A reawakening of Machine Learning Application in Unmanned Aerial Vehicle: Future Research Motivation

Wasswa Shafik¹, S. Mojtaba Matinkhah*, ², Fawad Shokoor² and Lule Sharif³

¹Computer Engineering Department, Intelligent Connectivity Research Lab, P.O. Box 89175-741, Yazd University, Yazd, Iran
²Computer Engineering Department, Yazd University, P.O. Box 89175-741, Yazd, Iran
³Department of Management Human Resource, Islamic University in Uganda, Kampala, Uganda

Abstract

Machine learning (ML) entails artificial procedures that improve robotically through experience and using data. Supervised, unsupervised, semi-supervised, and Reinforcement Learning (RL) are the main types of ML. This study mainly focuses on RL and Deep learning, since necessitates mainly sequential and consecutive decision-making context. This is a comparison to supervised and non-supervised learning due to the interactive nature of the environment. Exploiting a forthcoming accumulative compensation and its stimulus of machines, complex policy decisions. The study further analyses and presents ML perspectives depicting state-of-the-art developments with advancement, relatively depicting the future trend of RL based on its applicability in technology. It's a challenge to an Internet of Things (IoT) and demonstrates what possibly can be adopted as a solution. This study presented a summarized perspective on identified arenas on the analysis of RL. The study scrutinized that a reasonable number of the techniques engrossed in alternating policy values instead of modifying other gears in an exact state of intellectual. The study presented a strong foundation for the current studies to be adopted by the researchers from different research backgrounds to develop models, and architectures that are relevant.

Keywords: Machine learning, Reinforcement learning, supervised learning, non-supervised learning, semi-supervised learning

1. Introduction

Machine Learning (ML) entails many approaches like supervised (semi, and un) supervised learning that below a big entity of Artificial Intelligence. Reinforcement Learning (RL) in specific can be defined to be a promising and trending ML field that observes the environmental problems and practices [1]. This attempt to retro-feed its prototypical to enhance the desired execution of the IoTs. The network devices that are embedded with sensors, software, and extra technologies. This is mainly for connection, and data exchange with other devices and systems over the Internet, as was explicitly [2].

Several entities and fields have been seen embracing the ML including computer science, neurosciences, psychology, economics, mathematics, and engineering. Classical conditioning among others leads to increased detections on different social media platforms and networks, and increased technical control. Thus, making a better efficiency, increased operational capacity, autonomous and semi-autonomous operation [3-5].

RL has been mainly embraced in the global connectivity performing different duties to the developer and the end user's platforms connecting devices on the Internet, nowadays called the IoTs. These are used to habitually tag persons and recognize mutual kinds of stuff like breakthrughs in uploaded representations. It is used in product referencing, health analyses, generation division, rate expectation, fiscal analysis, predictive

*Corresponding author. Email: matinkhah@yazd.ac.ir
maintenance, and indistinguishable acknowledgment [6-8]. Early studies demonstrated that the inverse RL is formulated as trouble of a boosting classifier, comparable to amend the Ad boost set of rules for classification [9]. And the characteristic expectancies from experts’ demonstration, and the trajectory brought about via way of means of an agent's modern-day policy [10] and [11].

An early submission was focused on this current generation of technology, termed the fifth generation, which is expected to be focusing on the speed of execution and computation with pricing models [12]. Analyzed to different communication broadcasts and admission of requests [13] and [14]. It was revealed that security and privacy are the most persisting and existing major challenges as different nets prevail [15] and [16]. This is examined during performance evaluation that was carried out in modern technologies and packet loss possibilities, depicting ways of its management [17] and [18]. Besides the two identified issues of the internet, the IoTs-energy-based approaches were presented [19] with great focus on the RL in knowledge engineering [20]. Figure 1 and detailed Figure 2 illustrates mainly ML classifications, RL classifications, processes, and approaches.

A comprehensive and exhaustive systematic investigation of the main eight fundamental reinforcement learning processes. The six RL operations, the fundamental reinforcement learning tactics including artificial neuron networks, and learning automata among others. Designated current state-of-the-art opportunities are presented, approaches used in reinforcement algorithm implementation, its application, scrutinized and styled the relationship between the RL as a technique data training, approaches.

The performance of the Duisburg statistical object tracker (PETS) 2000 catalog indicated that the feature arguments are mined from the targets that are moving objects [7]. According to [21], [22], labor-intensive land patterning is seen to be time overwhelming, and extremely specialized work was current algorithms have flexibility deficiency and might be predisposed to facts multiplicity.

This study is structured as follows: In Section 2, the detailed literature on reinforcement learning in different areas is availed. The section presents the basic process of reinforcement learning. The operatives of this learning including iterations, learnings, Rewarding and penalties, action with others and learning techniques. In Section 3, approaches used in implementing RL algorithms are explained. Section 4, demonstrates deep learning application in Unmanned Aerial Vehicle (UAVs), techniques algorithmic implementation approach where the three approaches are simplified in summary including value-based, policy-based, and model-based approach, secondly, reinforcement learning applications are discussed that includes text mining, health operations, gaming, education mention but a few, thirdly reinforcement learning challenges are illustrated in details with possible learning opportunities and this section further depicts future research direction and depicting why the future of reinforcement learning is quite bright compared to other learnings. In Section 5, technical loopholes are presented in state-of-the-art models. In Section 6, future research direction is precisely demonstrated. Lastly, within section 7, conclusion of this paper with a clear display next research prospect. Bottlenecks, and challenges are classified and availed in simple tables and figures.

2. Reinforcement Learning Detailed

In this section, three main entities that include the fundamental procedures that are outstepped during the RL processing are presented. A brief demonstration how RL operates, and the RL techniques are shown how each category differs.

Agent - This is often the first step of RL that is seen to be a theoretical and a conjectural entity that accomplishes actions in an atmosphere to advance certain rewards. Action (ac)- This is repeatedly considered to be the second phase of RL that focuses on entirely the potential interchanges that the agent might consider. Environment- At this phase, the situation of an agent partakes to aspect.

State (St)- This is repetitively considered to be the phase at which the recent condition is reimbursed by the atmosphere. Reward (R)- An instantaneous reoccurrence of sent back from the atmosphere to assess and estimate the last exploit done by the agent. Policy (Po)- This is the strategy point, here strategy is seen and obtained in search of an agent engages to regulate the next accomplishment established in the existing state. Value (Va)- the anticipated long-term reoccurrence with deduction, as divergent to the short-term reward R. Va Po (st), is distracted as the expected long-term return of the existing states underneath policy Po.

Q-Value or Action-Value (Qv) lastly on the processes of RL list is the Qv related to or mainly proportional to the Value, except that it earns an extra constraint, the present action ac. Qv Po (st, ac) denoting the long-term reappearance of the present states, enchanting action a below policy Po; see [23]. A brief demonstration of how RL operates, and the last subsection shall entail the RL techniques, is discussed. Observation- it appears to be a phase of observing the environment where we interchange the environment to the atmosphere, the two are used interchangeably. Technically, the main point is to find out the back and flows of connectivity [24].

Decision-Making- Next, after the observation of the atmosphere, the environment focuses on the state of the agent as defined early, and the agent does the same thing but focusing on the action to be done [25]. Action- in this step, the action to be done is acted upon based on the first one and second phases, consequently per the observations like sharing [26]. Rewarding or Penalty- Putting aside policy in RL, the other important phase is rewarding.
Using the expected number of the rewards in each state of the action; the Bellman approach in the early years of the 1980s of paths [27]. Learning and strategizing- RL uses the real geographical setup in observing and learning the environments, these learnings from the experiences besides humanizing the desired strategy [28]. Iteration- the last phase is to recapitulate until an optimal strategy is initiated [29], see Figure 1.

2. Learning Methods

This section presents learning techniques illustrated. In this conception, the device is availed with a tradition of satisfactory actions, rubrics, and budding end states, for instance, the rules of the inclined are demarcated.

2.1. Markov Decision Process

Markov decision process (MDP) is simply considered to be a discrete-time regulator process stochastically and a memory-less random progression [30] and [31]. Several studies have been seen confusing up the RL and the MDP. There exists a thin line in-between [32], the problematic aspect of the concepts can be resolved as MDP has described early and the RL depends on the MDP depiction actuality a correct counterpart to the tricky and [33].

2.2. Q-Learning

QL can be simply defined as a model-free RL process to acquire strategies expressively informing the agent on what action to do depending on a given situation. Some functions are returned to be the recompressions rummage-sale to deliver the strengthening providing the quality of an action reserved in a specified situation [34] and [35].

The QL embraces diverse alternatives entailing the deep QL, Double QL, and Delayed QL. Based on [36]. The Greedy (GQL) rummage sale in the practice in amalgamation with occupation approximation and conjunction [37].

2.3. Learning Automata

This is one of the studies or learning techniques that were inverted in the early 1960s that can be deliberated to be LA where it selects and inferior its existing action per the past familiarities from the atmosphere. This makes it more suitable to tumble interested in the variety of RL in case the atmosphere is stochastically utilized; MDP is used [38] and [39]. LA is an adaptive and environmental decision-making component positioned in an arbitrary atmosphere that studies the optimum action via repetitive connections with its atmosphere. The actions are elected, permitting an explicit distribution of probabilities that are updated based on the atmosphere reply to the robotics in obtaining better and outstanding performance over a given specific action [40] and [41].

2.4. Artificial Neural Network

ANN at times is termed to be called connectionist systems are technical schemes ambiguously enthused by the natural neural nets that institute visceral brains. These kinds of schemes tend to absorb mainly to execute responsibilities via some considerable, for instance, without actualities of the automatic task-specific guidelines [42-44]. This technique of ANN mainly endeavours to ape the nets of neurons that sort up a human brain, hence enabling the processor to learn pieces of stuff and support decisions in a human-related-state [45].

3. Approaches Used in implementing the Reinforcement Learning Algorithm

In this section, the three main approaches that are used in implementing the RL algorithm are illustrated and described below, briefly discuss the RL applications, RL challenges, and opportunities. Number of articles in Web of Science, Scopus and Google Scholar citing the Citation Indexes for machine learning articles between 2010 and 2020; (see Figure 2).

![Figure 2. A few articles in Web of Science, Scopus and Google Scholar citing the Citation Indexes for machine learning articles between 2010 and 2020.](image)
3.1. Value-Based Approach

In this initial approach called a value-based RL technique, here the programmer tries attempts to exhaust all the possibilities optimal a value function. The foremost emphasis is to discover an optimal value per the desired decision and achievable under any policy [46, 47]. The exists a distinct difference between RL and other ML as precisely verified in Table 1.

3.2. Model-Based Approach

In this category of RL, the programmer tends to create a computer-generated archetypal for every environment, and the agent studies to accomplish in that explicitly planned environment [48].

4. Reinforcement Learning Applications

In this subsection, the main RL applications are discussed with a descriptive approach to show RL involvement and loophole that need enhancements.

4.1. Robotics and Education

This term has been in technical words for some generations, the robots are now substituting human operations and activities due to the autofocus of the machine and way of operation. Robotics entails reinforcement learning where autonomous expertise is used in making usage in multidiscipline research, vigorous scheme modelling and scrutiny, reckoning, ecology, engineering, and, robotics high technology [49]. Online studies have been too much visible as technology advances, reinforcement learning ensures adaptive

---

### Table 1. Distinct differences between machine learning classifications

<table>
<thead>
<tr>
<th>Basic Parameters</th>
<th>Mode of Learning</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Supervised</td>
</tr>
<tr>
<td>Definition</td>
<td>Learning based on input or output pairs</td>
</tr>
<tr>
<td>Probing</td>
<td>No-Probing</td>
</tr>
<tr>
<td>Teaching data</td>
<td>Necessity purview proficient to tag facts</td>
</tr>
<tr>
<td>Predilection</td>
<td>Monotonous tasking</td>
</tr>
<tr>
<td>Optimal strategy</td>
<td>Facts and erudition procedures</td>
</tr>
<tr>
<td>Extent</td>
<td>ML</td>
</tr>
<tr>
<td>Approach</td>
<td>Output</td>
</tr>
<tr>
<td>Prevalent algorithm</td>
<td>Linear regression, Logistic regression</td>
</tr>
<tr>
<td>Training</td>
<td>External supervision</td>
</tr>
<tr>
<td>Kind of Data</td>
<td>Labeled Data</td>
</tr>
<tr>
<td>Problem Kinds</td>
<td>Classification and Regression</td>
</tr>
<tr>
<td>Integration</td>
<td>Runs on any application</td>
</tr>
</tbody>
</table>
learning and smart education traits. The identified applications are not the only others to include inventory supervision, delivery management, power schemes, finance Subdivision, and manufacturing (industrial) [55-59].

4.2. Text Mining or Text Excavation

Typescript excavating also discussed as writing data insertion, unevenly comparable to text analytics, that involves the practice of springing superior info from text. Excellent gen is stereotypically resultant complete the formulating of arrangements and inclinations via means such as geometric pattern erudition. Automated customer care service, deception discovery through prerogative analysis, contextual publicizes, and amelioration [50].

4.3. Trade Execution and Healthcare

RL is evident due to the superlative accomplishment of duties given by an investment facilities firm, executing commands on behalf of clientele to guarantee the preeminent implementation imaginable for their clients' remits. RL helps to reduce when human decision-making, and facilities in return enabling better standards. An intellectual routine leading temperature monitoring, telemedicine, electronic health record management, and personal digital assistance [51] and [52].

4.3 Gaming, Smart Business, and Market

Manoeuvres like smart game controllers and toys progressively assimilate sensors to ensure gaming is done easy and user-friendly, identification of games that support brains in growing [53] and [54]. A smart marketing and smart business tend to be an intervallic auction that is vacant by the process's inquiry practice of scientific optimization, i.e., rectilinear indoctrination. The summarized RL approaches, applications, and challenges are displayed in Table 3.

5. Reinforcement Learning Challenges & Opportunities

In this subsection, the main RL challenges are opportunities are discussed with the eloquent approach to show RL participation and loophole that need improvements. This may improve the efficiency of systems with security and privacy are detailed below:

5.1. Multi-Task Learning

Persistence is a multi-task learning achievement mostly when it comes to resource pooling, formerly solitary of the gigantic encounters of AI and RL inexact of [60] and [61]. The fundamental of this task is scalability. An RL agent ought to avail of a library of all-purpose awareness and acquire wide-ranging expertise that may be used across a variability of errands [62].

5.2. Learning to Remember

The notable number of real-life responsibilities, surveillance merely seizures a small fragment of the full atmosphere state that regulates the best exploitation. In such moderately noticeable atmospheres, an agent must yield into justification just not the existing thought, but also previous annotations in mandate to regulate the best action. As individuals speak, people tend to shift from one topic to the other, altering the themes and entwining back again. Info is important, although supplementary info is more lateral [63].

5.3. Feature Selection

Feature selection for instance by self-paced learning regularization [65], a framework using robust 0-1 integer programming [66-68]. The progression of recognizing and distinguishing a minor subset of extremely predictive topographies out of a large set of contender geographies. The current situation is that in case the features are mixed up, the current algorithms cannot knowledge the features given several inputs.

5.4. Robust Representation Schemes and Interpretability

In general setup, it is still a prevailing problem to interpret and understand contexts, the clue behind this tactic is to contemplate and translate circumstantial connotation regardless of its importance to the owner; see [69] and [70]. In the same line, RL still has the challenge to robustly represent and learn algorithms and interpret them per the predestined outputs.

5.5. Transfer Learning Issues

TL is one of the recent is investigation problematic arenas in ML that emphasizes tidying away information obtained while resolving one unruly and smearing it to a dissimilar but associated delinquent. Consider an illustration of gen gained while eradiating to distinguish cars may besmear once irritating to identify trucks; see [71].

5.6. Continuous Learning Challenges

The constant extension of expertise and expertise sets technically via erudition and cumulative information is key in this challenge like in robotics. These continuous learning challenges are interactive in that there is back and forth of learning naturally [72].

5.7. Adaptability in Contexts
It is commonly known as adaptive teaching, which is an instructive technique that practices processor algorithms to compose the interface with the apprentice and distribute made-to-order possessions and learning actions to discourse the inimitable necessities of each apprentice. Its corporate's preparation of the future of operatives mainly in training, approaches, and culture traits that are inimitable to every apprentice [73].

5.8. Planning

Some work frames, models, and algorithms have been developed, for example, improved retrosynthetic planning [74]. A maintenance planning framework for mechanical, electrical, and plumbing components [75], human-scale greenway planning [76], cracking the motion planning issues [77] and planning support systems [78]. Due to poor planning of algorithm performance, specific policies, or schedules to influence from the start state to the goalmouth state tend to fail. Researchers are cautioned to consider all possible traits of adaptability, continuousness, and transfer among other factors during the planning of the algorithm.

5.9. Data Efficiency and Stability

The efficiency of data analytics and stability of the program, notable factors are seen into play, mainly operational and functional like eloquent whom routines come again sensitive data. Terminated principles yield redundant compliance projects, self-protective sensitive pamphlets fitting to their value, spring-cleaning up individual poisonous data junkyard, and subcontracting superficial info management [79].

5.10. Algorithm Policing

ML as the field is having a challenge of policies besides early definition. This is seen in international and intergovernmental operations. Different systems are developed to serve different purposes per different national [80] and [81]. This is a challenge in that a system might be of threat to a foreign of government. When adopted from a country that did not develop it, in long run will course increased cyber and technical technological problems among nations.

6. Deep Learning applications in UAV

In this section, the study focuses on the most current identified challenges that the Unmanned Aerial Vehicles face, considering the challenges in response to the proposed model, technical outcomes are process.

6.1. Resource Allocation of UAVs

The most current model proposed in search for better resource allocations called DDPG-FRAS. The DDPG-FRAS model focused on the optimization of the UAV flight controls, and collection of data scheduling structures within the specified time constraints. The DDPG-FRAS minimized the reasonable packet losses of the ground sensor nets and thus gradual converges, even though extending the holding capacities reduces the packet loss to approximately 50% [82].

6.2. Fresh Data Collections

The collection of onsite, and current data assembly in UAV-aided IoTs networks is seen to be a network challenge. MDP was used in determining an optimum UAV trajectory, and diffusion arrangement of the sensor nodes to minimize the prejudiced sum of the age of information [83].

6.3. Trajectory Designs and Mode Selection

Transportable devices by UAV-to-Device transportations over cellular networks, otherwise directly through the base station. To address this issue, MDP with a large state-action space, utilizing multi-agent DRL to estimate the state-action space, and complete proposing a multi-UAV trajectory design algorithm. This resulted in achieving a sophisticated total utility compared to the policy incline algorithms, and single-representative procedure among others [84] and [85].

6.4. Autonomous Task Offloading for UAVs

In approach to solving this offloading issue optimally in the decision-making of the tasks produced by the end-users. A distributed DNN several the training instances, for authentications, selecting the distributed deep neural network that provided at least the training loss. Trained distributed deep neural networks could accomplish near-optimal recital with numerous scheme stricture settings [86].

6.5. Energy Efficiency

A novel deep learning-based framework to challenge the energy delinquent was to formulate a group of UAVs energy-resourcefully. The proposed a deep reinforcement learning DRL model Convolution Neural Turing Machine (j-PPO+ConvNMT) model. It has the competence to style continuous route planning, and discrete that includes moreover to accumulate data action decisions instantaneously for all UAVs [87]. Other related approaches are detailed in [88-91], other within Table 2.
5. Technical Loophole in the State-of-art-proposed Models

In this section, the study focuses on the most current identified challenges that the Unmanned Aerial Vehicles face. Considering the challenges in response to the proposed model, technical outcomes.

5.1. Resource Allocation of UAVs

The UAV maintenance accurate time logs which in results in the enabling data collections and forecast data transmissions of the grounded nodes despite the fact monitoring UAV flight. The velocities reduce the multiple data packet losses that is after buffer overflows, and the channel fading. The most current model proposed in search for better resource allocations is DDPG-FRAS. The DDPG-FRAS model focused on the optimization of the UAV flight controls, and collection of data scheduling structures within the specified time constraints. The DDPG-FRAS minimized the reasonable packet losses of the grounded sensor networks. Gradual converges, even though extending the holding capacities reduces the packet loss to approximately 50% [82].

5.2. Fresh Data Collections

The collection of onsite and current data collection in UAV-assisted IoTs networks is seen to be a network challenge. This increases on safer storage, and UAV data backups, quick access to collectable data files. This is done often when the UAV collected data is transferred towards the network sensors in collecting points. This helps in collection, manipulation, retrieve, and update packets inside a specified time frame in conscious to the non-negative residual energy. The MDP was used in determining an optimum UAV trajectory, and transmission scheduling of the sensor nodes to minimize the prejudiced sum of the age of information [83].

5.3. Trajectory Designs and Mode Selection

The UAVs' sensing and transmission could influence their trajectories. The flight style drawback for UAVs in the

<table>
<thead>
<tr>
<th>Machine Learning Entities</th>
<th>Challenges</th>
<th>Proposed Models</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deterministic policy</td>
<td>Resource Allocation</td>
<td>DDPG-FRAS</td>
<td>[82]</td>
</tr>
<tr>
<td>Deep Reinforcement Learning</td>
<td>Fresh Data Collection</td>
<td>MDP Model</td>
<td>[83]</td>
</tr>
<tr>
<td></td>
<td>Trajectory Design &amp; mode Selection</td>
<td>MDP with consideration to sensing and transmission observation</td>
<td>[84]</td>
</tr>
<tr>
<td></td>
<td>Autonomous UAV Navigations</td>
<td>MDP with sparse Rewards</td>
<td>[85]</td>
</tr>
<tr>
<td>Distributed Deep Learning</td>
<td>Tasks Offloading</td>
<td>A distributed Deep Neural Networks</td>
<td>[86]</td>
</tr>
<tr>
<td>Deep Learning</td>
<td>Energy Efficiency</td>
<td>j-PP+ConvNTM Model</td>
<td>[87]</td>
</tr>
<tr>
<td></td>
<td>Anthropological Target Search</td>
<td>Autonomous UAV with Learning</td>
<td>[88]</td>
</tr>
<tr>
<td></td>
<td>Human / plant Detection</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deep Reinforcement Learning</td>
<td>Trajectory Designs</td>
<td>MDP Model</td>
<td>[89]</td>
</tr>
<tr>
<td></td>
<td>Power Allocation</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Efficient Data Collection</td>
<td>UAV flight Trajectory Model</td>
<td>[90]</td>
</tr>
<tr>
<td></td>
<td>Energy efficient Multiple UAV Navigation</td>
<td>A decentralized deep RL</td>
<td>[91]</td>
</tr>
<tr>
<td></td>
<td>Trajectory Planning</td>
<td>A partial Observable MDP</td>
<td>[92]</td>
</tr>
<tr>
<td>Deep Q-Learning</td>
<td>Network Resource Management</td>
<td>UAV aided WPT Framework.</td>
<td>[93]</td>
</tr>
</tbody>
</table>
thought of their sensing and transmission was considered. This is often what a mathematician call drawback with an outsized state-action house. A multi-UAV flight style algorithmic rule to resolve this drawback. Simulation results show that our planned algorithmic rule can do the next total utility than the policy gradient algorithmic rule and single-agent algorithms [84].

5.4. Autonomous UAV Navigation

An assumption was a previous policy (nonexpert helper) that may be of poor performance is accessible to the educational agent. The previous policy plays the role of guiding the agent in exploring the state area by reshaping the behavior policy used for environmental interaction; for details, see [85].

5.5. Task Offloading for UAVs

A distributed deep neural network was examined to search out the best offloading approach. Within the planned distributed DNN model, multiple distributed deep neural networks within the same coaching instance were trained, and eventually, for validation. For quicker convergence of the coaching method, used the best generated offloading call, employing a quadratically forced linear program with semi definite relaxation. The experiment demonstrated that the offloading call created by the trained distributed deep neural network can do a near-optimal performance with various system parameter settings [86].

5.6. Energy Efficiency

A new deep model referred to as j-PPO+ConvNTM contains a completely unique Spatio-temporal module ConvNTM to higher model long-sequence spatio-temporal knowledge. A deep reinforcement learning (DRL) model is referred to as “j-PPO”. Wherever it’s the potential to create continuous and distinct action choices at the same time for all UAVs. The study performed an intensive simulation to indicate its illustrative movement trajectories, hyperparameter standardization, and ablation study, and compare with four different baselines [87].

5.7. Human Target Search, and Detection

The application of developing nursing autonomous closed-circuit television victimization associate in tending UAV to spot a given target and/or objects of interest within the parcel over that it flies. The system may be employed in rescue operations, particularly in remote areas wherever physical access is tough. Optimum algorithms search and notice the target from the given search space. In recognition of the target, the UAV will either be wont to hold its position thus to have a video feed of the target or come to its base station. The coordinates are calculable victimization GPS modules or relay the GPS location to the bottom station [88].

5.8. Power Allocation in UAV Networks

To realize the joint optimum policy of mechanical phenomenon style and power allocation, the deep reinforcement learning approach is investigated. The continual action area of the MDP model, a deeply settled policy gradient approach, is bestowed. The proposed learning algorithmic program considerably scaled back the knowledge and packet drop rate, compared to the baseline greedy algorithms; refer to [89] and [90].

5.9. Multi-UAV Navigation

A conducted in-depth simulation was done and located the suitable set of hyperparameters, as well as expertise replay buffer size. Variety of neural units for fully connected hidden layers of actor, critic, and their target networks, and therefore the discount issue for memory the longer-term reward. The prevalence of the projected model over the progressive a deep reinforcement learning-based energy efficient control for coverage and connectivity approach supported deep settled policy gradient, and three different baselines [91].

5.10. Real-Time Trajectory Planning

Network states of battery levels and buffer lengths of the ground sensors, channel conditions, and location of the UAV. A flight trajectory planning optimization is formulated as a Partial Observable MDP (POMDP), where the UAV has partial observation of the network states. In practice, the UAV-enabled sensor network contains numerous network states and actions in POMDP while the up-to-date knowledge of the network states is not available at the UAV. An onboard deep RL algorithm to optimize the real-time trajectory planning of the UAV [92].

5.11. Resource Management

Formulation of the resource management of UAV-assisted WPT and information assortment as Mark off call method. Wherever the states carry with their battery levels and information queue lengths of the IoT nodes, channel qualities, and positions of the UAV. A deep Q-learning primarily based on resource management is planned to attenuate the general information packet loss of the IoT nodes. This is optimally deciding the IoT node for information assortment and power transfer, and the associated modulation theme of the IoT node [93]; details are illustrated in Table 3.
Table 3. Challenges of UAV, state-of-the-art proposed models, and loopholes

<table>
<thead>
<tr>
<th>Challenges</th>
<th>Proposed Model</th>
<th>Results</th>
<th>Loopholes</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resource Allocation</td>
<td>DDPG-FRAS</td>
<td>Gradual converge and Reduced Packet loss</td>
<td></td>
<td>[82]</td>
</tr>
<tr>
<td>Fresh Data Collection</td>
<td>MDP Model</td>
<td>Reduced weighted sums of Age of Information</td>
<td>Increased poor data quality, low emphasis was put on maintaining data quality.</td>
<td>[83]</td>
</tr>
<tr>
<td>Trajectory Design &amp; mode Selection</td>
<td>MDP with consideration to sensing and transmission observation</td>
<td>Higher total utilities</td>
<td>Increased computational rates</td>
<td>[84]</td>
</tr>
<tr>
<td>Autonomous UAV Navigations</td>
<td>MDP with sparse Rewards</td>
<td>Impressive navigation polices</td>
<td>Traditional models applied</td>
<td>[85]</td>
</tr>
<tr>
<td>Tasks Offloading</td>
<td>A distributed Deep Neural Networks</td>
<td>Close optimization performance</td>
<td>Current models are not used in comparisons</td>
<td>[86]</td>
</tr>
<tr>
<td>Energy Efficiency</td>
<td>j-PP+ConvNTM Model</td>
<td>Hyperparameter tuning</td>
<td>Data drop transfer rate</td>
<td>[87]</td>
</tr>
<tr>
<td>Human Target Search</td>
<td>Autonomous UAV and the Deep Learning</td>
<td>Simplified object detection</td>
<td>Increased complexities during detections</td>
<td>[88]</td>
</tr>
<tr>
<td>Human Detection</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trajectory Designs</td>
<td>MDP Model</td>
<td>Higher system capacity and faster processing speeds</td>
<td>Only Traditional models were used, and low data capacity transfer</td>
<td>[89]</td>
</tr>
<tr>
<td>Power Allocation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Efficient Data Collection</td>
<td>UAV flight Trajectory Model</td>
<td>Reduced Packet drop rates</td>
<td>Increased computation rate</td>
<td>[90]</td>
</tr>
<tr>
<td>Energy-efficient Multiple UAV Navigation</td>
<td>A decentralized deep Reinforcement Learning</td>
<td>Improved efficiency</td>
<td>Alternative machine learning was not considered for better performance evaluation</td>
<td>[91]</td>
</tr>
<tr>
<td>Trajectory Planning</td>
<td>A partial Observable MDP</td>
<td>Optimized real-time trajectory planning</td>
<td>Time complexities were not considered</td>
<td>[92]</td>
</tr>
<tr>
<td>Network Resource Management</td>
<td>UAV-Assisted WPT Framework.</td>
<td>Associative modulation schemes</td>
<td>Low complex model Comparison</td>
<td>[93]</td>
</tr>
</tbody>
</table>
5.14. Trajectory Design & Power Allocation

The non-convexity feature of joint flight style and power allocation issue. It's complex to realize the best joint policy in multi-UAV networks. A multi-agent deep RL-based approach was given attention utmost, a long network utility. This design satisfied the user equipment's QoS necessities; see [106]. Other approaches include partially explainable big data [94], secure UAV Communications [95], UAV pursuer and evader problem [96], Automatic Recognition [97], Path planning [98], Trajectory Design and Generalization [99], energy-efficient UAV [100-101], Multi-UAV Target-Finding [102], Real-time Data Processing [103], UAV Navigation [104], Anti-Intelligent UAV Jamming [105]. Summarized challenges of UAV, hi-tech proposed models, and loopholes are illustrated in Table 4.

Table 4. Challenges of UAV, state-of-the-art proposed models, and loopholes

<table>
<thead>
<tr>
<th>Challenges</th>
<th>Proposed Model</th>
<th>Results</th>
<th>Loopholes</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>Explainable big data</td>
<td>A double Dueling Deep Q-Learning Neural Network Model</td>
<td>Increased Interpretability</td>
<td>Time complexities are considered</td>
<td>[94]</td>
</tr>
<tr>
<td>Pursuers and Evader issue</td>
<td>Takagi-Sugeno fuzzy control system</td>
<td>Experiment replay</td>
<td>Only “And” was used</td>
<td>[95]</td>
</tr>
<tr>
<td>Secure Communication</td>
<td>A multi-agent deep RL Approach</td>
<td>Solved multi-UAV cooperation trajectory design</td>
<td>Low communications scenarios were used.</td>
<td>[96]</td>
</tr>
<tr>
<td>Auto-recognition</td>
<td>Neural Networks Model</td>
<td>Fine-tuning</td>
<td>Low communications scenarios were used.</td>
<td>[97]</td>
</tr>
<tr>
<td>Path Planning</td>
<td>MDP</td>
<td>Increased effectiveness, increased efficiency</td>
<td>Only conventional algorithms were compared too</td>
<td>[98]</td>
</tr>
<tr>
<td>Trajectory Design</td>
<td>A proximal Policy Optimization</td>
<td>a proximal policy optimization (PPO) - based algorithm (MSR-PPO) shows improvement, RT-PPO depicted efficiency in general performance</td>
<td>Only Deep Reinforcement Learning was used</td>
<td>[99]</td>
</tr>
<tr>
<td>Interface Management</td>
<td>Key Performance indicators</td>
<td>Spectral efficiency, less interface to a user</td>
<td>No other Complex learning scheme was compared with</td>
<td>[100]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Less Interference to user</td>
<td></td>
<td>[101]</td>
</tr>
<tr>
<td>Target finding</td>
<td>Partially Observation MDP</td>
<td>Increased number of Application</td>
<td>No current framework was used</td>
<td>[102]</td>
</tr>
<tr>
<td>Data Processing</td>
<td>Deep Q-Network Learning</td>
<td>Low interruption</td>
<td>Real-time compression</td>
<td>[103]</td>
</tr>
<tr>
<td>UAV Navigation</td>
<td>MIMO Techniques</td>
<td>Good coverage, convergence</td>
<td>Low model comparison</td>
<td>[104]</td>
</tr>
<tr>
<td>UAV Jamming</td>
<td>Partially Observation MDP</td>
<td>Defense strategy</td>
<td>Low time complexity</td>
<td>[105]</td>
</tr>
<tr>
<td>Power Design Allocation</td>
<td>Optimal Policy for JTDPA</td>
<td>Higher Network utility</td>
<td>Increased computational complexities</td>
<td>[106]</td>
</tr>
</tbody>
</table>
6. **Technical Trending of UAV**

In this section, the study focuses on the most current identified trends that the Unmanned Aerial Vehicles are facing, based on the response to the proposed model.

6.1. **Tasking in Complex Environments**

An imitation increased DRL learns the underlying complementary behaviors of UAVs from an indication dataset. Some straightforward situations with non-optimized methods. Analysis of the imitation increased the Deep RL approach during a visual game-based simulation platform. Conducted experiments showed effectiveness that allowed the coalition and cost-effectively accomplishes tasks [107].

6.2. **Manoeuvre Decision**

A high dimensional state and action area which needs a vast computation load for DQN coaching exploitation of ancient strategies. A phased coaching technique, known as "Basic-confrontation", relies on the concept. Learned from easy to complicated setup to assist cut back the coaching time whereas obtaining suboptimal however economical results. The coaching technique facilitated the UAV to come through autonomous call within the air combats [108].

6.3. **Minimum Throughput Maximization**

The joint remote-controlled aerial vehicle flight designing and time resource allocation for minimum output maximization during a multiple UAV-enabled wireless high-powered communication network (WPCN). The UAVs perform as base stations to broadcast energy signals within the downlink to charge IoT devices. Whereas, the IoT devices send their freelance data within the transmission by utilizing the collected energy [109].

6.4. **Persistent Communication Service, and Fairness**

To deal with the mix downside of third-dimensional quality of multiple UAVs and energy filling programming, that ensures energy-efficient and honest coverage of every user in an exceedingly massive region and maintains the persistent service. The model reveals that UC-Deep settled Policy Gradient shows an honest convergence and outperforms alternative programming algorithms in terms of knowledge volume, energy potency, and fairness [110].

6.5. **Segmentation Method**

The accuracy of single-tree purpose cloud segmentation of deep learning ways is quite ninetieth, and the accuracy is much higher than ancient flat image segmentation and purpose cloud segmentation. As a biological science survey tool, encompasses a massive house for promotion and attainable future development [111, 112].

6.6. **Wireless Networking**

Since the state of the UAV movement drawback has giant dimensions, studies intended to use the proposed dueling deep Q-network formula that familiarizes neural networks and dueling structures with Q-learning. The simulation results demonstrate the planned movement formula is ready to trace the movement of global technology systems. This gets period optimum capability and subjects it to the coverage constraint [113].

6.7. **Online Power Transfers, and Data Collections**

A Mark call method with the states of battery level and information queue length of devices, channel conditions, and waypoints given the mechanical phenomenon of the UAV. The model irrefutable showed that DRL algorithmic rule reduces the packet loss by a minimum of 69.3%, as compared to existing non-learning greedy algorithms [114].

6.8. **Efficient Animal Detection**

The CNN scores within the supply information set are accustomed to ranking the samples per their chance of being animals, and this ranking is transferred to the target information set. A new window cropping strategy that accelerates sample retrieval was used. The experiments showed that with each way combined, oracle-provided labels are enough to seek out nearly eighth of the animals in difficult sets of UAV pictures, beating all baselines by a margin [115].

6.9. **Target Tracking**

A coarse-to-fine deep theme to handle the ratio variation in UAV training. The coarse-tracker 1st produces an associate initial estimate for the target object. Then a sequence of actions square measure learned to fine-tune the four boundaries of the bounding box. It is trained together by sharing the perception network with associate end-to-end reinforcement learning design. The benchmark aerial information set proves that the approach outperforms existing trackers and produces important accuracy gains in coping with the ratio variation in UAV trailing [116].

6.10. **Node Positioning**
The placement of the relay node is set by each traffic quality-of-service necessities and the link conditions. The study intended to style a replacement queuing model, referred to as a multi-hop priority queue, to investigate the realizable QoS performance through multi-hop queue-to-queue accumulation modeling. To handle dynamic swarm topology and time-varying link conditions, the deep Q networking-based UAV link choice is computed within the powerful craft that maintains the graphs of the swarm topology [117]. Table 5 clearly depicts the current state-of-the-art challenge trends, and loopholes of UAVs.

Table 5. State-of-the-art challenge trends, and loopholes of UAV

<table>
<thead>
<tr>
<th>Challenges</th>
<th>Proposed Model</th>
<th>Results</th>
<th>Loopholes</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>Image multi-labeling</td>
<td>Neural Network</td>
<td>Increased Interpretability</td>
<td>Time complexities are considered</td>
<td>[107]</td>
</tr>
<tr>
<td>Urban traffic density estimation</td>
<td>Random Sampling DNN</td>
<td>Experiment replay</td>
<td>Only “And” was used</td>
<td>[108]</td>
</tr>
<tr>
<td>Situational Awareness system</td>
<td>Person-Action-Locator</td>
<td>Solved multi-UAV cooperation</td>
<td>Low communications scenarios were used.</td>
<td>[109]</td>
</tr>
<tr>
<td>Tasking</td>
<td>Imitation Augmented Deep RL Model</td>
<td>Fine-tuning</td>
<td>Low communications scenarios were used.</td>
<td>[110]</td>
</tr>
<tr>
<td>Maneuver Decision</td>
<td>MDP</td>
<td>Increased effectiveness, increased efficiency</td>
<td>Only conventional algorithms were compared too</td>
<td>[111]</td>
</tr>
<tr>
<td>Minimum Throughput Maximum</td>
<td>A proximal Policy Optimization</td>
<td>MSR-PPO shows improvement, RT-PPO depicted efficiency in general performance</td>
<td>Only Deep Reinforcement Learning was used</td>
<td>[112]</td>
</tr>
<tr>
<td>Persistent communication services &amp; fairness</td>
<td>The deep deterministic policy gradient (UC-DDPG) Model</td>
<td>Convergence and output forms</td>
<td>Increased data volume, and increased energy</td>
<td>[113]</td>
</tr>
<tr>
<td>Segmentation Method</td>
<td>Patch-based MVS Algorithm</td>
<td>Availed Forestry Survey tools</td>
<td>Low coverage rates</td>
<td>[114]</td>
</tr>
<tr>
<td>Crop Row Detection</td>
<td>CNN</td>
<td>Robust Performance</td>
<td>No one model was compared</td>
<td>[115]</td>
</tr>
<tr>
<td>Power Transfer</td>
<td>MDP</td>
<td>Low packet low</td>
<td>Over 39.8% was trusted compared to greedy learning</td>
<td>[116]</td>
</tr>
<tr>
<td>Efficient Animal Detection</td>
<td>CNN</td>
<td>Increased identification of Animals to over 80%</td>
<td>Low baseline</td>
<td>[117]</td>
</tr>
<tr>
<td>Target Tracking</td>
<td>A course to a fine deep scheme</td>
<td>Increased accuracy gains</td>
<td>Low ratio of UAV tracking variation</td>
<td>[118]</td>
</tr>
<tr>
<td>Node Positioning</td>
<td>Deep Q- Network</td>
<td>Efficiency</td>
<td>Limited position models</td>
<td>[119]</td>
</tr>
<tr>
<td>Beam selection</td>
<td>3-Dimensional Fast Machine Learning Algorithm</td>
<td>optimized utilization of biomedical engineering technology systems</td>
<td>More realistic medical situations are desired in the comparisons</td>
<td>[120]</td>
</tr>
</tbody>
</table>

[121]
6.11. Content Transmission

The projected rule permits every SBS to predict the users' responsibility, therefore, to realize the optimum contents to cache and content transmission format for every cellular-connected UAV as per the model. Results showed the varied network factors that impact content caching and content transmission format choice. Simulation results expressed that the projected rule yields close to 26%, and 14.7% gains, in terms of responsibility compared to Q-learning and a random caching method [118].

6.12. Interference Management

Based on echo state network cells projected, the introduced deep echo state specification is trained to permit every UAV to map every observation of the network state to associate action, to minimize a sequence of time-dependent utility functions. The results show that a higher wireless latency per UAV and rate per ground user requires a variety of steps that include a heuristic baseline [119].

6.13. Intelligent Monitoring

To facilitate the implementation of the system, addressing triple main challenges of deep learning in vision-based power cable inspection. Deep residual networks to detect tiny elements and faults in the planned system is quick. To correct in police work common faults on power cable elements, together with missing high caps, cracks in poles and cross arms, peckerwood injury on poles, and decay injury on cross arms [120].


To avoid the advanced feature extracting method and understand the classification of UAV-to-ground vehicles in several things. Transfer-learning of pre-trained DCNNs is achieved through mistreatment of measured knowledge, and classification underneath numerous conditions is complete mistreatment of the new-trained network. Once there's no noise, the classification accuracy of dual sorts of physicist signals, 3 sorts of physicist signals, four sorts of physicist signals, and five sorts of physicist signals has reached 100%, 97%, 97%, and 96% [121].

6.15. Situational Awareness

The Person-Action-Locator system addresses the primary issue by analyzing the video feed aboard the UAV, battery-powered by a supercomputer-on-a-module. As support for human operators, the person-action-locator system depends on deep learning models to automatically find folks and acknowledge their actions in close to a period [122].

6.16. Urban Traffic Density Estimation

Advanced deep neural network-based vehicle detection and localization, type (like the car, bus, and truck) recognition, tracking, and vehicle counting overtime. The study proved that the enhanced single shot multi-box detector, outperforms other DNN-based techniques and that deep learning techniques are more effective than traditional computer vision techniques in traffic video analysis [123]. Table 6 demonstrates the UAV trending key points and Modern UAV Models.

7. Future Research Directions

In this section, we provide some future research directions.

7.1. Advanced Self-Driving Cars

The world has so far enjoyed this respective regardless of its limitation that is to be worked on per RL. Generating and preserving atlasts aimed at self-driving cars is a major problem for the car to understand, lashing necessitates numerous multifarious societal interactions. It is still threatening when it comes to robotics, there will desire to strategize principles already to feed the automatic cars categorically are complex. Cyber-security and anomaly faults are prospectively possible issues via a conquerable factor that needs attention too, radar intrusions, accident liability, unmanned traffic situations, weather, and road circumstances [141-147].

7.2. Smart Traffic Signals

Smart traffic open a big factor in smart cities and their associated benefits that the future demands. The following are predictable to happen due to RL, including the smart way of managing traffic signals, and advanced traffic management radars. Smart traffic networking increased adaptive keen high-techs with the automatic expressway management system, indications for multifaceted metropolitan highway networks among others are considered [142-146].

7.3. Fully Automated Factory

High-tech factories are semi-automated, based on the advancement of the technologies. RL is hoped to be enhanced. For instance, there exist several factors that have influenced the production appreciation goes to robotics that has RL in them like user-guided vehicles to empower production flows. Replacing an increased number of manpower and doubling capacities [147]. Studies seek to compute the following increased computation, increased scalability, and, improved real-
time monitoring. The amplified smart technology available provides real-time factory details and the identification of real-time diagnosis of the faults. Increased human to machine, machine-to-machine interaction, increased speed, and user-friendly approaches; see the details [148]. Table 6 demonstrates the UAV trending key points and Modern UAV Models.

<table>
<thead>
<tr>
<th>Parameter Type</th>
<th>UAV Entity</th>
<th>Proposed Model</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deep Reinforcement Learning</td>
<td>Autonomous Target Search</td>
<td>Snake Algorithm</td>
<td>[122]</td>
</tr>
<tr>
<td>Deep Learning</td>
<td>Situational Awareness</td>
<td>Person-Action-Locator</td>
<td>[123]</td>
</tr>
<tr>
<td>Game-based</td>
<td>Complex Tasking Environments</td>
<td>Imitation Augmented Deep Reinforcement Learning</td>
<td>[125]</td>
</tr>
<tr>
<td>Deep Q-Network Learning</td>
<td>Maneuver Decision</td>
<td>DQN training Model</td>
<td>[126]</td>
</tr>
<tr>
<td>Deep Q-Learning</td>
<td>Minimum Throughput Maximization</td>
<td>A multiple UAV-enabled wireless powered communication network</td>
<td>[127]</td>
</tr>
<tr>
<td>Deep Learning</td>
<td>Persistent Communication Service, and Fairness</td>
<td>Deep Deterministic Policy Gradient</td>
<td>[128]</td>
</tr>
<tr>
<td>Cloud deep learning</td>
<td>Segmentation Methods</td>
<td>A forest sparse point cloud and patch-based MVS algorithm</td>
<td>[129]</td>
</tr>
<tr>
<td>A convolutional neural network</td>
<td>Detection</td>
<td>CRowNet, S-SegNet, and HoughCNet</td>
<td>[130]</td>
</tr>
<tr>
<td>Q-learning</td>
<td>Wireless Networking</td>
<td>Constrained Markov Decision Process</td>
<td>[131]</td>
</tr>
<tr>
<td>Q-Learning</td>
<td>Markov Decision Process</td>
<td>Online Power Transfers, Data Collections</td>
<td>[132]</td>
</tr>
<tr>
<td>Ratification Intelligence</td>
<td>Efficient Animal Detection</td>
<td>Transfer Sampling</td>
<td>[133]</td>
</tr>
<tr>
<td>Deep Learning</td>
<td>Target Tracking</td>
<td>A coarse-to-fine deep scheme</td>
<td>[134]</td>
</tr>
<tr>
<td>A deep Q-Learning model</td>
<td>Node Positioning</td>
<td>A multi-hop priority queue</td>
<td>[135]</td>
</tr>
<tr>
<td>A distributed deep learning algorithm</td>
<td>Content Transmission</td>
<td>UAV-SBS backhaul traffic</td>
<td>[136]</td>
</tr>
<tr>
<td>A deep reinforcement learning algorithm</td>
<td>Interference Management</td>
<td>A perfect Nash equilibrium upon convergence</td>
<td>[137]</td>
</tr>
<tr>
<td>Deep Residual Networks</td>
<td>Intelligent Monitoring</td>
<td>A novel automatic autonomous vision-based power line inspection system</td>
<td>[138]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Singular Values in UAV Automation control</td>
<td>[139]</td>
</tr>
<tr>
<td>Deep intensive learning</td>
<td>Autonomous Target Search</td>
<td>Ancient puzzle game Nokia Snake</td>
<td>[140]</td>
</tr>
</tbody>
</table>
7.4. Smart Prosthetic Limbs

In beauty aspects, RL is more principal. Studies have demonstrated electronic skins, for example Johns Hopkins created a skin with abilities of sensitivity where it fits over the natural finger of the prosthetic hands [149]. The e-dermis is alleged to have been made from the fabric that is entrenched with sensors, these sensors send signals to nerves at the end of an amputated limb. The e-dermis is modelled on the skin to react to pressure [150].

7.5. Intelligently Trade Stocks

Trades and trading are key in influencing the development of RL, as illustrated. Several issues are identified in this element, including smart trade decision-making. AI-powered algorithmic trading possibilities, intelligent daily trade drone deliveries, and automated trading systems are still semi-machine. The AI-trading signals, market trading bots, intelligent trade financing, and intelligent self-sufficient trading application among others are still having technical issues altogether [151].

8. Conclusion

In this paper, RL among the ML techniques was examined in detail. The recent advancements in the technologies, techniques, and future research directions were depicted mainly in the four perspectives. These four perspectives included the Markov decision process, q-learning, learning automata, and artificial neural networks. Also, the study detailed and demonstrated the main issues around ML within figures, and tables. A concise perspective on identified arenas in the analysis of RL is examined too. The study revealed that various methods capitivated on discontinuous policy values instead of modifying other gears in an exact state of intellectual. Furthermore, the study evidenced that contemporary research adopted by researchers from diverse research disciplines engages in AI-enabled models, architectures, and frames among others.

Acknowledgements.

The authors in nutshell would like to distinguish the support and comments shared with us from the computer engineering department members to attain this paper's quality.

Data Availability Statement: All the data that was used to support the results of this study are encompassed within the paper.

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding this study.

References


A reawakening of Machine Learning Application in Unmanned Aerial Vehicle: Future Research Motivation


[40] W. Shafik, "A fast machine learning for beam selection in 5g unmanned aerial vehicle communications” M.S. dissertation, Yazd University, Iran, 2020.


[82] K. Li, Y. Emami, W. Ni, E. Tovar, and Z. Han, “Onboard deep deterministic policy gradients for


A reawakening of Machine Learning Application in Unmanned Aerial Vehicle: Future Research Motivation


