, A*, and Dijkstra [10], as well as machine learning techniques, are commonly used [11].

3. Reinforcement Learning Overview

RL [12] is a computational approach where an agent learns to make decisions via interactions with its environment. The agent identifies an optimal policy (π^*) that links situations to actions by employing a trial and error strategy to maximize a numerical reward signal. This reward function aligns with the application's objectives. The RL problem comprises an environment and an agent that perceives and modifies the environment's state driven by one or more objectives. The environment in RL is typically modeled as

a Markov Decision Process (MDP) or a Partially Observable Markov Decision Process (POMDP) [13].

Model-free methods, RL algorithms more common in complex environments, involve the agent learning the policy or value directly through trial and error without relying on an environmental model. Model-free algorithms can be classified into value-based, policy search, and actor–critic algorithms, as shown in Figure 1b.

Value-based methods, like Q-learning and deep Q-networks (DQNs), derive policies from value functions that estimate the benefits of specific states or actions [14]. Monte Carlo (MC) [15] and State–Action–Reward–State–Action (SARSA) [11] learn the value function by using actions from the current policy. On the other hand, policy search methods learn policies directly from the agent’s interactions with the environment, such as policy gradient (PG) [16]. Deterministic policy gradient (DPG) and its variations, such as deep deterministic policy gradient (DDPG) and twin delayed deep deterministic policy gradients (TD3), are other significant examples. Meanwhile, actor–critic methods, including soft actor–critic (SAC) and asynchronous advantage actor–critic (A3C), utilize a hybrid approach where the critic evaluates the policy choices that the actor predicts [17]. Proximal policy optimization (PPO) combines both approaches by using a critic to evaluate and an actor to predict policies [18]. Proximal policy optimization (PPO) and Trust Region Policy Optimization (TRPO) are policy gradient algorithms designed to refine policies while avoiding significant parameter changes, thereby preventing performance degradation [19].

4. RL Algorithms in Exploration Strategies for AVs

4.1. Value-Based Algorithms

Value-based algorithms, particularly deep Q-learning methods, are widely adopted for single-agent exploration. The robustness of deep neural networks in function approximation has enabled the application of classic Q-learning in complex tasks [20]. Additionally, actor–critic algorithms, such as the asynchronous advantage actor–critic (A3C) developed by Google DeepMind in 2016, are also significantly utilized [21]. These algorithms excel in decision making tasks like path planning and obstacle avoidance in real time. In multi-agent systems, deep Q-learning (DQN) and actor–critic methods are common, with DQN being preferred for discretization or function approximation. These algorithms are suitable for coordination tasks like cooperative navigation, where communication between vehicles is limited.

4.2. Two-Stage Strategies

In single-agent systems, two-stage strategies typically employ an RL algorithm to decide the next location, paired with methods for path guidance. For example, deep reinforcement learning (DRL) with frontier-based exploration uses an A3C network to process the known map and determine the next goal frontier [22]. Similarly, A3C and DQN networks guide movements, with the A* algorithm handling path planning [23]. For example, DRL combined with A* excels in guiding vehicles through complex environments while minimizing computational costs. In contrast, multi-agent systems use strategies that include RL for moves, dynamic Voronoi partitions for targets, and dual RL algorithms for decision making and guidance [24]. These strategies are often used for scenarios involving fleet management, where multiple AVs must coordinate movements while avoiding collisions.

4.3. End-To-End Strategies

End-to-end strategies are increasingly used for goal-guided exploration in single-agent systems by employing neural networks with convolutional layers to translate pixels into actions [25]. The Rainbow network architecture generates local paths in rough terrain, while DQN variations, such as double DQNs and D3QN, improve performance [26]. Additionally, non-CNN solutions like backpropagation networks and actor–critic models with deterministic policy gradients offer reactive exploration solutions.

In multi-agent systems, end-to-end strategies use sensor data to generate control actions via convolutional and fully connected layers. Discrete action spaces may cause abrupt behaviors in complex scenarios, and they require real-time computing [27]. End-to-end strategies facilitate seamless collaboration between AVs in environments like traffic-dense cities.

5. Common Challenges in Single- and Multi-Agent Systems

Exploration–Exploitation Dilemma: In single-agent systems, the exploration–exploitation dilemma is typically managed using the ϵ -greedy algorithm or its variations [28]. The ϵ value determines the agent's behavior, encouraging exploratory actions early in training and optimizing actions over time, which are suitable for AVs operating in well-mapped urban environments where exploitation of learned behaviors can be prioritized after sufficient exploration. Similarly, in multi-agent systems, the ϵ -greedy method is commonly used. Researchers have explored switching strategies based on collision risk and self-decay probability, as well as the Boltzmann distribution mechanism, to smooth action selection [29]. The ϵ -greedy approach is often paired with techniques to reduce collisions, which are essential for AV coordination in these scenarios.

Curse of Dimensionality: Deep neural networks (DNNs) are commonly employed to overcome the curse of dimensionality in single-agent systems. Moreover, techniques like Fixed Sparse Representation (FSR) and discretizing state and action spaces are also used [30]. DNNs are applied to mapping or obstacle avoidance in unstructured spaces. In multi-agent systems, RL algorithms are similarly applied to tasks with limited and discrete states and actions. Efficiently reducing space representation is a common strategy; for example, FSR approximation maps original Q-values to a low-dimension parameter vector. Furthermore, deep reinforcement learning (DRL) solutions learn low-dimensional state features from high-dimensional sensory data, providing robust function approximation [31]. Additionally, convolutional neural networks (CNNs) are increasingly used as feature extractors for raw images or other high-dimensional sensor data, such as platooning or cooperative driving in smart cities.

Reward Shaping: Reward shaping is critical in both single- and multi-agent systems. For example, rewards incentivize information gathering in search and rescue (SAR) missions, while goal-driven navigation tasks promote reducing the distance to the target. Moreover, collisions are universally penalized for RL for AVs. Although complex to define, defense rewards are commonly used because they expedite the learning process. In contrast, sparse rewards are simpler but tend to increase convergence time. The same reward-shaping principles apply to multi-agent systems. Positive rewards are given for discovering targets or unexplored areas. Dense rewards are often employed to speed up learning [32]. Scenarios like swarm robotics, where multiple AVs work together to map an environment, require dense reward signals to ensure efficient cooperation.

Learning Convergence: Several strategies are employed to accelerate learning convergence in both single-agent and multi-agent AV systems. One approach is discretizing state and action spaces to simplify the learning problem. Curiosity-driven intrinsic rewards also play a role by motivating agents to explore new areas and avoid deadlocks. Additionally, experience replay (EP) enhances data efficiency by reusing past experiences and reducing sample correlation to smooth the learning process [33]. Attention mechanisms improve data efficiency further by analyzing the importance of input.

Moreover, transfer learning techniques, such as initializing actor–critic networks with dynamic programming or supervised learning models trained on real-world data, are often used. In multi-agent systems, curriculum learning (CL) techniques expose agents to progressively challenging scenarios, which helps them acquire new skills. Additionally, experience sharing and prioritizing EP enhance learning efficiency [34].

6. Conclusions

This survey paper has reviewed the current state of exploration techniques in RL for Avs by categorizing them as single-agent and multi-agent systems. It highlights the basic ideas, methods, and challenges of each strategy. It shows the significant advantage that RL offers for Avs while discussing key algorithms, such as value-based methods, policy search, and actor–critic approaches, in the context of AV exploration. The findings of this survey have important implications for AV development. Integration of RL into AV systems has the potential to greatly enhance their ability to navigate complex and dynamic environments. Despite these advancements, several limitations remain. The curse of dimensionality continues to challenge the efficiency and scalability of RL algorithms. Additionally, ensuring safe exploration in real-world settings is difficult and costly, thus limiting the direct application of RL-trained models from simulations to real-world scenarios. Challenges in reward shaping and learning convergence also require innovative solutions.

Future research should prioritize the development of more efficient RL algorithms capable of addressing high-dimensional spaces and ensuring safe exploration in the real world. Improving the transferability of RL models from simulations to practical application is crucial. Hybrid strategies combining RL with traditional methods, advanced reward-shaping techniques, and enhanced learning convergence will drive future progress.

In summary, applying RL in AV exploration holds great promise for revolutionizing the transportation system. This survey provides a foundation for future work by offering useful insights into current trends and guiding the direction of future advancements in this rapidly evolving field.

Author Contributions: Conceptualization, A.K. and Á.B.; methodology, A.K. and Á.B.; writing—original draft preparation, A.K. and Á.B.; writing—review and editing, Á.B.; visualization, A.K.; supervision, A.K. and Á.B. All authors have read and agreed to the published version of the manuscript.

Funding: This study received no additional funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data for the study are available through the referenced sources.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Wijanto, E. Design of Deliberative and Reactive Hybrid Control System for Autonomous Stuff-Delivery Robot Rover. *Elkomika J. Tek. Energ. Elektr. Tek. Telekomun. Tek. Elektron.* **2023**, *11*, 1. [\[CrossRef\]](#)
2. Yarovoi, A.; Cho, Y.K. Review of simultaneous localization and mapping (SLAM) for construction robotics applications. *Autom. Constr.* **2024**, *162*, 105344. [\[CrossRef\]](#)
3. Makarenko, A.A.; Williams, S.B.; Bourgault, F.; Durrant-Whyte, H.F. An experiment in integrated exploration. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, Lausanne, Switzerland, 30 September–4 October 2002; pp. 534–539.
4. Placed, J.A.; Strader, J.; Carrillo, H.; Atanasov, N.; Indelman, V.; Carlone, L.; Castellanos, J.A. A survey on active simultaneous localization and mapping: State of the art and new frontiers. *IEEE Trans. Robot.* **2023**, *39*, 1686–1705. [\[CrossRef\]](#)
5. Cole, M.; Kulkarni, K.B.; Ewing, J.; Tau, S.; Goodin, C.; Jayakumar, P. Exploration of robust and intelligent navigation algorithms to ensure off-road autonomous vehicle mobility. *Int. J. Veh. Perform.* **2024**, *10*, 239–267. [\[CrossRef\]](#)
6. Sui, T. Exploring moral algorithm preferences in autonomous vehicle dilemmas: An empirical study. *Front. Psychol.* **2023**, *14*, 1229245. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Alitappeh, R.; Jeddisaravi, K. Multi-robot exploration in task allocation problem. *Appl. Intell.* **2022**, *52*, 2189–2211. [\[CrossRef\]](#)
8. Yan, X.; Huang, J.; He, K.; Hong, H.; Xu, D. Autonomous exploration through deep reinforcement learning. *Ind. Robot Int. J. Robot. Res. Appl.* **2023**, *50*, 793–803. [\[CrossRef\]](#)
9. Ansari, Y.; Yasmin, S.; Naz, S.; Zaffar, H.; Ali, Z.; Moon, J.; Rho, S. A deep reinforcement learning-based decision support system for automated stock market trading. *IEEE Access* **2022**, *10*, 127469–127501. [\[CrossRef\]](#)
10. Karur, K.; Sharma, N.; Dharmatti, C.; Siegel, J.E. A survey of path planning algorithms for mobile robots. *Vehicles* **2023**, *3*, 448–468. [\[CrossRef\]](#)

11. Swain, N.R.; Kulkarni, A.J.; Sonawwanay, P.D.; Babu, A.; Kale, J.G. Machine Learning Algorithms for Autonomous Vehicles. In *Handbook of Formal Optimization*; Springer Nature: Singapore, 2024; pp. 1–54.
12. Sutton, R.S.; Barto, A.G. *Reinforcement Learning: An introduction*; MIT Press: Cambridge, MA, USA, 2018.
13. Liang, Q.; Liu, J.; Jiang, Z.; Yin, J.; Xu, K.; Li, H. Limited Information Aggregation for Collaborative Driving in Multi-Agent Autonomous Vehicles. *IEEE Robot. Autom. Lett.* **2024**, *9*, 6624–6631. [[CrossRef](#)]
14. Byeon, H. Advances in Value-based, Policy-based, and Deep Learning-based Reinforcement Learning. *Int. J. Adv. Comput. Sci. Appl.* **2023**, *14*, 348–354. [[CrossRef](#)]
15. Wu, J.; Yang, H.; Yang, L.; Huang, Y.; He, X.; Lv, C. Human-Guided Deep Reinforcement Learning for Optimal Decision Making of Autonomous Vehicles. *IEEE Trans. Syst. Man Cybern. Syst.* **2024**; *accepted*. [[CrossRef](#)]
16. Cai, X.; Bai, X.; Cui, Z.; Hang, P.; Yu, H.; Ren, Y. Adversarial Stress Test for Autonomous Vehicle Via Series Reinforcement Learning Tasks With Reward Shaping. *IEEE Trans. Intell. Veh.* **2024**; *accepted*. [[CrossRef](#)]
17. Chen, Y.; Wu, S. Framework of active obstacle avoidance for autonomous vehicle based on hybrid soft actor-critic algorithm. *J. Transp. Eng. Part A Syst.* **2023**, *149*, 04023002. [[CrossRef](#)]
18. Wu, Y.; Yuan, X. Proximal Policy Optimization-based Reinforcement Learning for End-to-end Autonomous Driving. In Proceedings of the 2023 38th Youth Academic Annual Conference of the Chinese Association of Automation (YAC), Hefei, China, 27–29 August 2023; pp. 844–849.
19. Shi, H.; Zhou, Y.; Wu, K.; Wang, X.; Lin, Y.; Ran, B. Connected automated vehicle cooperative control with a deep reinforcement learning approach in a mixed traffic environment. *Transp. Res. Part C Emerg. Technol.* **2021**, *133*, 103421. [[CrossRef](#)]
20. Yang, J.; Zhang, J.; Xi, M.; Lei, Y.; Sun, Y. A deep reinforcement learning algorithm suitable for autonomous vehicles: Double bootstrapped soft-actor-critic-discrete. *IEEE Trans. Cogn. Dev. Syst.* **2021**, *15*, 2041–2052. [[CrossRef](#)]
21. Darapaneni, N.; Paduri, A.R.; Sudha, B.G.; Bendre, V.; Chandran, M.; Mohana Priya, M.; Jacob, V. Controlling the Steering Wheel Using Deep Reinforcement Learning: A Survey. In Proceedings of the International Conference on Paradigms of Communication, Computing and Data Analytics (PCCDA 2023), Delhi, India, 22–23 April 2023; pp. 805–815.
22. Muzahid, A.J.M.; Kamarulzaman, S.F.; Rahman, M.A.; Alenezi, A.H. Deep reinforcement learning-based driving strategy for avoidance of chain collisions and its safety efficiency analysis in autonomous vehicles. *IEEE Access* **2022**, *10*, 43303–43319. [[CrossRef](#)]
23. Li, J.; Chen, Y.; Zhao, X.; Huang, J. An improved DQN path planning algorithm. *J. Supercomput.* **2022**, *78*, 616–639. [[CrossRef](#)]
24. Xue, X.; Li, Z.; Zhang, D.; Yan, Y. A Deep Reinforcement Learning Method for Mobile Robot Collision Avoidance based on Double DQN. In Proceedings of the 2019 IEEE 28th International Symposium on Industrial Electronics (ISIE), Vancouver, BC, Canada, 12–14 June 2019; pp. 2131–2136.
25. Josef, S.; Degani, A. Deep reinforcement learning for safe local planning of a ground vehicle in unknown rough terrain. *IEEE Robot. Autom. Lett.* **2020**, *5*, 6748–6755. [[CrossRef](#)]
26. Zhelo, O.; Zhang, J.; Tai, L.; Liu, M.; Burgard, W. Curiosity-driven exploration for mapless navigation with deep reinforcement learning. *arXiv* **2018**, arXiv:1804.00456.
27. Palanisamy, P. Multi-agent connected autonomous driving using deep reinforcement learning. In Proceedings of the 2020 International Joint Conference on Neural Networks (IJCNN), Glasgow, UK, 19–24 July 2020; pp. 1–7.
28. Yijing, Z.; Zheng, Z.; Xiaoyi, Z.; Yang, L. Q Learning Algorithm Based UAV Path Learning and Obstacle Avoidance Approach. In Proceedings of the 2017 36th Chinese control conference (CCC), Dalian, China, 26–28 July 2017; pp. 3397–3402.
29. Kwa, H.L.; Leong Kit, J.; Bouffanais, R. Balancing collective exploration and exploitation in multi-agent and multi-robot systems: A review. *Front. Robot. AI* **2022**, *8*, 771520. [[CrossRef](#)] [[PubMed](#)]
30. Pham, H.X.; La, H.M.; Feil-Seifer, D.; Van Nguyen, L. Reinforcement Learning for Autonomous UAV Navigation Using Function Approximation. In Proceedings of the 2018 IEEE international symposium on safety, security, and rescue robotics (SSRR), Philadelphia, PA, USA, 6–8 August 2018; pp. 1–6.
31. Lin, J.; Yang, X.; Zheng, P.; Cheng, H. End-to-end Decentralized Multi-robot Navigation in Unknown Complex Environments via Deep Reinforcement Learning. In Proceedings of the 2019 IEEE International Conference on Mechatronics and Automation (ICMA), Tianjin, China, 4–7 August 2019; pp. 2493–2500.
32. Li, J.; Wu, X.; Xu, M.; Liu, Y. Deep reinforcement learning and reward shaping based eco-driving control for automated HEVs among signalized intersections. *Energy* **2022**, *251*, 123924. [[CrossRef](#)]
33. Fan, T.; Long, P.; Liu, W.; Pan, J.; Yang, R.; Manocha, D. Learning Resilient Behaviors for Navigation Under Uncertainty. In Proceedings of the 2020 IEEE International Conference on Robotics and Automation (ICRA), Paris, France, 31 May–31 August 2020; pp. 5299–5305.
34. Chen, Z.; Subagdja, B.; Tan, A.H. End-to-end Deep Reinforcement Learning for Multi-agent Collaborative Exploration. In Proceedings of the 2019 IEEE International Conference on Agents (ICA), Jinan, China, 18–21 October 2019; pp. 99–102.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.