



Article

Reinforcement Learning-Enabled UAV Itinerary Planning for Remote Sensing Applications in Smart Farming

Saeid Pourroostaei Ardakani ^{1,*}  and Ali Cheshmehzangi ² 

¹ School of Computer Science, University of Nottingham Ningbo, Ningbo 315100, China

² Department of Architecture and Built Environment, University of Nottingham Ningbo, Ningbo 315100, China; Ali.Cheshmehzangi@nottingham.edu.cn

* Correspondence: saeid.ardakani@nottingham.edu.cn; Tel.: +86-(0)-574-8818-0000 (ext. 8102)

Abstract: UAV path planning for remote sensing aims to find the best-fitted routes to complete a data collection mission. UAVs plan the routes and move through them to remotely collect environmental data from particular target zones by using sensory devices such as cameras. Route planning may utilize machine learning techniques to autonomously find/select cost-effective and/or best-fitted routes and achieve optimized results including: minimized data collection delay, reduced UAV power consumption, decreased flight traversed distance and maximized number of collected data samples. This paper utilizes a reinforcement learning technique (location and energy-aware Q-learning) to plan UAV routes for remote sensing in smart farms. Through this, the UAV avoids heuristically or blindly moving throughout a farm, but this takes the benefits of environment exploration–exploitation to explore the farm and find the shortest and most cost-effective paths into target locations with interesting data samples to collect. According to the simulation results, utilizing the Q-learning technique increases data collection robustness and reduces UAV resource consumption (e.g., power), traversed paths, and remote sensing latency as compared to two well-known benchmarks, IEMF and TBID, especially if the target locations are dense and crowded in a farm.

Keywords: UAV; reinforcement learning; Q-learning; path planning; remote sensing



Citation: Pourroostaei Ardakani, S.; Cheshmehzangi, A. Reinforcement Learning-Enabled UAV Itinerary Planning for Remote Sensing Applications in Smart Farming. *Telecom* **2021**, *2*, 255–270. <https://doi.org/10.3390/telecom2030017>

Academic Editors: Sotirios K. Goudos, Demosthenes Vouyioukas, Shaohua Wan, Paolo Rocca and Achilles D. Boursianis

Received: 30 April 2021

Accepted: 22 June 2021

Published: 1 July 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 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

Remote sensing applications aim to remotely capture and report environmental data samples using sensory devices for further processing and/or decision making. Robust, fast, and accurate data collection plays a critical role in agriculture remote sensing applications. This means farming and/or natural resources are threatened/wasted if late (out-of-date), meaningless and/or inaccurate sensory recordings are reported [1]. For example, farm animals fall into danger and/or are threatened by environmental risks if they are not continuously monitored. According to [2], there are two paradigms to address remote sensing applications: client/server and mobile agent (MA). The former deploys a remote sensing infrastructure to collect and forward environmental data samples through Zigbee, Bluetooth and/or internet links (e.g., 5G). In contrast, the latter forwards a mobile object (e.g., drones) for sensory data collection throughout a farm. However, MA remote sensing increases the deployment cost and risk, especially in large and wide areas such as farms [3].

Unmanned Aircraft Vehicles (UAVs) are increasingly used for environmental remote sensing in urban and rural areas such as farms [4]. There are a number of UAV remote sensing applications such as animal behavior monitoring, farm surveying, and herd tracking [5]. UAVs are usually equipped with sensory devices—mainly cameras to collect and report high-resolution environmental data samples. They utilize GPS information on-the-fly and move without a human pilot on-board. However, UAVs can be navigated by users using remote-controllers.

UAV path planning aims to form itineraries to forward either a single or multiple UAVs to do a particular mission such as remote sensing. Path planning algorithms usu-

ally suffer from restricted power (and bandwidth) resources, which provide them short traveling distance and limited communication and computation capacities [6]. Due to this, remote sensing UAVs should fly through best-fitted and cost-effective routes to avoid overlapped, duplicated, irrelevant and/or out-of-interest data collection [7]. Yet, they need to utilize efficient path planning approaches aiming to minimize the total flight length and avoid random/blind movement [4]. However, UAVs should avoid performing complex computing algorithms (i.e., path planning and remote sensing computation/aggregation) and/or frequent wireless communication (i.e., network interconnections and flight synchronizations/navigations) to conserve battery resources [8]. Hence, there is still a major gap in the usage/utilization of UAVs in some practices, such as in smart farming. This paper aims to respond to this specific matter.

UAV paths are usually planned according to two categories: proactive and reactive [9]. Proactive path planning provides routes in advance. This is a fixed route plan and supports no update during the flight. By this, proactive routing fits no dynamic remote sensing application—mainly mobile object tracking. Reactive routing dynamically establishes the UAV's routes on-the-fly according to the updating mission or changing environment. However, this may result in increased route planning overhead and enhanced UAV power consumption, especially if the mission/environment is frequently or continuously updated.

Machine learning techniques offer UAV path planning a number of benefits—mainly autonomous and reactive routing in remote sensing application [10]. Reinforcement learning (RL) [11] is a machine learning technique that has the potential to be commonly used for real-time decision-making problems such as UAV route planning. By this, UAVs are able to explore the environment and learn the best-fitted and/or cost-effective paths to the target areas for sensory data collection. Indeed, RL-enabled UAV route planning provides a trial-and-error environment interaction paradigm to figure out optimal routes to fly through. Q-learning [12] is a model-free and value-based algorithm of reinforcement learning that works according to an action-reward approach. This can be used for UAV route planning, as this has the potential to return good-fitting routes (maximized reward) according to the actions (i.e., route plan).

This research proposes a location-aware RL-enabled UAV path planning for remote sensing in smart farms. The study's novelty is based on its contribution to optimize the use of UAVs for farming practices. This is explored specifically from a combination of power-aware UAV path planning and location-based reinforcement learning, which also responds to key factors in cost-effective UAV flights and dynamic field labeling and partitioning processes. Such an approach enables us to explore optimized practical methods for remote sensing in the farming applications. This research utilizes the Q-learning technique for reactive path planning on-the-fly over a labeled region-based farm. For this, the algorithm dynamically labels a grid-partitioning farm according to the static or mobile sensory targets such as plant, crop and/or farm animals. In turn, it utilizes an energy-aware exploration–exploitation paradigm to interact with the farm and learn the remote sensing target regions. This returns a table of routes/directions according to a weighting function of path length (euclidean distance) and consuming energy. Finally, it forwards the UAV through the optimal routes (shortest with minimized consumed energy) to capture the sensory recordings. The proposed approach offers the following contributions:

- To take the benefits of RL to interact with the environment and dynamically route the UAV.
- To propose an energy-aware routing by labeling the farm zones as no-fly, safe-fly and target to avoid blind walk and/or loop.
- To use a weighting function of path length and consumed energy to minimize the remote sensing cost.
- To utilize a grid-based field partitioning/labeling to avoid overlapped data collection.
- To compare and contrast the performance of the proposed approach with infrastructure-based and to heuristic UAV path planning to highlight their superiorities and/or inferiorities.

This paper is organized into six sections, as shown here. Section 2 outlines a review of UAV route-planning algorithms and applications in smart farming. This addresses the existing drawbacks and advances in this field of research. Section 3 introduces the RL enabled UAV path planning approach. This highlights the key features and techniques of the proposed approach. Section 4 explains the research simulation plan and shows how the performance of the proposed approach is tested and evaluated. Section 5 discusses the evaluation of the proposed approach according to four key metrics: (1) average End-to-End delay (ETE), (2) average number of captured data, (3) average battery consumption and (4) average traversed distance. The performance of the proposed approach is compared with two well-known conventional protocols IEMF [13] and TBID [14]. Section 6 summarizes the benefits of the proposed approach and highlights the key points of this research that have the potential to be addressed as further work.

2. Literature Review

UAVs are identified as one of the technology enablers related to IoT (Internet of Things) systems [15] and towards enhancing the smart farming applications [16]. In recent years, a growing number of studies focus on UAV-based approaches for smart farming, suggesting sustainable agricultural methods [17], enhancing smart farming communications [18], and transforming traditional farming practices [19]. The combination of UAVs and IoT-based approaches is recognized as a new paradigm [20,21], leading to further research activities and sensory-based approaches to enhanced smart farming. Some studies highlight the UAV-based applications for better communications in agriculture [22], including methods of digitizing agriculture and achieving precision agriculture. As suggested by [23], there is still scope for further development to utilize deep learning, especially when combined with key technological advances such as UAVs. From a much broader understanding, we could see UAVs to be applicable to overcome smart farming challenges; for instance, by using the visible light sensors (RGB sensor) and UACE camera capture [24] to enhance communications in an easy and low-cost way. In other studies, the use of distributed simulation [25] is an effective approach in managing connections and communications, which are then more effective for smart farming practices.

In the area of smart family, we see more studies focusing on UAV integration in the IoT or IoIT systems. Such integration is helpful, as it enables us to optimize the IoT solutions [26], provides better data management for smart farming practices [27], and embraces a wide technological development in the field of smart farming [28]. The architecture model, as described by [29], highlights the role of communication technologies and IoT-based platforms, enabling sensory systems to provide “immediate monitoring and optimisation of crops”. In doing so, we note a range of challenges and shortfalls regarding controls and applications [30], and limitations for multi-sensory systems and technologies, and battery life. Thus, this study aims to address some of these common limitations and help to consider ways that could be used for optimizing the UAV application in current and future smart farming practices.

Existing literature already highlights the role of smart farming as a crucial approach to developing sustainable agriculture [31]. More recently, deep reinforcement learning has been used to optimize UAV navigation and to provide better connectivity and communication [32]. This approach has broadened areas of innovation such as modeling and coverage of field areas [33], target tracking [34], and battery management [35]. The cooperation between the management aspect and enhancing the IoT system is useful for UAV-enabled or UAV-based methods, particularly that we could go beyond just the common areas of irrigation, fertilization, and disease and weed detection [36]. An important area is to manage battery life, which could be done by enhancing the navigation and route decision-making of the UAVs. In this study, we cover key aspects of coverage path planning, such as the one identified by [37] and move towards more integrated approaches that include UAVs, IoT, and battery management. The study also highlights the role of reinforcement learning-based approaches in itinerary planning, particularly to be used for smart farming

applications and practices. In doing so, the study suggests optimization methods for routing and pathfinding, which will be directly linked to battery life management for the UAVs when used in the agricultural field.

UAV itinerary planning in smart farms aims to address four key issues: (1) minimizing journey delay, (2) flight distance reduction, (3) increased UAV resource conservation (mainly power), and (4) maximizing the number of collected data samples. For this, a number of route planning approaches have been proposed to forward UAVs through best-fitted and cost-effective routes. They are categorized into two classes including heuristic and infrastructure-based. Heuristic path planning is a widely used technique to reactively find the paths for UAVs to move. They usually utilize a greedy function to find routes according to the applications or user requirements, such as reduced delay. For example, IEMF (Itinerary Energy Minimum for First-source-selection) [13] is a greedy algorithm that heuristically finds the closest visiting location via minimum battery consumption path. This allows UAVs to start from a starting location, and finds the next location with minimum distance and consumed energy. On the other hand, infrastructure-based UAV path planning forwards UAVs through an infrastructure (e.g., tree or chain). Tree-Based Itinerary Design (TBID) [14] proposes a tree-based infrastructure (spanning tree) for mobile agents (e.g., UAVs) to move through and collect sensory data. For this, TBID forms a set of concentric zones around the (single) user access point. In turn, a tree is formed from the most inner zone to the outer ones through minimized euclidean distance links. The inter-zone links form the tree trunk, whereas the intra-zone links shape the tree branches.

RL-enabled UAV route planning approaches offer remote sensing applications a number of benefits as compared to heuristic and infrastructure-based path planning. Heuristic routing addresses increased UAV resource consumption and data collection latency, especially when the field (e.g., farm) is wide. This forwards UAVs to find the path according to a greedy function with no environment exploration and interaction. On the other hand, infrastructure-based path planning needs to set-up a routing infrastructure for the UAV to go through. This results in significantly increased UAV resource consumption, especially if the environment is highly dynamic. Due to these issues, this paper aims to design a RL-enabled path planning algorithm through which UAVs interact with the environment and learn how to move throughout the field to complete the mission with minimized recourse consumption or delay, and maximized data collection.

3. Research Method

This paper aims to utilize reinforcement learning—mainly Q-learning techniques to route UAVs for remote sensing in farms. This allows UAVs to learn the environment and find the best-fitted and/or cost-effective paths according to an action-reward fashion. Indeed, UAVs autonomously move throughout the farm to capture sensory recordings with minimum cost (e.g., power) and delay. The key objectives of this research are outlined as: (1) minimize path length, (2) decrease data collection delay, (3) reduce UAV's power consumption, and (4) increase the number of captured data samples.

3.1. Environment Model

This research addresses an environment (farm) model consisting of three keys: sensory points (targets), UAV, and base station. Sensory Points (SPs) are target locations in the farm that should be visited by UAVs for remote sensing. They contain interesting agricultural data such as soil PH, crop growth and/or animal nesting, which should be collected and reported by using UAVs. SPs can be either static or mobile. The former is a particular fixed location in the farm (e.g., sensing plant/crop growth), whereas the latter moves throughout the field (e.g., monitoring the behavior of animal herds).

UAV moves over the farm to visit SPs and remotely collect/aggregate agricultural data. This is highly power constrained, and equipped with on-board cameras and/or sensors to capture data samples and a GPS to find the location information. To reduce power consumption, the UAV avoids continuously transmit taken images and/or collected

data via wireless/internet links to the user's access point/server. However, it utilizes a lightweight aggregation function to combine collected data samples and report the aggregated results when it returns home.

The base station is the end-user access point that manages remote sensing missions. This factor usually has no power or communication constraints. Base station forwards the UAV into the farm for remote sensing journey, and collects sensory recordings reported by the UAV when it returns to deliver the collected data and/or re-charge the battery.

As Figure 1 shows, the farm is partitioned as a grid according to the acquired location information from GPS. Each partition is labeled as no-fly, safe-fly, or SP zones according to the application. SP zones are the target areas to be visited by the UAV for remote sensing. Safe-fly zones have no data for the UAV to collect. However, a safe-fly zone might become an SP zone if a mobile SP (e.g., animal herds) moves into it. The UAV should avoid flying over no-fly zones, whereas it should minimize the flight time over safe-fly zones. The partitions are discovered, and the labels are updated during the environment exploration phase.

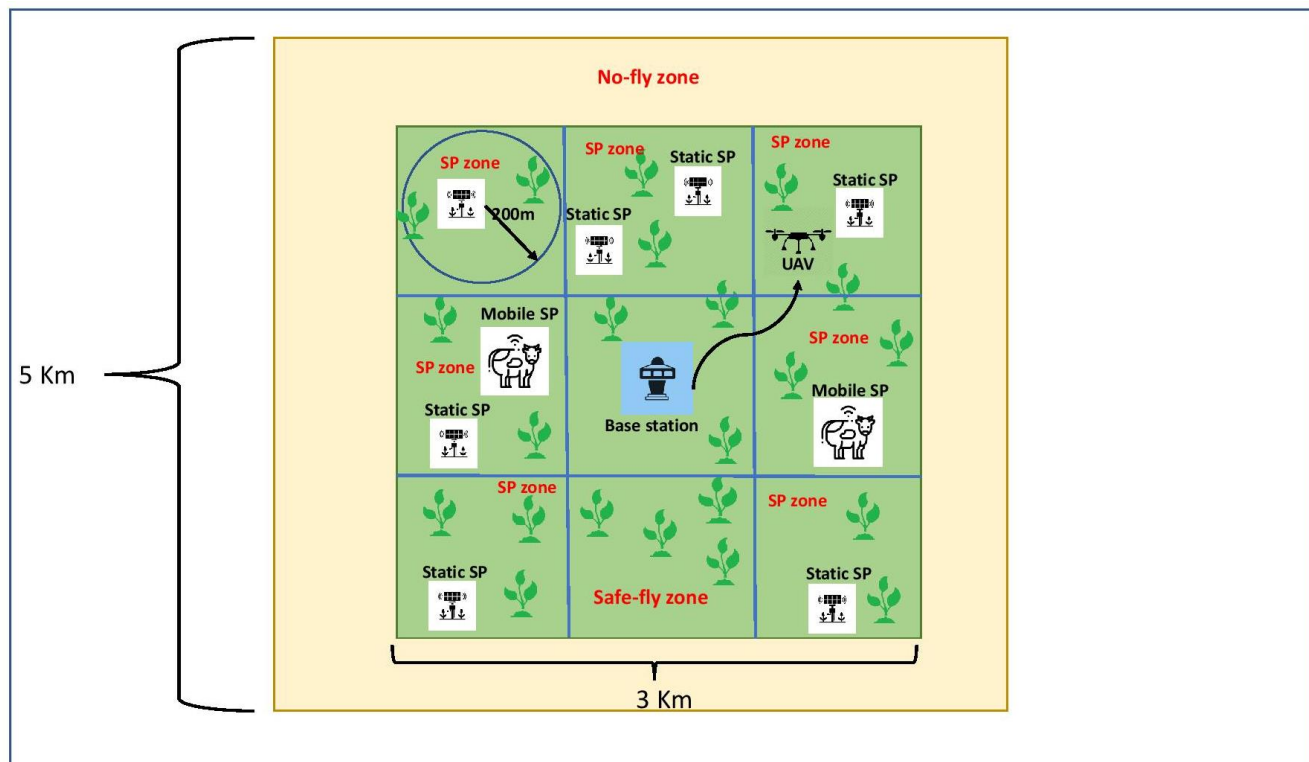


Figure 1. The farm model.

3.2. Reinforcement Learning

Reinforcement learning (RL) [11] is a machine learning technique that allows agents to learn suitable behaviors through environmental interactions. This is widely used to solve problems aiming to achieve specific goals by using interaction-enabled learning strategies.

Q-learning [12] is a model-free and value-based algorithm of reinforcement learning. This works according to an action-reward approach. Q-learning updates a Q-table (including a number of Q-values) according to taken actions until the best-fitted reward is achieved. For this, a Q value, which is considered as an agent, takes action A at a state S to get a particular reward according to the application. Q-learning works in an iterative fashion, which increasingly updates the Q-values according to the taken actions until the best (or required) reward is achieved. The results of Q-values update Q-table as Table 1 shows.

Table 1. Q-table.

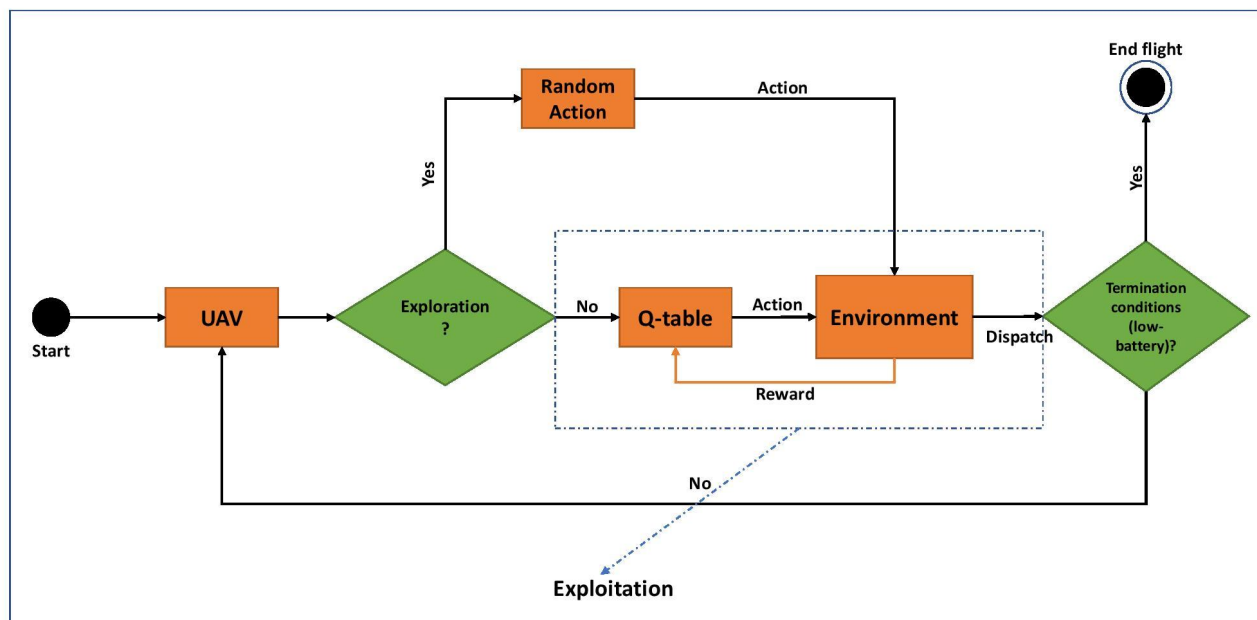
State	Action ₁	Action ₂	...	Action _n
S ₁	Q(S ₁ , A ₁)	Q(S ₁ , A ₂)	...	Q(S ₁ , A _n)
S ₂	Q(S ₂ , A ₁)	Q(S ₂ , A ₂)	...	Q(S ₂ , A _n)
...
S _n	Q(S _n , A ₁)	Q(S _n , A ₂)	...	Q(S _n , A _n)

Equation (1) formularizes Q-value computation according to the taken action and for each state. According to this, $\alpha \in (0, 1)$ is the learning rate that allows the algorithm to explore the environment faster or slower, R_t is the action reward received at time t , and $\gamma \in (0, 1)$ is the discount future reward factor that controls the sum of an infinite number of rewards to stay finite.

$$Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \alpha [R_t + \gamma \max_{A+1} Q(S_{t+1}, A_{t+1}) - Q(S_t, A_t)] \quad (1)$$

3.3. Proposed Approach

This research focuses on location-aware RL-enabled UAV path planning aiming to autonomously learn/find best-fitted and cost-effective flight directions/paths to optimize the results of UAV path planning for remote sensing such as reduced data collection delay and UAV power consumption. There are two operational strategies under RL-enabled UAV route planning: exploration and exploitation. The former allows the UAV to take a random path to fly toward the closest SP, whereas the latter forwards the UAV according to what was learned (e.g., achieved reward values) from previous flights. Figure 2 shows how a UAV takes the benefits of RL (Q-learning) to explore and exploit the environment (e.g., farm).

**Figure 2.** Exploration–exploitation for the RL enabled UAV.

RL assumes UAV as an agent to find optimal routes through constant “trial-and-error” training/interaction with the environment (e.g., the farm and location of SPs). The following outlines the key components of RL enabled UAV path planning approach:

- Objective: the UAV flies over a farm aiming to visit a constant number of SPs through the shortest path (minimum Euclidean distance) and with minimized delay. Each journey/path has a maximum flight time (e.g., 20 min) to meet the battery constraints.

- States: UAV's location and residual battery.
- Action: flight direction and/or route.
- Environment: a map of labeled regions (e.g., SP or no-fly), the reward algorithm, and mission termination function.
- Reward: two rewards, including a positive if a SP is visited and a negative if the UAV moves into no-fly zones or loop.
- Termination: if the UAV visits all SPs, or runs out of battery.

The RL-enabled UAV path planning algorithm aims to explore/exploit the farm and update the Q-table with the rewards achieved. The UAV gets positive rewards if it flies over SP zones and captures the remote sensing recordings, whereas it receives a penalty if moves to no-fly areas. To avoid loops, the UAV receives a negative reward if it continuously moves over safe-fly zones for a while with no SP-visit success. This results in reduced UAV power consumption, and data collection delay. The UAV explores the farm and moves until no SP is left or residual battery drops below a threshold. Otherwise, it returns to the base station to re-charge the battery, deliver the remote sensing recordings and plan the next journey. The remote sensing journey is finished if all SPs are visited. Figure 3 depicts the UAV route planning approach.

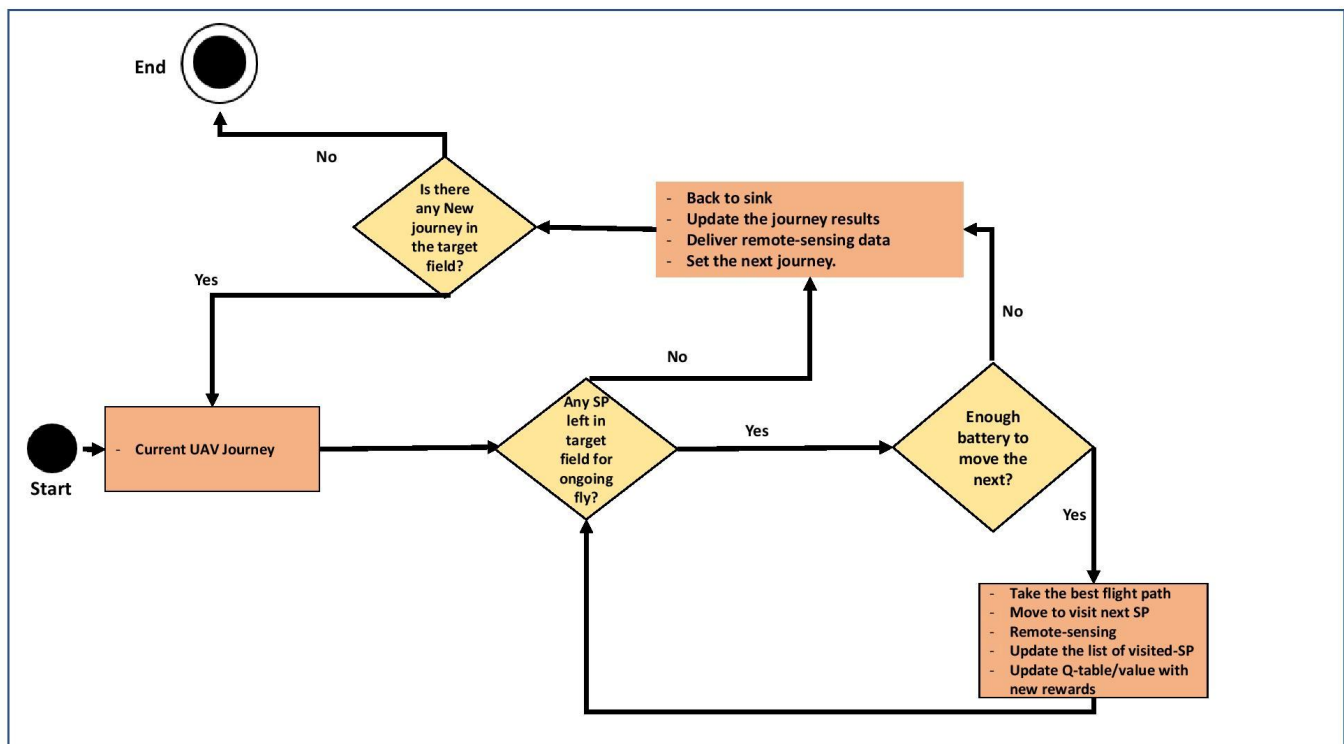


Figure 3. UAV path planning.

The UAV returns the base station to deliver the results in three cases: (1) all SPs are visited and the collected data should be delivered to the base station. The UAV might be able to transmit remote sensing recordings (e.g., photos) to the base station via internet links. However, this results in increased communication overhead and consequently power consumption, especially if the farm is wide, data volume (e.g., image quality) is large and/or SPs are crowded and dense. By this, UAV should return the sink to deliver the data with minimized power consumption, and (2) the UAV's battery meets the threshold. For this, the UAV cannot move forward and visit the next SPs, but it should only return home for battery re-charge, and (3) a (part of) remote sensing mission in a particular/target field/farm is finished. The proposed approach has the potential to support remote sensing with multiple UAVs. Under this, each UAV labels the zones and autonomously detects and/or recognizes its own target areas (e.g., particular SP zones). This should terminate the

remote sensing mission and return it to the base station if there is no SP left unvisited in the allocated target zone. This point addresses multi-UAV remote sensing with no overlapping or uninteresting data collection.

4. Simulation

Simulation is broadly used to implement and test UAV route planning approaches, as real-world UAV routing implementations are expensive and risky. OMNET++ [38] is used in this article to study the performance of the proposed approach according to a simulated remote sensing scenario. This is a component-based network simulator and utilizes INET [39] to support communication models and node mobility.

This simulation aims to test and study the performance of the RL-enabled UAV path planning approach for remote sensing in smart farms. The simulation results are compared with the performance of two well-known UAV (or MA) path planning approaches, including IEMF (a power-aware version of LCF [40]) and TBID, which are widely used in the literature [2,41–43]. Four key metrics are measured to study the performance of the proposed approach: average End-to-End delay (ETE), average number of captured data, average battery consumption and average traversed distance. These metrics are commonly used to evaluate the performance of UAV path planning algorithms [6,9,34,44].

4.1. Simulation Setup

This addresses a 2D simulation to model UAV remote sensing applications in smart farms. Since this is a 2D simulation, UAV's altitude has no impact on the results of the study. However, in a real scenario, UAV can fly with a constant operational altitude and speed during the whole data collection process. As Figure 1 shows, a $3 \times 3 \text{ km}^2$ farm is simulated with three types of network nodes, including UAV, SPs and base station in a field of $5 \times 5 \text{ km}^2$. Through this, the base station is localized in the center of the farm to collect and aggregate remote sensing recordings from the UAV. The central location of the UAV station would enable better accessibility for data collection, more route options, and better flexibility for different directions. Since UAVs are mobile devices, they can move to any other allocated spots for services and maintenance. SPs are either static or mobile (constant speed of 1 m/s) and randomly distributed throughout the farm. They generate a random value as sensory recordings at a particular timestamp.

According to the simulation deployment, a single base station is localized in the center of a farm, which includes a network of randomly distributed SPs. For this, the SP distribution model generates a number of simulation seeds to randomly localize SPs in the farm. By this, the UAV should fly and visit SPs, which are randomly localized in each simulation iteration. This situation is similar to the simulation scenarios, which randomly localize an UAV in a farm with fixed SPs. This minimizes the impact of the UAV's (initial) location on the simulation results and supports a realistic scenario in which the location of SPs randomly changes according to a fixed UAV localization.

The UAV is mobile (with a constant speed of 20 m/s) and moves using a route planning approach. It starts its journey from the base station and utilizes a battery consumption model according to Matrice 200 (0.055 unit/s in ideal/clear-flight conditions) [45]. The UAV is able to remotely sense targets within a limited range of 200 m. This means each (static or mobile) SP forms a circle of 200 m around as SP zone.

This simulation supports IEEE 802.11 and Carrier Sense Multiple Access (CSMA/CA) MAC protocol to address a collision-free wireless channel sharing and communication between the base station and the UAV. Yet, this simulation uses a Line-Of-Sight (LOS) wireless signal propagation model, which assumes no wireless interferences caused by an obstacle (e.g., trees) and/or the environment (e.g., weather).

Statistical power analysis technique [46] is used to compute experiment repetitions required for this experimental design. This technique calculates the number of repetitions according to a target confidence level by measuring the population standard deviation for a (randomly selected) subset of experiments. According to the results, 300 repetitions

are required to achieve a confidence level of 90% confidence based on the simulations of 50 experimental repetitions under this simulation scenario.

A varying number of SPs including 64, 128 and 256 (of which 25% are mobile and 75% are static) is randomly distributed throughout the farm field. The UAV visits the SP zones to collect the value. Table 2 summarizes the simulation setup parameters:

Table 2. Simulation parameters.

Parameter	Range	Parameter	Range
Simulation time	3600 s	network initialization	50 s
repetition	300	ambient noise and obstacles	disabled
Remote sensing range	200 m	SP distribution model	random
Base station location	Farm's center	Area size	3×3 km
MAC	CSMA/CA	Communication protocol	IEEE 802.11
UAV battery threshold	10%	UAV Max flight (one battery)	30 min
Mobile SP speed	1 m/s	Mobile SP ratio	25%

4.2. RL Routing Simulation

This simulation allows studying of the performance of RL enabled UAV path planning for remote sensing applications in smart farms. The simulation addresses a ϵ -greedy routing algorithm aiming to maximize the reward value according to the environment exploration–exploitation paradigm. Through this, UAV starts to explore the area and update the reward using random flight plans. This takes the random actions (depending on ϵ value, which is exploration probability) and moves throughout the farm and receives a positive reward if the route ends at an SP zone. In contrast, a negative reward is recorded if the UAV blindly flies over either a fly or no-fly zone with no SP visit. This algorithm is repeated until either all the SPs are visited or simulation time is finished. Then, this aims to forward the UAV for remote sensing according to an exploitation fashion. According to Algorithm 1, Q-learning values are computed using Equation (1), where either a positive or negative reward (rt) is applied with learning rate ($\alpha = 0.01$), future reward discount factor ($\gamma = 0.1$) and exploration probability ($\epsilon = 0.1$) [47].

Algorithm 1: Q-learning for UAV Routing.

```

/*initialize Q table at base station*/
Initialize Q (S, A);
/*Repeat Q-learning calculation until the whole set of target region
is explored.*/
while  $R_{set} \neq \text{Null}$  do
    /*initialize States for a particular exploring region.*/
    Initialize S; /*Repeat Q-learning value calculation until available
states are addressed.*/
    while  $S_{set} \neq \text{Null}$  do
        Select action (A) from state (S) according to exploration probability ( $\epsilon$ );
        Measure R and observe  $S_{next}$ ;
         $Q(S_t, A_t) \leftarrow \{Q(S_t, A_t) + \alpha [R_t + \gamma \max_{next} Q(S_{next}, A_{next}) - Q(S_t, A_t)]\}$ ;
         $S \leftarrow S_{next}$ ;
    end
end
end

```

5. Results and Discussions

This section outlines and discusses the simulation results to study the performance of RL enabled UAV path planning for remote sensing applications by measuring four route planning metrics including (1) average End-to-End delay (ETE), (2) average number of captured data, (3) average battery consumption and (4) average traversed distance. This compares the results of RL-enabled UAV path planning with two well-known benchmarks

(IEMF and TBID) to highlight the benefits of RL-based route planning against heuristic (IEMF) and tree-based (TBID) routing techniques.

5.1. Average End-to-End Delay (ETE)

End-to-End delay (ETE) is measured since a UAV starts a remote sensing mission until it completes data collection and returns to the base station. This metric is used to study the performance of route planning algorithms to forward UAVs through short and loop-free paths. This approach offers a number of benefits—mainly enhanced data freshness.

As Figure 4 shows, Q-learning outperforms IEMF if the number of SPs is increased. This stems from the exploration–exploitation capacity of Q-learning that allows environment interaction aiming to learn the best-fitted actions and find a greater number of SPs (maximize the reward). Hence, Q-learning reduces ETE as compared to IEMF if SP density is increased and a greater number of SPs is visited during exploration flights. However, IEMF works according to a heuristic and greedy approach, through which SPs are visited one-by-one based on their distance to the current UAV's location. This increases ETE when the number of SPs is increased.

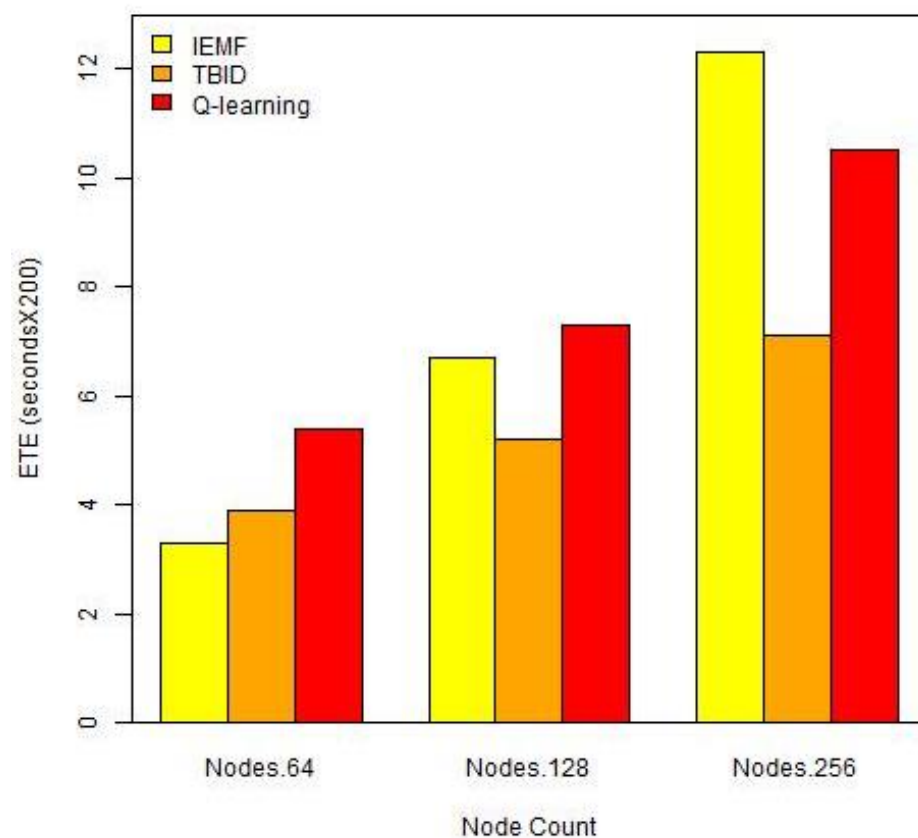


Figure 4. The average ETE (based on synthesized data of the model generated for this study).

IEMF outperforms Q-learning (and TBID) when SPs are rare. This is because the UAV wrongly finishes the remote sensing mission and returns to the base station, as it is unable to find any close SP to the current location. Indeed, IEMF forwards UAV into the base station with a minimized latency, as this suffers a lack of environment interaction to discover the farm fully. Figure 5 supports this and shows how IEMF addresses longer ETE for each visited SP if SPs are rare (e.g., 64).

TBID outperforms Q-learning and IEMF UAV route planning as this establishes a spanning tree in advance to move UAV through. By this, the UAV needs to compute no route while flying over the farm for remote sensing. This results in reduced ETE.

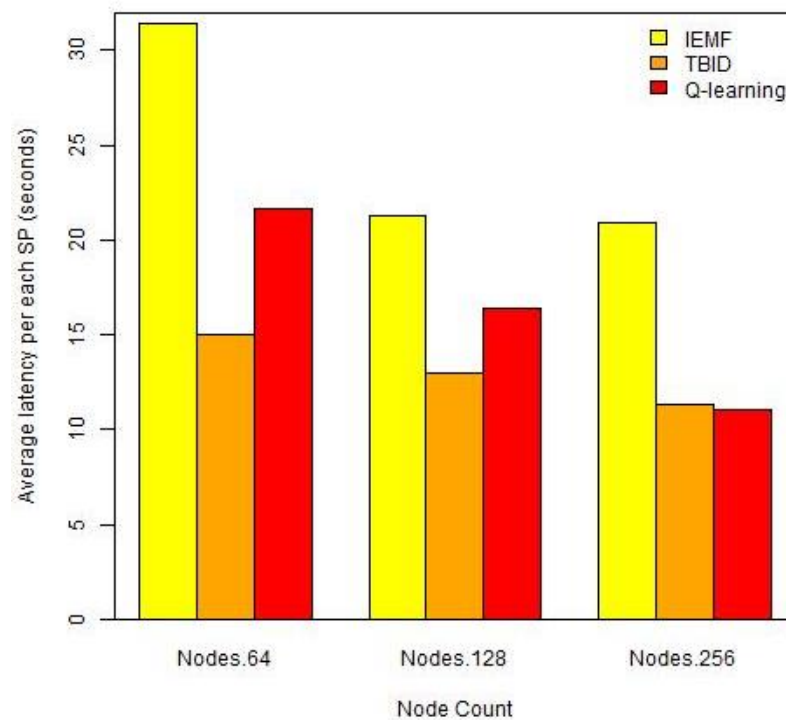


Figure 5. The average ETE per each visited SP (based on synthesized data of the model generated for this study).

5.2. Average Number of Captured Data

UAV remote sensing aims to collect, aggregate, and report the maximum number of sensory recordings (e.g., farming values). According to this, better performance is achieved if route-planning algorithms forward UAVs through best-fitted paths into target areas. By this, the UAVs are able to discover a greater number of SPs and collect maximized number of data samples. This results in increased data robustness and data collection accuracy.

According to Figure 6, Q-learning addresses a better data collection performance if SPs are crowded. This is because Q-learning has the capacity to interact with the environment and discover SPs. It takes location-aware actions that forward UAVs through loop-free paths to explore the environment and update the Q-table. This allows remote sensing to capture a greater number of sensory data samples. However, the number of captured data is reduced in Q-learning as compared to TBID if SPs are rare. This is because TBID forms a tree-based infrastructure to forward the UAV through and collect the data. This links SPs as a tree just before stating UAV remote sensing mission. By this, TBID-enabled UAV routing can visit a greater number of SPs, especially if they are rare and difficult to find through environment exploration.

IEMF underperforms Q-learning and TBID when SPs are rare. This is because IEMF utilizes a heuristic route planning paradigm to find and visit SPs. This is unable to find (mobile) SPs if they are (or move) far from the current location (e.g., Farm's center). By this, IEMF wrongly forwards the UAV to the base station to finish the remote sensing task. However, IEMF has a better performance when SPs are crowded.

5.3. Average Battery Consumption

Power resource consumption (e.g., battery) is a key in UAV remote sensing as UAVs are highly power constraint. This is measured to study the capacities of the UAV route planning algorithm to move UAVs through cost-effective paths with minimized power consumption to collect and report sensory recordings. Average battery consumption measures the amount of power consumed for the whole the procedure since the routing infrastructure is deployed until the remote sensing mission is finished.

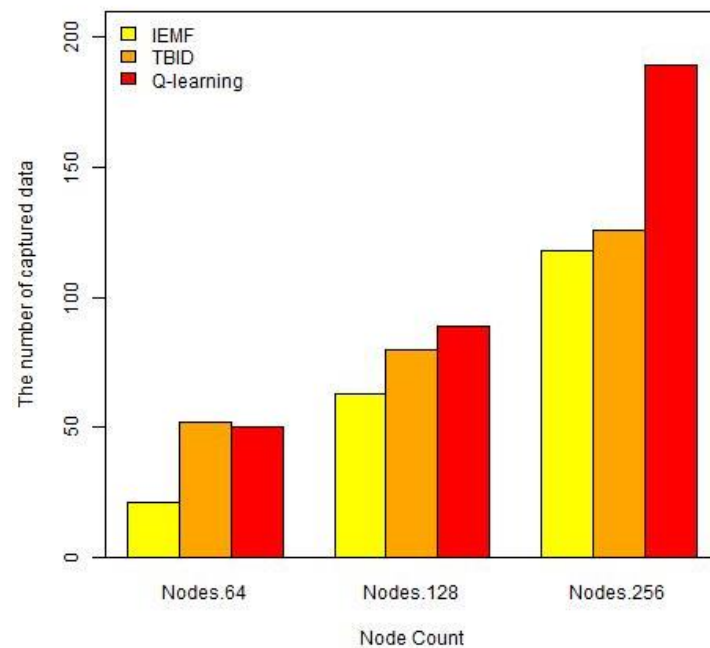


Figure 6. The average number of captured data (based on synthesized data of the model generated for this study).

As Figure 7 depicts, Q-learning outperforms both the benchmarks when SPs are crowded and dense. This is because Q-learning aims to find the best-fitted actions (flight location and direction) according to the current states to maximize the rewards. Thus, the proposed UAV routing algorithm takes the residual energy level and flight distance into account to select minimum cost routes, consequently reducing battery consumption. However, IEMF reduces consumed battery if SPs are rare. This is because IEMF has a limited ability to discover the farm and SP's locations. By this, the UAV leaves the remote sensing mission uncompleted and wrongly returns to the base station even though there are still unvisited SPs on the farm.

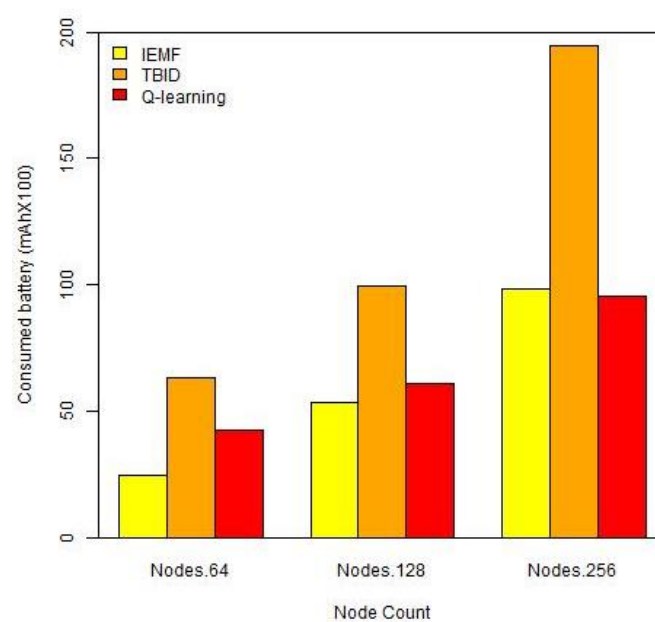


Figure 7. The average of consumed battery (based on synthesized data of the model generated for this study).

TBID underperforms Q-learning and IEMF due to the cost of tree-based infrastructure establishment. This forms a spanning tree to forward the UAV through for remote sensing. This results in increased power consumption, especially when the number of (mobile) SPs is increased.

5.4. Average Traversed Distance

Flight traversed distance reduction plays a key role in UAV route planning algorithm design. UAVs may traverse to discover the area, find SPs and/or setup the routing infrastructure. Increased traversed distance addresses enhanced battery consumption and increased ETE. Hence, a better route planning performance is achieved if the algorithm/technique is able to forward UAVs through minimized distances to complete a remote sensing mission.

Figure 8 shows that Q-learning reduces the UAV's average traversed distance as compared to IEMF if SPs are dense and crowded. This is because Q-learning takes the flight distance into account during environment exploration. Hence, the UAV learns the environment and finds the best movement to go through if the reward (minimized distance) is maximized. However, IEMF is able to only heuristically discover the SPs using a greedy algorithm. Hence, this increases the UAV movement, especially when the number of SPs is increased.

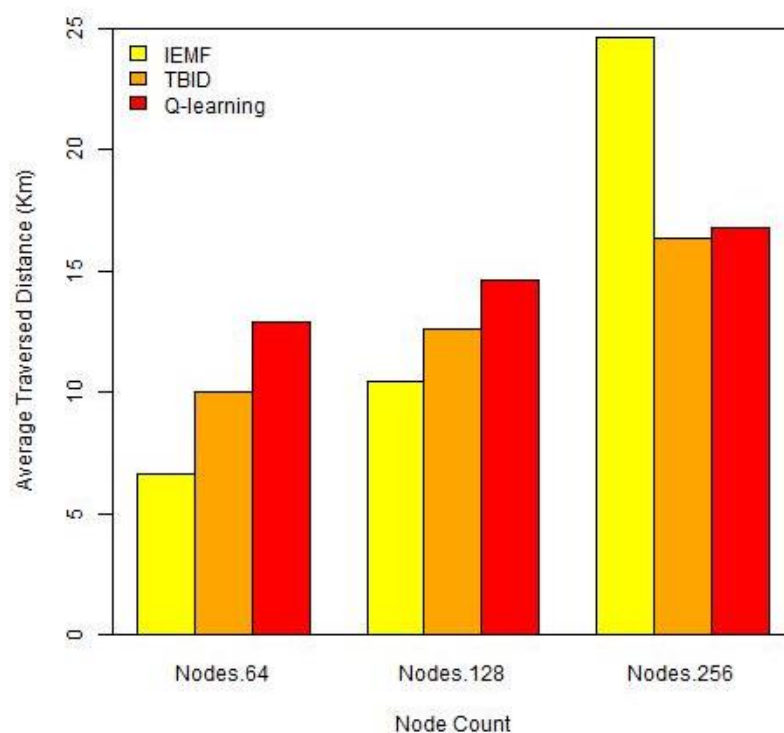


Figure 8. The average traversed distance (based on synthesized data of the model generated for this study).

TBID outperforms both Q-learning and IEMF as this forms a spanning tree to forward the UAV through. This results in the reduced traversed distance because the UAV movements are restricted into the tree infrastructure. However, this forwards UAV through longer and re-established paths to collected data from unvisited SPs when the number of (mobile) SPs is increased.

6. Conclusions and Future Work

This paper utilizes the reinforcement learning technique (Q-learning) in UAV path planning for smart farm remote sensing. This addresses an autonomous UAV itinerary

planning, which is established based on the Q-learning exploration-exploitation paradigm. Under this study, a single UAV is programmed to autonomously fly, discover the environment (autonomous zone-based labelling) and record the sensory points (SPs) with minimum cost/delay and maximum accuracy. Yet, this UAV collects, aggregates, and reports the sensory recordings to the base station for the end-user further processing and/or decision making.

The performance of Q-learning enabled UAV routing is compared with two benchmarks, including IEMF and TBID. This allows us to study the superiority and efficiency of reinforcement learning and environment exploration techniques compared to infrastructure-based (e.g., tree) and/or heuristic UAV route planning approaches. According to the results, Q-learning enabled UAV remote sensing offers a number of benefits, including reduced energy consumption, ETE, and traversed distance, and an enhanced number of captured sensory recordings, especially if SPs are crowded and dense. This shows that Q-learning allows us to find the best-fitted (e.g., minimized ETE and distance) and cost-effective (e.g., power consumption) paths to visit SPs and capture sensory recordings. However, infrastructure-based UAV route planning approaches reduce ETE and traversed distance, although they address increased power consumption required to establish the infrastructure. The energy consumption is significantly increased in infrastructure-based UAV route planning if SPs are highly mobile and the infrastructure needs to be repeatedly updated. Yet, heuristic UAV path planning increases power consumption, ETE, and traversed distance as it forwards UAV according to greedy algorithm. According to the simulation results, this addresses wrong flight directions and leaves remote sensing missions incomplete because the IEMF enabled UAV fails to find the next visiting SP when SPs are rare.

This study responds to the existing research gap on the use of UAV for smart farming. It models and addresses optimizing itinerary planning, path finding, and battery life management. The findings are novel, as the study uses reinforcement learning technique to autonomously partition a farm and dynamically forward UAVs through minimum distance and power composition routes for non-overlapped data collection. In doing so, the study feeds into future research work that aims to enhance the use and integration of UAV use in smart farming practices.

Further investigation on Q-learning UAV path planning optimization is still required. For this, the award function may consider a set of other features such as data-type to address data-centric UAV remote sensing. By this, the UAV can be programmed to fly over a farm and capture particular sensory recordings according to the user interest and/or remote sensing requirement. Moreover, tuned reinforcement learning algorithms, such as Deep Q Network, Proximal Policy Optimization and meta-reinforcement learning, can enhance path planning performance.

Future research studies need to consider the effects of the relation of the field's dimensions on the suggested algorithm. While this study uses a simplified model, future studies could benefit from utilizing this study's findings to explore a more realistic situation, such as irregularities of the field shape, field size, and target areas of the field.

Author Contributions: Conceptualization, S.P.A. and A.C.; methodology, S.P.A.; software, S.P.A.; validation, S.P.A. and A.C.; formal analysis, S.P.A.; investigation, S.P.A. and A.C.; resources, S.P.A. and A.C.; visualization, S.P.A.; project administration, S.P.A.; funding acquisition, S.P.A. All authors have read and agreed to the published version of the manuscript.

Funding: The authors would like to extend sincere thanks to University of Nottingham Ningbo China for supporting this research project under UNNC FOSE Grant: I01200900006 and I01200300007.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

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

References

- Joseph, R.B.; Lakshmi, M.; Suresh, S.; Sunder, R. Innovative analysis of precision farming techniques with artificial intelligence. In Proceedings of the 2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), Bangalore, India, 5–7 March 2020; pp. 353–358.
- Ardakani, S.P. MINDS: Mobile Agent Itinerary Planning Using Named Data Networking in Wireless Sensor Networks. *J. Sens. Actuator Netw.* **2021**, *10*, 28. [\[CrossRef\]](#)
- Ardakani, S.P.; Padget, J.; Vos, M.D. CBA: A cluster-based client/server data aggregation routing protocol. *Ad Hoc Netw.* **2016**, *50*, 68–87. [\[CrossRef\]](#)
- Chen, X.; Tang, J.; Lao, S. Review of Unmanned Aerial Vehicle Swarm Communication Architectures and Routing Protocols. *Appl. Sci.* **2020**, *10*, 3661. [\[CrossRef\]](#)
- Chen, S.; Laefer, D.F.; Mangina, E. State of Technology Review of Civilian UAVs. *Recent Patents Eng.* **2016**, *10*, 160–174. [\[CrossRef\]](#)
- Ghazzai, H.; Ghorbel, M.B.; Kadri, A.; Hossain, M.J.; Menouar, H. Energy-Efficient Management of Unmanned Aerial Vehicles for Underlay Cognitive Radio Systems. *IEEE Trans. Green Commun. Netw.* **2017**, *1*, 434–443. [\[CrossRef\]](#)
- Thibbotuwawa, A.; Bocewicz, G.; Nielsen, P.; Banaszak, Z. Unmanned Aerial Vehicle Routing Problems: A Literature Review. *Appl. Sci.* **2020**, *10*, 4504. [\[CrossRef\]](#)
- Khan, M.F.; Yau, K.L.A.; Noor, R.M.; Imran, M.A. Routing Schemes in FANETs: A Survey. *Sensors* **2020**, *20*, 38. [\[CrossRef\]](#)
- Ardakani, S.P. Wireless Sensor Network Routing Protocols for Data Aggregation. Ph.D. Thesis, Computer Science Department, University of Bath, Bath, UK, 2014.
- Nazari, M.; Oroojlooy, A.; Snyder, L.V.; Takac, M. Reinforcement Learning for Solving the Vehicle Routing Problem. In Proceedings of the 32nd Conference on Neural Information Processing Systems (NeurIPS 2018), Montreal, QC, Canada, 3–8 December 2018.
- Sutton, R.S.; Barto, A.G. *Reinforcement Learning: An Introduction*; The MIT Press: Cambridge, MA, USA, 2014.
- van Hasselt, H.; Guez, A.; Silver, D. Deep reinforcement learning with double Q-Learning. In Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, Phoenix, AZ, USA, 12–17 February 2016.
- Chen, M.; Leung, V.; Mao, S.; Kwon, T.; Li, M. Energy-Efficient Itinerary Planning for Mobile Agents in Wireless Sensor Networks. In Proceedings of the 2009 IEEE International Conference on Communications, Dresden, Germany, 14–18 June 2009; [\[CrossRef\]](#)
- rao, R. G.; China Babu, G. Implementation of Tree Based Itinerary Design (TBID) Algorithm Over Wireless Sensor networks by Using Mobile Agent. *IOSR J. Comput. Eng.* **2013**, *9*, 108–112; [\[CrossRef\]](#)
- Bacco, M.; Berton, A.; Ferro, E.; Gennaro, C.; Gotta, A.; Matteoli, S.; Paonessa, F.; Ruggeri, M.; Virone, G.; Zanella, A. Smart farming: Opportunities, challenges and technology enablers. In Proceedings of the 2018 IoT Vertical and Topical Summit on Agriculture—Tuscany (IOT Tuscany), Tuscany, Italy, 8–9 May 2018; [\[CrossRef\]](#)
- Lottes, P.; Khanna, R.; Pfeifer, J.; Siegwart, R.; Stachniss, C. UAV-based crop and weed classification for smart farming. In Proceedings of the 2017 IEEE International Conference on Robotics and Automation (ICRA), Singapore, 29 May–3 June 2017; [\[CrossRef\]](#)
- Tripicchio, P.; Satler, M.; Dabisias, G.; Ruffaldi, E.; Avizzano, C.A. Towards Smart Farming and Sustainable Agriculture with Drones. In Proceedings of the 2015 International Conference on Intelligent Environments, Prague, Czech Republic, 15–17 July 2015. [\[CrossRef\]](#)
- Nintanavongsa, P.; Pitimon, I. Impact of sensor mobility on UAV-based smart farm communications. In Proceedings of the 2017 International Electrical Engineering Congress (iEECON), Pattaya, Thailand, 8–10 March 2017; [\[CrossRef\]](#)
- Boursianis, A.D.; Papadopoulou, M.S.; Diamantoulakis, P.; Liopa-Tsakalidi, A.; Barouchas, P.; Salahas, G.; Karagiannidis, G.; Wan, S.; Goudos, S.K. Internet of Things (IoT) and Agricultural Unmanned Aerial Vehicles (UAVs) in smart farming: A comprehensive review. *Internet Things* **2020**, 100187. [\[CrossRef\]](#)
- Mukherjee, A.; Misra, S.; Sukrutha, A.; Raghuwanshi, N.S. Distributed aerial processing for IoT-based edge UAV swarms in smart farming. *Comput. Netw.* **2020**, *167*, 107038. [\[CrossRef\]](#)
- Islam, N.; Rashid, M.M.; Pasandideh, F.; Ray, B.; Moore, S.; Kadel, R. A Review of Applications and Communication Technologies for Internet of Things (IoT) and Unmanned Aerial Vehicle (UAV) Based Sustainable Smart Farming. *Sustainability* **2021**, *13*, 1821. [\[CrossRef\]](#)
- Bacco, M.; Barsocchi, P.; Ferro, E.; Gotta, A.; Ruggeri, M. The Digitisation of Agriculture: A Survey of Research Activities on Smart Farming. *Array* **2019**, 3–4, 100009. [\[CrossRef\]](#)
- Unal, Z. Smart Farming Becomes Even Smarter With Deep Learning—A Bibliographical Analysis. *IEEE Access* **2020**, *8*, 105587–105609. [\[CrossRef\]](#)
- Maddikunta, P.K.R.; Hakak, S.; Alazab, M.; Bhattacharya, S.; Gadekallu, T.R.; Khan, W.Z.; Pham, Q.V. Unmanned Aerial Vehicles in Smart Agriculture: Applications, Requirements, and Challenges. *IEEE Sens. J.* **2021**, *1*. [\[CrossRef\]](#)
- Malik, A.W.; Rahman, A.U.; Qayyum, T.; Ravana, S.D. Leveraging Fog Computing for Sustainable Smart Farming Using Distributed Simulation. *IEEE Internet Things J.* **2020**, *7*, 3300–3309. [\[CrossRef\]](#)
- Navarro, E.; Costa, N.; Pereira, A. A Systematic Review of IoT Solutions for Smart Farming. *Sensors* **2020**, *20*, 4231. [\[CrossRef\]](#) [\[PubMed\]](#)
- Debauche, O.; Trani, J.P.; Mahmoudi, S.; Manneback, P.; Bindelle, J.; Mahmoudi, S.A.; Guttadauria, A.; Lebeau, F. Data management and internet of things: A methodological review in smart farming. *Internet Things* **2021**, *14*, 100378. [\[CrossRef\]](#)
- Raja, L.; Vyas, S. The Study of Technological Development in the Field of Smart Farming. In *Advances in Environmental Engineering and Green Technologies*; IGI Global: Hershey, PA, USA, 2019; pp. 1–24; [\[CrossRef\]](#)

29. Triantafyllou, A.; Tsouros, D.C.; Sarigiannidis, P.; Bibi, S. An Architecture model for Smart Farming. In Proceedings of the 2019 15th International Conference on Distributed Computing in Sensor Systems (DCOSS), Santorini, Greece, 29–31 May 2019; [CrossRef]
30. Kim, J.; Kim, S.; Ju, C.; Son, H.I. Unmanned Aerial Vehicles in Agriculture: A Review of Perspective of Platform, Control, and Applications. *IEEE Access* **2019**, *7*, 105100–105115. [CrossRef]
31. Walter, A.; Finger, R.; Huber, R.; Buchmann, N. Opinion: Smart farming is key to developing sustainable agriculture. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 6148–6150. [CrossRef] [PubMed]
32. Theile, M.; Bayerlein, H.; Nai, R.; Gesbert, D.; Caccamo, M. UAV Coverage Path Planning under Varying Power Constraints using Deep Reinforcement Learning. In Proceedings of the 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Las Vegas, NV, USA, 24 October 2020–24 January 2021; [CrossRef]
33. Faryadi, S.; Velni, J.M. A reinforcement learning-based approach for modeling and coverage of an unknown field using a team of autonomous ground vehicles. *Int. J. Intell. Syst.* **2020**, *36*, 1069–1084. [CrossRef]
34. Li, B.; Gan, Z.; Chen, D.; Aleksandrovich, D.S. UAV Maneuvering Target Tracking in Uncertain Environments Based on Deep Reinforcement Learning and Meta-Learning. *Remote Sens.* **2020**, *12*, 3789. [CrossRef]
35. Conti, S.; Faraci, G.; Nicolosi, R.; Rizzo, S.A.; Schembra, G. Battery Management in a Green Fog-Computing Node: A Reinforcement-Learning Approach. *IEEE Access* **2017**, *5*, 21126–21138. [CrossRef]
36. Balasubramanian, M.; Navaneethan, C. Applications of Internet of Things for smart farming—A survey. *Mater. Today Proc.* **2021**. [CrossRef]
37. Tsouros, D.C.; Bibi, S.; Sarigiannidis, P.G. A Review on UAV—Based Applications for Precision Agriculture. *Information* **2019**, *10*, 349. [CrossRef]
38. OMNET++. OMNET++ Simulator. 2019. Available online: <http://www.omnetpp.org/> (accessed on 10 March 2019).
39. INET. INET Framework. 2019. Available online: <https://inet.omnetpp.org/Introduction.html> (accessed on 10 March 2019).
40. Qi, H.; Wang, F. Optimal Itinerary Analysis for Mobile Agents in Ad Hoc Wireless Sensor Networks. In Proceedings of the 13th International Conference on Wireless Communication, Calgary, Alberta, Canada, 9–11 July 2001; pp. 147–153.
41. Ardakani, S.P.; Padget, J.; Vos, M.D. A Mobile Agent Routing Protocol for Data Aggregation in Wireless Sensor Networks. *Int. J. Wirel. Inf. Netw.* **2017**, *24*, 27–41. [CrossRef]
42. Cai, W.; Chen, M.; Hara, T.; Shu, L.; Kwon, T. A Genetic Algorithm Approach to Multi-Agent Itinerary Planning in Wireless Sensor Networks. *Mob. Netw. Appl.* **2010**, *16*, 782–793. [CrossRef]
43. Wang, J.; Zhang, Y.; Cheng, Z.; Zhu, X. EMIP: Energy-efficient itinerary planning for multiple mobile agents in wireless sensor network. *Telecommun. Syst.* **2015**, *62*, 93–100. [CrossRef]
44. Yousefi, S.; Derakhshan, F.; Karimipour, H.; Aghdasi, H.S. An efficient route planning model for mobile agents on the internet of things using Markov decision process. *Ad Hoc Netw.* **2020**, *98*, 102053. [CrossRef]
45. DJI. Matrice 200. 2021. Available online: <https://www.dji.com/cn/matrice-200-series> (accessed on 20 March 2019).
46. AUSVET. Sample Size to Estimate a Single Mean with Specified Precision. 2014. Available online: <https://epitools.ausvet.com.au/onemean?page=1Mean&Stdev=45&Conf=0.95&Error=20> (accessed on 10 February 2020).
47. Fotouhi, A.; Ding, M.; Hassan, M. Deep Q-Learning for Two-Hop Communications of Drone Base Stations. *Sensors* **2021**, *21*, 1960. [CrossRef] [PubMed]