

Review

Variations in Multi-Agent Actor–Critic Frameworks for Joint Optimizations in UAV Swarm Networks: Recent Evolution, Challenges, and Directions

Muhammad Morshed Alam ¹, Sayma Akter Trina ¹, Tamim Hossain ¹, Shafin Mahmood ¹, Md. Sanim Ahmed ¹ and Muhammad Yeasir Arafat ^{2,3,*}

¹ Department of Electrical and Electronic Engineering, American International University-Bangladesh, Dhaka 1229, Bangladesh; dr.malam@aiub.edu (M.M.A.); 22-47535-2@student.aiub.edu (S.A.T.); tamim@aiub.edu (T.H.); 22-46665-1@student.aiub.edu (S.M.); 22-46670-1@student.aiub.edu (M.S.A.)

² IT Research Institute, Chosun University, Gwangju 61452, Republic of Korea

³ School of Computer Science and IT, University College Cork, T12 YN60 Cork, Ireland

* Correspondence: myarafat@chosun.ac.kr or marafat@ucc.ie

Abstract: Autonomous unmanned aerial vehicle (UAV) swarm networks (UAVSNs) can efficiently perform surveillance, connectivity, computing, and energy transfer services for ground users (GUs). These missions require trajectory planning, UAV-GUs association, task offloading, next-hop selection, and resource allocation, including transmit power, bandwidth, timeslots, caching, and computing resources, to enhance network performance. Owing to the highly dynamic topology, limited resources, stringent quality of service requirements, and lack of global knowledge, optimizing network performance in UAVSNs is very intricate. To address this, an adaptive joint optimization framework is required to handle both discrete and continuous decision variables, ensuring optimal performance under various dynamic constraints. A multi-agent deep reinforcement learning-based adaptive actor–critic framework offers an effective solution by leveraging its ability to extract hidden features through agent interactions, generate hybrid actions under uncertainty, and adaptively learn with scalable generalization in dynamic conditions. This paper explores the recent evolutions of actor–critic frameworks to deal with joint optimization problems in UAVSNs by proposing a novel taxonomy based on the modifications in the internal actor–critic neural network structure. Additionally, key open research challenges are identified, and potential solutions are suggested as directions for future research in UAVSNs.

Keywords: actor–critic frameworks; joint optimizations; multi-agent deep reinforcement learning; UAV swarm networks; trajectory control; task offloading; resource allocation



Academic Editor: Emmanouel T. Michailidis

Received: 26 December 2024

Revised: 10 February 2025

Accepted: 15 February 2025

Published: 19 February 2025

Citation: Alam, M.M.; Trina, S.A.; Hossain, T.; Mahmood, S.; Ahmed, M.S.; Arafat, M.Y. Variations in Multi-Agent Actor–Critic Frameworks for Joint Optimizations in UAV Swarm Networks: Recent Evolution, Challenges, and Directions. *Drones* **2025**, *9*, 153. <https://doi.org/10.3390/drones9020153>

Copyright: © 2025 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

Owing to the advantages of high maneuverability, autonomy, and flexibility, unmanned aerial vehicle (UAV) swarm networks (UAVSNs) can effectively execute various emergent missions. It includes simultaneous energy transfer and data collection from ground Internet of Things (IoT) devices/ground users (GUs), providing mobile edge computing services to resource-constrained IoT/GUs [1], aerial surveillance [2,3], and many more. These autonomous mission objectives are minimizing the average age of information (AoI) [4–6] or average delay in data collection and task execution [7], maximizing area coverage [8,9], maximizing fairness aware throughput [10], and minimizing energy consumption for both UAV [11] and IoT/GUs [12]. Thus, achieving the above objectives in

UAVSNs requires jointly optimizing several continuous and discrete decision variables. It may include controlling the continuous trajectory of UAVs [13], binary UAV-GUs association, determining binary or partial task-offloading, next-hop selection [14], and continuous resource allocation (i.e., transmit power, bandwidth/timeslot, caching, and computing).

In these optimization processes, several mixed integer nonlinear dynamic constraints are involved related to UAV's mobility, resource limitations (i.e., bandwidth, transmit power, onboard energy, and maximum queue size), UAV-GUs association, signal-to-interference-plus-noise ratio (SINR) threshold, and minimum delay or AoI threshold. Furthermore, owing to the mobility of UAVs and GUs, the network topology and channel state becomes very dynamic. Thus, the network state and decision space become very large. Additionally, the environment becomes both cooperative and competitive in terms of collaborative trajectory planning and resource sharing in multiple layers along with multiple decision-makers (i.e., UAVs, IoT devices/GUs, high altitude platforms, and satellites), as illustrated in Figure 1.

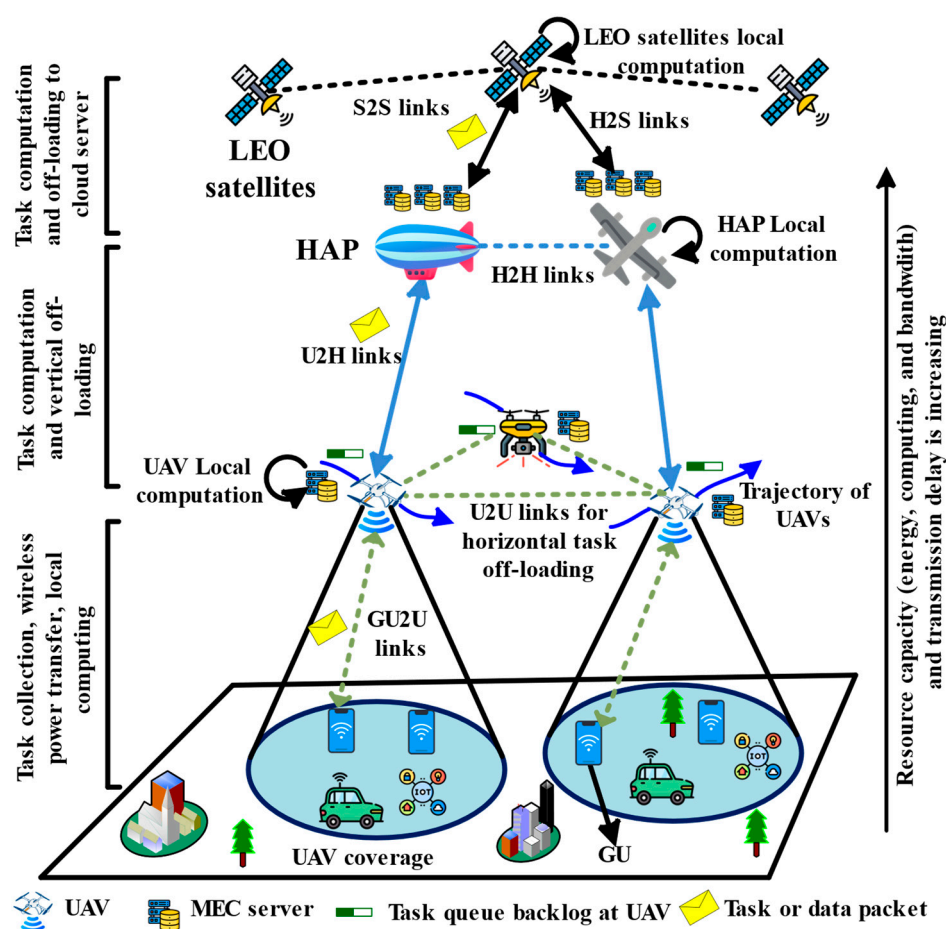


Figure 1. An illustration of a UAVSN-assisted mission.

For UAVSNs communication links are classified as several bidirectional links such as GUs to UAVs (GU2U) or UAV to GUs (U2GU), UAV to UAV (U2U), UAV to high-altitude platform (HAP) (U2H), HAP to HAP (H2H), HAP to satellites (H2S), and satellites-to-satellites (S2S) links, as shown in Figure 1. Particularly, GU2U or U2GU links encounter obstacles, but the probability of obtaining line-of-sight paths is high owing to the mobility adjustment of UAVs. In contrast, U2U links are highly dynamic due to the mobility of UAVs. Both GU2U and U2U are very resource-constrained in terms of transmit power, computing, and bandwidth resources [15,16]. On the other hand, U2H and H2S links have higher transmission delays, as these links have very long-distance free-space propagation

paths [17]. Additionally, owing to the mobility of LEO satellites, H2S links are dynamic in nature but predictable [18]. U2H, H2S, and S2S links are backhaul links for UAVSNs to route their sensed data to the cloud server and ensure global access [19]. In [13,20], UAVs are classified into fixed-wing, rotary-wing, and tethered UAVs. Rotary-wing UAVs can support vertical take-off and landing, whereas fixed-wing UAVs can move very fast with more payload capacity and endurance [21,22]. The type of UAVs should be selected based on the mission and application requirements. For instance, rotary-wing UAVs are highly suitable for tasks requiring hovering, such as data collection at specific points. In contrast, fixed-wing UAVs are ideal for surveillance and fast mission execution due to their higher speed and endurance. Additionally, multiple UAVs can be deployed as a swarm in either homogeneous or heterogeneous systems for efficient task allocation and execution [23].

Thus, UAVSN behavior can be treated as multi-agent systems [24,25]. The action space with neighboring agents becomes highly coupled; more specifically, actions taken by the neighboring agents have a higher impact on the network utility of a particular agent. As a result, existing fully centralized model-based iterative algorithms are not suitable for this unknown and highly dynamic environment. Here, collecting the centralized information on the dynamic topology is cumbersome, and this outdated centralized information-based network optimization can trigger performance degradation. Furthermore, the computational complexity and control overhead increase with the increasing number of UAVs and GUs, which limits scalability.

To combat the above-mentioned challenges, model-free data-driven multi-agent deep reinforcement learning (MA-DRL) can be efficiently utilized [26]. By leveraging the partially observable Markov decision process (POMDP), each agent can iteratively interact with the dynamic UAVSN to optimize the long-term cumulative reward based on their current observation. MA-DRL approaches can deal with multiple dynamic constraints by imposing penalty terms in the reward for choosing unfeasible actions by the agent without requiring convexity. Unlike the single-agent DRL algorithm, MA-DRL can avoid the environmental nonstationary problem by sharing observation–action via centralized training and distributed execution. Existing value-based MA-DRL approaches such as double deep Q-learning (DDQN) [27,28] and Q-learning-based mixing network (QMIX) [29] are suitable only for small-scale discrete action spaces. Hence, these algorithms are not suitable for large hybrid action spaces and extracting features from the dynamic UAVSNs. In contrast, policy gradient-based actor–critic learning can efficiently tackle large observation spaces and hybrid action spaces [30]. Here, the actor network is responsible for generating continuous and discrete actions. Consequently, the critic network is responsible for evaluating the action by estimating the value function, which is further utilized to train the actor network and critic network [31]. The discrete action spaces such as binary task off-loading ratio, discrete frequency band selection, discrete next hop selection to relay data packet, and binary UAV-GUs association can be mapped into continuous probability distribution using softmax activation function at the output layer of the actor-network. Then, during exploitation, the possible action can be sampled according to the maximum probability [32].

According to the existing literature, policy-based multi-agent actor–critic algorithms can be classified as off policy-based deep deterministic policy gradient (DDPG) [8], on policy-based trust region policy optimizations (TRPO) [33], and on policy-based proximal policy optimization (PPO) [31]. Among them, DDPG can handle only continuous action spaces such as trajectory control of UAVs or resource allocation in continuous domains. TRPO and PPO can tackle both continuous and discrete action spaces. Although these policy-based MA-DRL approaches have superior performance compared with the DDQN or QMIX, they encounter instability during policy updates due to the variability of the step-size parameter in policy optimization [4]. Small step sizes trigger slow learning, while

large ones can degrade performance. To address this, TRPO defines a trusted region for updating policy using a complex second-order method. PPO simplifies it by employing a first-order method, which bounds policy updates within a range instead of TRPO's hard constraint. PPO has two key variants: PPO-penalty and PPO-clip. PPO-penalty replaces TRPO's hard constraint with a penalty in the objective function. In contrast, PPO-clip utilizes a clipping mechanism to limit policy changes without a hard constraint [34].

1.1. Related Work and Motivations

In this section, related survey articles are reviewed along with the motivation for writing this survey article to highlight our key contributions. In [35], existing distributed machine learning-based approaches for UAVSNs are reviewed and classified into federated learning and multi-agent reinforcement learning for trajectory optimization, sensing, edge computing, and resource allocation. In [13], the authors reviewed the existing trajectory and topology control algorithms by classifying them based on their underlying mathematical model. In [36], the existing edge computing algorithms for the metaverse environment are reviewed, focusing on optimization techniques. In [37], the authors reviewed the existing edge computing algorithms for UAVSNs to support various artificial intelligence-based applications in computer vision, navigation, power control, security and privacy, and communication. In [38], the authors investigated the application of transformer-based encoder–decoder networks in UAV-assisted wireless networks to perform compression, schematic communication, and channel estimation. In [39], the existing task offloading algorithms in UAVSNs are reviewed and classified based on value-based and policy-based methods. Additionally, they have discussed the tasks in UAVSNs by classifying them as dependent tasks and independent tasks. The independent task can be processed in parallel, whereas dependent tasks must be processed sequentially by assuming a directed acyclic graph of task arrival. Task size, CPU cycle requirements, and task execution latency requirements depend on the application scenario. In [40], the existing trajectory-aware task offloading algorithms are reviewed and classified into general iterative optimization and reinforcement learning-based optimization methods. In [41], the existing data aggregation techniques for UAVSN-assisted IoT devices are reviewed. However, all of this existing survey did not focus on joint optimization problems based on modifications of actor–critic MA-DRL framework neural network structure in UAVSNs, which has drawn significant research attention.

Owing to the dynamic topology of UAVSNs, actor–critic-based MA-DRL algorithms are highly suitable for sequential decision-making. Nevertheless, an actor–critic network solely depending on a fully connected network (FCN) or multi-layer perceptron (MLP) cannot deal with time series sequential data such as UAV trajectories and channel states. Additionally, FCN requires a higher number of learning parameters, and this FCN-based learning is less scalable and not adaptive to environmental changes [42]. Fortunately, the recurrent neural network (RNN) such as long short-term memory (LSTM) and gated recurrent unit (GRU) can remember a finite amount of the most recent historical sequential data and utilize it to predict more accurate state in the next timeslot by mining the temporal relationship with the state information in the current timeslot [43]. Nevertheless, RNN-based LSTM or GRU-based encoder–decoder networks may encounter gradient vanishing problems during backpropagation and cannot efficiently deal with lengthy time series data. Moreover, RNN cannot adopt parallelization in feature extraction.

The transformer-based encoder–decoder networks can overcome these limitations via multi-head attention mechanisms [44]. However, how to employ these networks in actor–critic frameworks to address joint optimization problems in dynamically resource-constrained UAVSNs remains a research question. Although plenty of research has been

conducted to address joint optimization problems in UAVSNs by leveraging various actor–critic-based MA-DRL frameworks, no in-depth survey has focused on the variations in actor–critic frameworks, their key limitations, research challenges, and potential solutions as future research directions. It motivates us to write this survey. According to the above discussions, a comparison of key contributions in this article relative to existing survey articles is summarized in Table 1.

Table 1. Comparison of key contributions in existing survey articles related to this survey.

Ref.	Year	Key Contributions
[35]	2023	Existing distributed machine learning algorithms are reviewed and classified as federated learning and multi-agent reinforcement learning, to optimize trajectory and resource allocation in UAVSNs.
[13]	2022	Existing trajectory and topology control algorithms in UAVSNs are reviewed and classified based on their underlying mathematical model.
[36]	2024	Existing edge computing algorithms for metaverse environments are reviewed based on optimization techniques.
[37]	2022	Existing edge computing algorithms for UAVSNs are reviewed to support various artificial intelligence-based applications in computer vision, navigation, power control, security, and communication.
[38]	2024	The application of transformer-based encoder–decoder networks in UAV-assisted wireless networks is investigated to perform compression, schematic communication, and channel estimation.
[39]	2024	Existing task offloading algorithms in UAVSNs are reviewed by classifying them based on value-based and policy-based methods.
[40]	2024	Existing trajectory-aware tasking offloading algorithms are reviewed by classifying them into general iterative optimization and reinforcement learning-based optimization methods.
[41]	2023	Existing data aggregation techniques are reviewed for UAVSN-assisted IoT devices.
This survey		• An overview of UAVSN environments is provided, focusing on mission-oriented optimization objectives, control variables, available links, and the potential to adopt actor–critic-based MA-DRL frameworks. • The evolution of existing actor–critic frameworks used to solve joint optimization problems in UAVSNs is extensively studied and classified based on modifications to the internal actor–critic neural networks. • Additionally, a comprehensive comparative study is presented, considering optimization objectives, control variables with constraints, POMDP formulation, training methods, modifications in policy gradient, loss function, state embedding, advantages, and limitations in addressing joint optimization problems in UAVSNs. • Key open research issues are discussed, along with potential solutions for UAVSNs.

1.2. Key Contributions

The major contributions of this paper are summarized as follows:

- The evolution of recently proposed state-of-the-art actor–critic frameworks is extensively reviewed based on their internal neural network architecture, optimization objectives, control variables with constraints, POMDP formulation, training methods, policy gradient, modifications in the loss function, key contributions, advantages, and limitations in addressing joint optimization problems in the UAVSNs environment.
- A novel taxonomy for recently proposed multi-agent actor–critic frameworks is provided based on their internal neural network structure to address joint optimization problems in UAVSNs while considering reasonable key constraints.

- It will facilitate researchers to choose and design appropriate actor–critic frameworks with internal neural network cell and activation function, policy gradient, loss, and reward with penalty terms for obtaining optimal solutions in complex UAVSNs environment with reasonable training in a high-fidelity digital twin environment. Additionally, it will assist researchers in understanding how to extract key features from large dynamic time series state parameters in UAVSNs, facilitating improved decision-making in multi-agent cooperative–competitive UAVSN environments.
- According to the comparative study, key open issues and research challenges are identified. Additionally, each possible research direction is addressed with potential solutions as future research directions for the UAVSN environment.

1.3. Outline of This Paper

The remainder of this paper is arranged as follows. In Section 2, the recent evolution of actor–critic frameworks in UAVSNs is briefly discussed based on the proposed taxonomy, the target optimization objective, constraints, POMDP formulation, modifications of the internal actor–critic network, policy gradient, loss function, and computational complexity. In Section 3, based on the proposed taxonomy, existing algorithms are qualitatively compared. In Section 4, related research challenges and potential solution directions are summarized. Finally, this paper is concluded in Section 5.

2. Recent Evolution in Actor–Critic Frameworks

In this section, based on the internal neural network architecture, the recent evolution of actor–critic frameworks to deal with joint optimization problems in UAVSNs are classified as follows: solely FCN or MLP-based, RNN-based, hybrid-structure-based, transformer-based encoder–decoder, and quantum neural network (QNN)-based, as shown in Figure 2. Based on this classification, each proposed framework is briefly reviewed in terms of its optimization objectives, constraints, POMDP formulation, training method, policy gradient, and computational complexity.

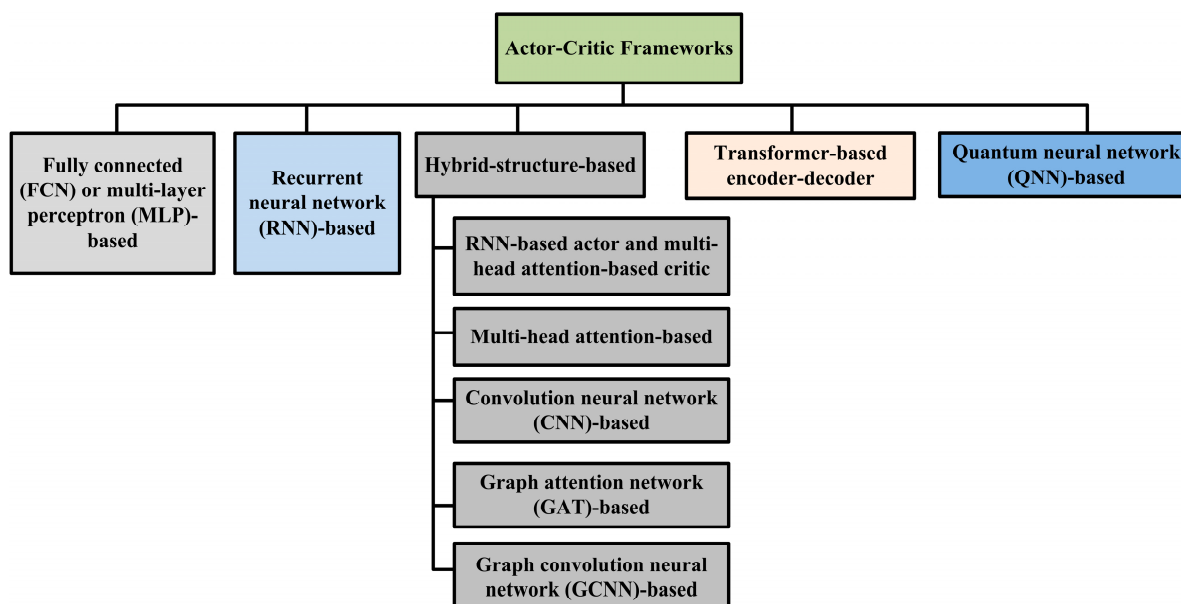


Figure 2. Taxonomy of actor–critic frameworks in UAVSNs.

2.1. Solely Fully Connected Network (FCN) or Multi-Layer Perceptron (MLP)-Based Frameworks

This type of actor–critic framework utilizes only FCN or MLP in both actor and critic networks. It struggles to capture key features from dynamic UAVSNs, and training

is not adaptive to varying numbers of UAVs or GUs. Thus, it has less scalability and higher computational complexity. In [45], minimizes the total system cost consisting of weighted energy consumption and delay-aware task priority-based utility function in GUs-UAVSNs assisted edge computing mission. They have considered binary task offloading decisions, continuous UAV mobility adjustment, transmit power, and computing resource allocation in the hybrid action space. To deal with hybrid action space, existing algorithms approximate either solely in discrete or continuous probability distributions, which may degrade optimization performance. Thus, considering the hybrid action space, they have designed a learnable conditional variational auto-encoder (VAE), and decoder network based on FCN in the presence of Gaussian latent distribution in the encoder. This VAE can approximate better latent space for hybrid action representation in the continuous domain to ensure policy gradient for training the FCN-based actor-critic framework utilizing twin-delayed MA-DDPG.

2.2. Recurrent Neural Network (RNN)-Based Frameworks

In RNN-based actor-critic frameworks, actor-critic networks are mainly constructed based on either LSTM or GRU cells [46]. In some cases, an LSTM or GRU-based state encoder-decoder is formed to predict the dynamic topology of UAVSNs. In [4], a mean field game (MFG) hybrid PPO-based actor-critic framework is proposed to minimize the average AoI by jointly optimizing UAV trajectories and data collection scheduling of GUs in a hybrid action space. To predict dynamic topology and stabilize the training process, this framework utilizes a learnable LSTM-based state encoder network. Then, LSTM-based state encoder output is fed into the two isolated actor networks to generate continuous and discrete actions related to the UAV's trajectory and transmission scheduling of GUs, respectively. In [24], LSTM is utilized in both actor-critic networks to control the trajectory and maintain the adaptive formation of UAVSNs in continuous action space (i.e., acceleration) in the presence of dense obstacles, with the help of an improved potential field-based reward function. The model is trained using twin-delayed MA-DDPG. This model performs well owing to the advantages of capturing long-term dependencies in sequential trajectory data by using LSTM.

In [47], the authors utilize an FCN-based DDPG actor-critic framework to maximize energy efficiency and throughput for UAV-aided data collection with the help of energy harvesting at UAV using a rectangular array-based terrestrial base station with power beamforming toward UAV. In [48], an MA-DRL-based actor-critic framework is employed to make cooperative packet routing decisions in a hierarchical topology-based UAVSN. The objective of this framework is to minimize packet routing latency along the end-to-end path by selecting the least congested shortest path. It incorporates a weighted reward function based on the queue backlog difference and distance progress toward the destination. Additionally, the framework trains an MLP-based binary classifier to determine whether to flood information within the local zone of UAVs based on their observations. Furthermore, a GRU-based communication channel is trained using the actor network's policy function and the critic network's Q-value function to significantly compress communication data.

2.3. Hybrid Structure-Based Frameworks

Hybrid structure-based actor-critic frameworks are classified into five categories according to the type of neural network as follows:

2.3.1. RNN-Based Actor and Multi-Head Attention-Based Critic

This type of actor-critic framework employs either LSTM or GRU in the actor network to capture the sequential behavior of dynamic states. Subsequently, utilizes a multi-head cross attention-based critic network to precisely estimate the value function. In [2], the

authors proposed a behavior-based mobility integrated LSTM-based actor and multi-head cross-attention-based critic framework to maximize a multi-metric link utility to deal with packet routing problems in dynamic UAVSNs. They have considered the cross-layer design consisting of a large hybrid action space, which includes collaborative trajectory control, frequency allocation, and next-hop selection. The model is trained using distributed MADDPG consisting of two-hop neighbor information-based POMDP. The actor network is constructed based on an LSTM-based three-layer state encoding layer and an FCN. LSTM-based state encoding layers extracted key features from dynamic topology by detecting temporal continuity utilizing current and historical information. Subsequently, a multi-head cross-attention-based critic network determines the precise value function to train the actor network by paying attention to the neighboring agent's observation–action space according to their degree of influence. It provides training stability and overcomes environmental nonstationary, training adaptivity, and algorithm convergence.

In [1], a swarming-behavior-integrated GRU-based actor and a multi-head attention-based critic network are proposed for an air-ground integrated network. Here, mobile edge computing-enabled UAVSN acts as agents to jointly minimize the task execution delay for mobile GUs and energy consumption for UAVs by controlling UAV's trajectory, UAV-GUs association, partial task off-loading ratio, computing, and bandwidth allocation for G2U links under several key constraints. This framework introduced behavior-based motion models and adaptive penalty terms to enhance decision-making. The GRU-based actor network is capable of coping with the dynamic observation of each agent. Similarly, the multi-head attention-based critic network utilizes a cross-attention mechanism to identify the influence of other neighboring agent's dynamic observation–action space for each individual UAV agent. Moreover, it provides training adaptivity to different numbers of UAVs and GUs, system scalability, and helps to avoid environmental nonstationary.

2.3.2. Multi-Head Attention-Based

This type of actor–critic framework employed multi-head attention mechanisms in both actor–critic networks to provide both self and cross-attention in multi-agent environments. In [10,12], a joint optimization on trajectory planning and communication design in air-ground communication systems is investigated by adopting an actor–critic-based heterogeneous collaborative decision-making algorithm, considering UAV base stations and GUs as intelligent agents. Thus, the action space for UAV base stations includes trajectory planning and bandwidth allocation for G2U links to maximize long-term fair throughput. Similarly, the action space for GUs includes binary UAV-GUs association to compete for limited UAV resources in each timeslot to maximize their own long-term throughput. To quickly identify the correlation with other agent's complex observation–action space, as well as enhance the model adaptability and the decision-making process, they introduced a learnable multi-head attention mechanism in both the actor and critic network of all agents.

In [49], UAVSNs are employed for a cooperative target search mission, with a multi-objective reward function designed to include four components: area coverage, exploration of unexplored regions, target discovery reward, reward for target discovery time, and avoidance of collisions with unknown threats and neighboring UAVs. This method utilizes a multi-head attention mechanism within the actor networks to focus on the most important features in observations from neighboring UAVs. Furthermore, to bridge the gap between simulation and real-world implementation, this approach leverages a digital twin of the physical environment to effectively train the agent UAVSNs in a high-fidelity simulation environment.

2.3.3. Convolution Neural Network (CNN)-Based

In joint optimization problems, typically, the observation and action space are very large and dynamic. Thus, convolution neural networks (CNN) can be utilized to extract the key features from the observation–action spaces by reducing the dimensionality. Then, CNN is followed by LSTM or GRU layers and, MLP can learn temporal dependencies more efficiently by utilizing historical knowledge of sequential data. Nevertheless, the computational complexity and vanishing gradient problem are key limitations of this network.

In [5], a federated heterogeneous actor–critic framework is utilized to minimize AoI for computing tasks in a multi-agent mobile GU–UAV–cloud environment by jointly considering computation offloading, UAV trajectory, and bandwidth allocation. Since tasks arrive at mobile GUs following a Poisson distribution, along with a dynamic task queue and topology, CNN- and MLP-based actor networks were considered. CNNs can extract the areas of GUs from dynamic observation that have higher computation demand and AoI; thus, accordingly, the trajectory of UAVs can be designed.

2.3.4. Graph Attention Network (GAT)-Based

In this type of framework, an actor–critic network is designed based on a graph attention network (GAT). GU and UAVSN topology can be represented as a dynamic graph attention network, and it can be efficiently trained using the MA-DRL method [50]. The learnable attention weights in GAT facilitate UAVs to pay more attention to the nearby UAVs according to their degree of influence in decision-making via communication range restricted partial observation [51]. In [52], the authors addressed the joint nondivisible task offloading and resource allocation (i.e., bandwidth and computing) in a hierarchical model considering mobile devices, edge, and cloud servers. This model leverages a learnable adaptive attention-weighted graph RNN-based state encoder utilizing the local knowledge followed by an actor–critic framework. Owing to the learnable state encoding using spatiotemporal graph RNN, the model can generalize different numbers of mobile devices and edge servers via fixed-size embedding, which leads to higher scalability and training adaptability to the unknown workload on the edge server.

2.3.5. Graph Convolution Neural Network (GCNN)-Based

This type of framework utilizes the graph convolution neural network (GCNN) as the state embedding layer to represent the dynamic state information such as channel state in GU2U or U2GU links, task attributes, GUs location, and UAVs location in a multi-relational graph-structured data [53]. Additionally, CNN-based pooling in GCNN helps to reduce the large-dimensional state information to the low-dimensional vector, which is suitable for feeding to the FCN-based actor network for action generation. Moreover, the graphical relationship integrates each agent state information with the neighboring agent state information, which enhances decision-making via multi-agent interaction and facilitates extracting key features inherent in graphical structural data. Thus, a notable advantage of this type of framework is to effectively capture and leverage the complex relationships and spatial dependencies among various devices in UAVSNs.

Based on this concept, in [53], a multi-relational graph convolutional state embedding-based actor–critic DDGP framework is applied to jointly control the trajectory of UAVs and the partial task offloading ratio of GUs to minimize task execution latency. The total battery capacity of UAVs, flight angle, velocity of UAVs, and partial task offloading ratio are considered the key constraints. Nevertheless, the algorithms neglected the communication and computing-related resource allocation, which is highly required to ensure optimal performance. Additionally, it requires higher computational complexity to train the two dif-

ferent agent's (GUs and UAVs) GCNN-based actor–critic frameworks in real-time because GUs have very limited resources.

2.4. Transformer-Based Encoder–Decoder Frameworks

This type of framework utilizes a transformer-based encoder–decoder structure in both actor and critic networks to learn from long observation–action spaces in UAVSNs. It can avoid vanishing gradient problems, remember long sequences, and train adaptivity to different sizes of networks via fixed-size embedding. However, it has a higher computational complexity. In [32], transformer-based encoder–decoder actor–critic frameworks are proposed to jointly minimize the long-term average weighted sum of task execution delay and energy consumption for multi-UAV and multi-edge server networks. Here, the action space consists of a binary task offloading policy and continuous resource allocation that includes transmit power, bandwidth, and computing resources. The model employed a DDPG-based policy gradient during training and adopted a transformer-based encoder–decoder structure to accelerate algorithm convergence due to the capabilities of handling large state–action decision space and avoidance of vanishing gradient problems. Furthermore, owing to the encoder–decoder-based key feature extractions, it is adaptive to different sizes of UAVs and task attributes.

In [8], an area coverage problem for multi-UAVs is investigated by utilizing transformer-based state encoder-based actor–critic frameworks. The model can make long-term trajectory decisions for UAVs to maximize fairness-aware area coverage toward points of interest. The transformer-based encoder network considers the observation of the neighbor UAV's state as key and value, each UAV's state as the query to generate a state embedding feature vector. This extracted feature vector from the dynamic environment is fed into an MLP-based actor–critic framework for training the model. Through feature extraction, the model is adaptive to different state features and different numbers of UAVs. Additionally, instead of random initialization of learnable weights in actor–critic networks, the designed model adopted a baseline-assisted pre-training mechanism, which can accelerate algorithm convergence.

2.5. Quantum Neural Network (QNN)-Based Frameworks

This type of framework employs QNNs in both the actor and critic networks to adopt powerful quantum computing for making quicker decisions while extracting key features via quantum state encoder from complex high-dimensional state–action spaces in UAVSNs [54]. By utilizing the principles of superposition, entanglement, and quantum interference, it has powerful computational benefits that are highly suitable for dealing with learning-based complex optimization problems.

The work in [55] leverages the unique capabilities of quantum computing in distributed actor and centralized critic framework to obtain the optimal trajectories of UAVs by utilizing millimeter wave (mmWave) communication. This framework overcomes the scalability challenges and mitigates quantum errors produced by the increasing number of qubits during the noisy intermediate-scale quantum process. Additionally, it introduces noise both in state and action spaces to make the UAVSN environment more realistic and the learning process more effective. In [56], the authors utilize QNNs in an actor–critic framework to execute task scheduling global access services given by geometry-aware coordinated large-scale cube satellites toward terrestrial ground stations. Owing to the advantages of modified QNN-based MA-DRL, it can reduce not only the scheduling action spaces but also the number of qubits. Moreover, due to application-oriented reward function design, the model ensures better quality of service (QoS) performance in global access in terms of maximizing transmission capacity and the residual energy of cube satellites.

QNNs also can ignore unworthy results and maintain focus on the best-suited outputs, which results in better optimization along with being resourceful in computing because of their flexibility to changes in the environment. It includes an enhanced level of data security due to quantum encryption. This is one of the major concerns for modern UAVs, particularly in military and sensitive commercial applications. Despite the several benefits of QNNs, there are a couple of challenges impeding the practical application of QNNs. One of the most critical limitations is the availability of hardware resources to support QNNs. The current quantum hardware is still in its infancy, with problems like noise, qubit decoherence, and a limited number of qubits affecting the reliability and scalability of QNNs. Quantum systems are also susceptible to high error rates that may compromise the accuracy of decisions made by QNNs. These issues demonstrate the need for further developments in quantum hardware and techniques for error correction. Another challenge is the very limited availability of software tools and libraries designed for QNN development. Whereas classical neural networks can rely on a mature ecosystem of libraries and frameworks, QNNs are still lacking similar support. This lack of public development tools causes more complexity in designing, training, and deploying QNN-based systems.

Despite these challenges, quantum-enhanced actor–critic frameworks are especially promising for UAVSs that require high-dimensional control and adaptive resource allocation under stringent QoS and resource-constrained dynamic environments. It aims to tackle traditional MA-DRL challenges like slow convergence and local optima by exploiting quantum acceleration in policy optimization in actor networks and value evaluation processes in critic networks.

3. Comparisons of Actor–Critic Frameworks

In this section, according to the given taxonomy, the existing actor–critic frameworks are qualitatively compared in terms of their optimization objectives, considered learning agents, state spaces, action spaces, policy and training methods, key contributions, advantages, and limitations as summarized in Tables 2 and 3. According to Table 2, most of the algorithms consider dynamically changing parameters in state space, and multiple discrete and continuous control variables are considered in the action space since it requires dealing with joint optimization. In most cases, some algorithms consider dual agents consisting of UAVs and GUs. All of these algorithms consider similar neural network architecture for both UAVs and GUs even though GUs have very less computational and energy resources. The hybrid-structure-based frameworks proposed in [1,2] utilize virtual forces for each UAV actor network, inspired by nature inspired Boids flocking theory to efficiently control the flying movement, velocity, and acceleration of UAVs. Owing to the advantage of adhering to physics, these physics-informed neural networks should provide more practical and smoother trajectory for UAVs. Most algorithms typically utilize the following activation functions in the output layer of the actor and critic networks: softmax, tanh, and sigmoid. Softmax is used for generating both discrete and continuous actions, as well as value functions, as it produces a probability distribution over all possible actions. The tanh and sigmoid activation functions are widely used for continuous action spaces, where the valid ranges are $[-1, 1]$ for tanh and $[0, 1]$ for sigmoid, respectively.

Nevertheless, none of these algorithms considered the significant modifications in the actor policy and critic loss function modification to make the algorithm more robust, which is highly required to train the model in a dynamic environment. During training, the model priority-aware data sampling and data augmentation may further improve the performance of the model, which requires further investigation. Additionally, most of the algorithms utilize state and action noise only based on Gaussian probability distribution; other probability distributions or any deterministic method can be utilized according to the type of action or control parameters.

Table 2. Summary of actor–critic frameworks to deal with joint optimization problems in UAVSNs.

Type	Ref.	Optimization Objective	Agent	State Space	Action Space
Solely FCN or MLP-based	[45]	To minimize total cost consisting of energy consumption and delay-aware task priority-based utility in UAVSNs assisted edge computing mission for GUs	GUs and UAVs	3D position of UAVs and task attributes	Binary task offloading, transmit power of GUs, UAV position, and computing resource allocation at UAVs
	[47]	To maximize energy efficiency and throughput for GU2U links with energy harvesting at UAVs via beamforming from terrestrial base station	UAVs	UAV position, energy level, and channel gain between UAV to base station and U2GU links and SINR	UAV movement adjustment and transmit power control from UAV
RNN-based	[4]	To minimize average AoI in UAVSNs assisted data collection mission from GUs	UAVs	2D position of UAVs and AoI of GUs	Scheduling data transmission for GU and UAV mobility
	[48]	To minimize packet routing latency in the end-to-end path while avoiding network congestion	UAVs	Location, ID number, and queue length of neighboring UAVs, and location of destination packet	Next hop selection from local neighbor list
	[24]	To plan trajectory for UAVSNs while maintaining formation in an unknown dense obstacle environment	Leader and follower UAVs	3D position of UAVs, pitch and yaw angle, and relative distances with nearest obstacles	Acceleration of UAVs
	[2]	To maximize a multi-metric link utility to route data packets in UAVSNs by utilizing two-hop neighbor information	UAVs	Virtual forces, binary frequency band paring matrix and link duration up to two hops, residual energy, and queue backlog size	UAV trajectory control, frequency band, and next-hop selection
Hybrid structure-based	[1]	To minimize energy consumption of UAVs and delay in air-ground integrated edge computing networks	UAVs	Virtual forces, SINR, GU-UAVs association matrix, task attributes, queue size, and remaining energy of UAVs	Trajectory of UAVs, UAV-GUs association, bandwidth, and computing resource allocation
	[10]	To maximize throughput for G2U links to provide communication service	GUs and UAVs	For GUs: GU and UAV position, and GU transmit power For UAVs: UAV position, and GU-UAVs paring matrix	UAVs: trajectory planning and bandwidth allocation, and GUs: binary UAV-GUs association
	[49]	To minimize target searching time and maximize the area coverage while avoiding collision avoidance	UAVs	Probability map of target, visiting time, neighboring UAVs and threats, and location of UAVs	Selecting flying direction of UAVs from a discrete set
	[5]	To minimize AoI for computing stochastic tasks in a multi-agent mobile GUs-UAV–cloud environment	GUs and UAVs	For GUs: its position and queue backlog size. For UAVs: its position, queue backlog size, and available bandwidth	Computation offloading, UAV trajectory, and bandwidth allocation
	[52]	To minimize long-term system costs consisting of energy and delay in a mobile device–edge server–cloud networks	GUs	Remaining computing resource, queue backlog size, channel gain, GUs position, and task attributes	Binary task offloading, computing, and bandwidth allocation
	[53]	To minimize the task execution latency for GUs in UAVs by jointly controlling UAV movement and partial task offloading ratio	GUs and UAVs	For GUs: position of GUs, task attributes, and channel state For UAVs: UAV position and remaining energy	GU selects the task offloading ratio, and UAV selects the flight angle and flight speed
	[32]	To minimize long-term average weighted task execution delay and energy consumption for multi-UAV and multi-edge server networks	UAVs	Task attributes and distance between UAVs	Binary task offloading, transmit power, bandwidth, and computing resources
QNN-based	[55]	To maximize data rate for U2GU links under QoS constraints	UAVs	GU-UAV paring matrix, and QoS requirement	UAVs movement

Table 3. Summary of actor–critic frameworks to deal with joint optimization problems in UAVSNs (continued).

Type	Ref.	Policy Method	Training Method	Key Contributions	Advantages	Limitations
Solely FCN or MLP-based	[45]	MA-DDPG	CTDE	Designing a conditional VAE-decoder-based continuous hybrid latent space	VAE provides better approximation for hybrid action space in continuous domain and ensures policy gradient for twin-delayed MA-DDPG	FCN has higher computational complexity, less scalability, and less adaptive to dynamic topology
	[47]	DDPG	CTDE	Designing UAV-assisted data collection mission with energy harvesting at UAVs using terrestrial base station	Considered UAV residual energy level with energy harvesting for executing successful UAV mission	Ignores the mobility prediction of UAVs from base station for beam steering and adaptive beamwidth control
RNN-based	[4]	PPO	CTDE	Solving an MFG by applying PPO-based continuous-discrete actor–critic	To predict the dynamic topology and stabilize the training process, it utilizes a learnable LSTM-based state encoder network	LSTM faces a vanishing gradient problem, and defined MFG can support only homogenous environments
	[48]	MA-DDPG	DTDE	Utilizing MLP and GRU-based adaptive communication channels for UAVSNs	MLP and GRU-based adaptive communication channels significantly reduce network flooding and compress communication data	UAV's relative trajectory knowledge and energy are not taken into consideration
	[24]	MA-DDPG	CTDE	Utilizing LSTM in both actor–critic networks to adaptively control the trajectory	An improved potential field-based reward accelerates the convergence to an optimal policy for twin-delayed MA-DDPG	LSTM cannot efficiently remember long trajectory sequences
Hybrid structure-based	[2]	MA-DDPG	DTDE	Proposing a swarming behavior-based mobility integrated LSTM-based actor and multi-head cross attention-based critic framework	State decoupling in LSTM-based actor network helps to achieve better policy and avoid environmental nonstationarity via multi-head cross-attention	LSTM-based actor may face overfitting and vanishing gradient problem
	[1]	MA-DDPG	CTDE	Proposing a behavior-integrated GRU-based actor and a multi-head attention-based critic network in hybrid action space	GRU-based actor network is capable of coping with dynamic observation. Multi-head attention-based critic network utilizes cross-attention to identify the influence of other neighboring agents	Vanishing gradient problem and struggle to remember large observation
	[10]	MA-DDPG	CTDE	Proposing a heterogeneous collaborative actor–critic framework based on multi-head attention	Learnable multi-head attention enhances the model's adaptability and the decision-making process	Higher computational complexity
	[49]	CTDE	MA-PPO	Leveraging multi-head attention in actor networks to extract key features from observations for exploring more targets	Addresses the simulation-to-real-world environment problem of MA-DRL algorithms by employing digital twin-assisted training	Action space for UAV flying directions is discrete, limiting smooth trajectory generation
	[5]	Probabilistic policy gradient	Federated learning	Proposing a federated heterogeneous CNN-based actor–critic frameworks	CNN can extract the areas of GUs that have higher computation demand and AoI	Higher computational complexity
	[52]	Probabilistic policy gradient	DTDE	Proposing an attention-weighted graph RNN-based state encoder utilizing the local knowledge	Highly scalable, utilizes log probability distribution-based policy gradient to support hybrid action space and fully decentralized	Higher computational complexity
	[53]	DDPG	CTDE	Applying a multi-relational graph convolutional state embedding-based actor–critic DDPG framework	Graph-structured information feeding with pooling in CNN helps to extract key features and reduce the dimension of graph-represented data	Requires higher computational complexity to train in real-time because GUs have very limited resources
Transformer-based encoder–decoder	[32]	MA-DDPG	CTDE	Proposing transformer-based encoder–decoder actor–critic frameworks based on MA-DDPG	Capable of remembering variable-length observation–action sequences, providing better scalability and adaptive state embedding	Higher computational complexity, critic network provides only self-attention and a long training time
QNN-based	[55]	Probabilistic policy gradient	CTDE	Mitigating quantum errors produced by the increasing number of qubits	Provides faster convergence in model training and enables real-time decision-making	Quantum hardware is still in its infancy, with problems like noise, qubit decoherence, and scalability of QNNs

According to Table 3, most of the algorithms considered MA-DDPG, MA-PPO, TRPO, and log probability policy to compute the policy gradient to train the actor network. Particularly, in [1,52], log probability-based policy distribution is considered with the help of the softmax activation function for discrete action and deterministic policy for continuous action. It helps to ensure a gradient for training the actor network in a hybrid action space. In [45], VAE is utilized to generate a hybrid latent space for both discrete and continuous action space from which samples are taken to obtain the policy gradient for training the actor network. Most of the algorithms applied centralized training and distributed execution (CTDE), distributed training and distributed execution (DTDE), and federated learning as training methods. In [48], adaptive communication channels based on MLP and GRU are designed and trained using a DTDE-based multi-agent actor–critic framework to reduce network flooding and communication data volume. In [49], the simulation-to-real-world implementation of MA-DRL-based actor–critic frameworks with multi-head attention mechanisms in the actor network is addressed through the introduction of digital twin-assisted training. The work in [55] leverages QNNs to accelerate the training convergence of MA-DRL algorithms while overcoming the limitations of scalability in QNN.

Even though existing frameworks have advantages in respective optimization objectives, one of the major challenges is the higher computational complexity of the neural network training and overfitting. It can be resolved by designing lightweight neural networks suitable for each agent’s computational resource and sampling the encoder–decoder structure by reducing the redundant computations as investigated in [57]. Additionally, a hybrid loss function can be designed based on the existing motion model of physics and data samples generated by the model to effectively control the flying direction, flight angle, velocity, and acceleration of UAVs in complex environments especially present of wind and obstacles.

4. Research Challenges and Directions

Based on the discussions in Sections 2 and 3, research challenges and potential solutions are summarized in this section as outcomes of this comprehensive survey.

4.1. Physics-Informed Neural Networks for Trajectory Control of UAVs

Existing model-free data-driven algorithms applied to control the trajectory of UAVs are trained neural networks based on either supervised learning or reinforcement learning. Physics-informed neural networks (PINNs) introduce UAV motion dynamics in the form of partial differential equations [58] or motion models into the neural networks as loss functions, augmenting the accuracy, generalization, and robustness of the model by ensuring predictions adhere to known physical principles. In addition, PINNs can assist in tackling uncertainties related to wind disturbance, UAV failures, or unexpected obstacles by embedding robust physical laws. Hence, PINNs can provide better training and algorithm convergence with fewer data samples considering real-life scenarios, which need further exploration.

4.2. Model Training, Knowledge Transfer, and Heterogeneous Environment

Existing multi-agent actor–critic frameworks have adopted centralized training and decentralized or distributed execution. However, considering the limited communication range of GUs and UAVs, distributed or federated training still requires further studies. An intelligent agent can use an attention mechanism to obtain the key features in multi-agent scenarios via a cross-attention mechanism in the form of query, key, and value by taking only partial observation–action as input. The hierarchical space air–ground integrated networks introduce multiple links and decision-makers as intelligent agents at

each layer [59]. Hence, a heterogeneous collaborative–competitive environment appears, which requires special neural networks for each agent according to its resource capacity to make the algorithm more robust and practical. Here, the potential for exploring transfer learning and knowledge distillation approaches from higher-capacity devices to lower-capacity devices is significant to enhance the overall system’s adaptability and efficiency [60,61]. The environment should be a high-fidelity environment for training, which can be achieved by employing digital twin-assisted training [12] and introducing noise not only in action space but also in state space to introduce uncertainties.

4.3. Transformer-Based Encoder–Decoder Multi-Agent Actor–Critic Frameworks

Actor neural networks require self-attention on their long observation sequences to predict the dynamic topology state by generating fixed-size state embedding and mapping it into the optimal action. In contrast, critic networks require cross-attention to pay attention to not only its own observation–action spaces but also to the neighboring agents in a multi-agent cooperative–competitive environment to precisely estimate the value function. This type of framework can deal with joint trajectory control, task offloading, and resource allocation problems that require hybrid action spaces in UAVSNs. Additionally, in space–air–ground integrated networks, computationally intensive and delay-critical tasks of GUs are offloaded based on binary or partially only to the hierarchical locations (UAVs, HAPs, or low earth orbit (LEO) satellites). Here, neighboring UAVs might have ideal resources that can be traced by paying attention to their observations to introduce both vertical and horizontal offloading [62]. Thus, low-complexity transformer-based self-attention encoder–decoder in actor and cross-attention-based encoder–decoder in critic networks require further studies to deal with these problems [57,63].

The high mobility of LEO satellites causes dynamic coverage toward UAVSNs or HAPs, which creates challenges for maintaining stable satellite-to-UAV or H2S links. Accurate prediction of LEO satellite mobility and pairing with accurate LEO satellite based on link quality [17], considering satellite constellations in orbit, is critical to ensure consistent coverage and reliability for these backhaul connections. To tackle this situation, the potential of generative adversarial learning within an actor–critic framework could be explored to enhance the adaptability and robustness of link prediction and resource allocation [64]. Additionally, the pre-trained large language models (LLMs) powered by transformer modules in the form of multi-agent actor–critic networks can be utilized for optimal trajectory control, task offloading, and resource allocation, which requires further investigations [65]. It is because LLMs are promising, owing to their capability to handle and understand vast amounts of data, capture complex time series patterns, and generalize across diverse tasks.

4.4. Channel State Estimation in UAVSNs

Owing to mobility and dense obstacles, both GU2U and U2U links are highly dynamic, and frequent link outages may occur. To adopt the mmWave and free space optic in UAVSNs, the power allocation, beamforming, beam training, adaptive beamwidth control, and controlling the phase shift from the reflected intelligent surface (RIS) require accurate channel state information by sending pilot signals [66–68]. Nevertheless, the frequent transmission of pilot signals triggers a waste of resources and delays in communication. Thus, channel state information in highly dynamic UAVSNs can be predicted by designing transformer-based encoder–decoder frameworks with less pilot transmission, which can be trained further by leveraging reinforcement learning [38,69].

4.5. Joint RIS-Assisted Adaptive Beamforming, Power, Caching, Computing, and Communication Resource Allocation in Dynamic UAVSNs

Considering the next generation of wireless communication, driven by the demand for computationally intensive and latency-sensitive applications such as virtual reality, augmented reality, and metaverse platforms, high-frequency spectrums like mmWave and terahertz are promising solutions to meet data rate and QoS requirements. However, these high-frequency solutions come with challenges such as complex antenna geometry design [70,71] and narrow beamforming in U2GU, U2U, and U2H links, which are critical for ensuring reliable communication and maintaining network topology [72]. Narrow beamforming also results in reduced coverage and introduces more complex issues in high-mobility scenarios involving ground users or UAVs. RIS-assisted integrated sensing and adaptive beamwidth-based beamforming from UAVs can mitigate these challenges by reducing frequent link outages and extending coverage [73–76]. Moreover, efficient resource allocation, including transmit power, bandwidth, computing, caching, and UAV trajectory optimization, significantly enhances overall system performance. Addressing the complexity of such optimizations, particularly in environments with multiple agents, dynamic state–action spaces, and nonlinear objectives and dynamic constraints, makes MA-DRL-based actor–critic frameworks highly suitable, which require further research attention.

4.6. Joint Model-Based and Model-Free MA-DRL Algorithms Development for UAVSNs

UAVSN-assisted missions, such as joint tasks or data collection, beam forming, task offloading, resource allocation (e.g., computing, caching, power, bandwidth, or timeslot), and trajectory optimization, involve large and dynamic state–action spaces with stringent short- and long-term constraints related to QoS and dynamic resource availability [77]. To address such joint optimization problems efficiently, formulating sub-problems is necessary to reduce computational complexity and enable real-time decision-making. A combination of model-based optimization and MA-DRL algorithms is highly suitable for these situations. Specifically, long-term constraints and queue stability under energy constraints can be simplified using the Lyapunov drift-plus-penalty method [7,78–80]. Alternatively, problems that allow for convex relaxation can be solved iteratively through model-based Lagrange relaxation, leveraging Karush–Kuhn–Tucker conditions and an adaptive step-size-based gradient descent method [81]. Simultaneously, control parameters requiring adaptive decision-making in dynamic environments with decentralized knowledge can be addressed using a meta-learning-inspired adaptive actor–critic MA-DRL framework [65].

During recursive state estimation for UAV trajectory prediction, particularly for tasks such as routing data packets [82] or task offloading, beamforming with neighboring UAVs [83] or between base stations and UAVs [84] can be enhanced by integrating model-free MA-DRL with model-based Kalman filters (e.g., extended Kalman filter or unscented Kalman filter) [85]. This hybrid approach can accelerate DRL convergence, improve optimal decision-making, and make it highly suitable for online execution. Kalman filter variations are well-suited for predicting nonlinear state dynamics, utilizing both motion models and noisy sensor data (e.g., inertial measurement units, gyroscopes, accelerometers, and light detection and ranging sensors) to enhance estimation accuracy [86]. Consequently, hybrid algorithms that combine model-based and model-free MA-DRL methods hold significant potential but require further investigations.

4.7. Joint Energy-Harvesting and Resource Allocation for Task Computation or Data Collection Using UAVSNs

UAVSNs can be leveraged to execute missions such as AoI-aware data collection or task computation for highly intensive, low-latency applications of low-power IoT devices. To extend the battery life of IoT devices, especially those located in remote areas,

one effective solution is energy harvesting at IoT devices using low-altitude UAVs with proper beamforming. Since most IoT devices are equipped with a single antenna, timeslot optimization is required along with optimal resource allocation for energy harvesting from UAVs and data/task uploading to UAVs, subject to AoI and resource constraints. Existing algorithms primarily use convex approximations and iterative methods to solve such dynamic optimization problems [87–91]. These approaches often rely on centralized knowledge and involve numerically computing the gradient of the objective function with respect to the control variables, resulting in higher time complexity, limited scalability, and reduced adaptability to environmental changes. Moreover, most existing studies assume that energy harvesting is linear in nature, whereas, in reality, it is nonlinear and depends on channel state conditions. To address the joint energy harvesting and resource allocation problem for AoI-aware task/data collection or computation, further research is required to develop adaptive MA-DRL-based frameworks [92]. These frameworks should account for the nonlinear nature of energy harvesting and provide scalable, real-time solutions that adapt to dynamic environments.

5. Conclusions

In this paper, the evolution of existing state-of-the-art multi-agent actor–critic frameworks is reviewed based on their internal neural network structure that is employed to solve various joint optimization problems in UAVSNs along with their advantages and limitations. According to this survey, for joint optimization in UAVSNs, particularly for UAV trajectory control and resource allocation, hybrid model-based and model-free MA-DRL-based actor–critic frameworks are more suitable for achieving better convergence and enabling online execution. Considering the heterogeneous environment, multiple decision-makers, and the limited computational resources of GUs/IoTs and UAVSNs, parameter-optimized neural network designs are essential, and federated learning methods can be utilized for efficient model training. Additionally, a learnable lightweight encoder–decoder structure based on transformer models appears more suitable for accurately estimating the dynamic state of UAVSNs, though further investigation is required. We believe this article will assist researchers in improving network performance in UAVSNs by addressing identified open issues through data-driven MA-DRL-based actor–critic frameworks.

Author Contributions: The individual contributions of the authors are as follows: M.M.A., S.A.T., T.H., S.M. and M.S.A. reviewed, analyzed, and summarized the actor–critic frameworks for performing joint optimizations in UAVSNs. This paper was drafted by M.M.A., S.A.T., T.H., S.M. and M.S.A., and was subsequently revised and approved by M.Y.A. All authors have read and agreed to the published version of the manuscript.

Funding: This work received no external funding.

Data Availability Statement: Not applicable.

Acknowledgments: The authors thank the editor and anonymous referees for their comments that helped improve the quality of this manuscript. The author M.M.A. thanks American International University—Bangladesh for providing a research incentive for this manuscript.

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

References

1. Morshed Alam, M.; Moh, S. Joint Optimization of Trajectory Control, Task Offloading, and Resource Allocation in Air-Ground Integrated Networks. *IEEE Internet Things J.* **2024**, *11*, 24273–24288. [\[CrossRef\]](#)
2. Alam, M.M.; Moh, S. Joint Trajectory Control, Frequency Allocation, and Routing for UAV Swarm Networks: A Multi-Agent Deep Reinforcement Learning Approach. *IEEE Trans. Mob. Comput.* **2024**, *23*, 11989–12005. [\[CrossRef\]](#)

3. Arafat, M.Y.; Moh, S. Localization and Clustering Based on Swarm Intelligence in UAV Networks for Emergency Communications. *IEEE Internet Things J.* **2019**, *6*, 8958–8976. [\[CrossRef\]](#)
4. Emami, Y.; Gao, H.; Li, K.; Almeida, L.; Tovar, E.; Han, Z. Age of Information Minimization Using Multi-Agent UAVs Based on AI-Enhanced Mean Field Resource Allocation. *IEEE Trans. Veh. Technol.* **2024**, *73*, 13368–13380. [\[CrossRef\]](#)
5. Li, H.; Zhang, J.; Zhao, H.; Ni, Y.; Xiong, J.; Wei, J. Joint Optimization on Trajectory, Computation and Communication Resources in Information Freshness Sensitive MEC System. *IEEE Trans. Veh. Technol.* **2024**, *73*, 4162–4177. [\[CrossRef\]](#)
6. Zhu, B.; Bedeer, E.; Nguyen, H.H.; Barton, R.; Gao, Z. UAV Trajectory Planning for AoI-Minimal Data Collection in UAV-Aided IoT Networks by Transformer. *IEEE Trans. Wirel. Commun.* **2023**, *22*, 1343–1358. [\[CrossRef\]](#)
7. Lee, W.; Kim, T. Multiagent Reinforcement Learning in Controlling Offloading Ratio and Trajectory for Multi-UAV Mobile-Edge Computing. *IEEE Internet Things J.* **2024**, *11*, 3417–3429. [\[CrossRef\]](#)
8. Chen, D.; Qi, Q.; Fu, Q.; Wang, J.; Liao, J.; Han, Z. Transformer-Based Reinforcement Learning for Scalable Multi-UAV Area Coverage. *IEEE Trans. Intell. Transp. Syst.* **2024**, *25*, 10062–10077. [\[CrossRef\]](#)
9. Alam, M.M.; Moh, S. Q-Learning-Based Routing Inspired by Adaptive Flocking Control for Collaborative Unmanned Aerial Vehicle Swarms. *Veh. Commun.* **2023**, *40*, 100572. [\[CrossRef\]](#)
10. Lu, Z.; Jia, Z.; Wu, Q.; Han, Z. Joint Trajectory Planning and Communication Design for Multiple UAVs in Intelligent Collaborative Air-Ground Communication Systems. *IEEE Internet Things J.* **2024**, *11*, 31053–31067. [\[CrossRef\]](#)
11. Arafat, M.Y.; Moh, S. Bio-Inspired Approaches for Energy-Efficient Localization and Clustering in Uav Networks for Monitoring Wildfires in Remote Areas. *IEEE Access* **2021**, *9*, 18649–18669. [\[CrossRef\]](#)
12. Liu, W.; Li, B.; Xie, W.; Dai, Y.; Fei, Z. Energy Efficient Computation Offloading in Aerial Edge Networks With Multi-Agent Cooperation. *IEEE Trans. Wirel. Commun.* **2023**, *22*, 5725–5739. [\[CrossRef\]](#)
13. Alam, M.M.; Arafat, M.Y.; Moh, S.; Shen, J. Topology Control Algorithms in Multi-Unmanned Aerial Vehicle Networks: An Extensive Survey. *J. Netw. Comput. Appl.* **2022**, *207*, 103495. [\[CrossRef\]](#)
14. Alam, M.M.; Moh, S. Survey on Q-Learning-Based Position-Aware Routing Protocols in Flying Ad Hoc Networks. *Electronics* **2022**, *11*, 1099. [\[CrossRef\]](#)
15. Sun, H.; Zhou, Y.; Zhang, H.; Ale, L.; Dai, H.; Zhang, N. Joint Optimization of Caching, Computing and Trajectory Planning in Aerial Mobile Edge Computing Networks: A MADDPG Approach. *IEEE Internet Things J.* **2024**, *11*, 40996–41007. [\[CrossRef\]](#)
16. Arafat, M.Y.; Moh, S. A Q-Learning-Based Topology-Aware Routing Protocol for Flying Ad Hoc Networks. *IEEE Internet Things J.* **2022**, *9*, 1985–2000. [\[CrossRef\]](#)
17. Zhang, S.; Cai, T.; Wu, D.; Schupke, D.; Ansari, N.; Cavdar, C. IoRT Data Collection With LEO Satellite-Assisted and Cache-Enabled UAV: A Deep Reinforcement Learning Approach. *IEEE Trans. Veh. Technol.* **2024**, *73*, 5872–5884. [\[CrossRef\]](#)
18. Paul, A.; Singh, K.; Nguyen, M.T.; Pan, C.; Li, C.-P. Digital Twin-Assisted Space-Air-Ground Integrated Networks for Vehicular Edge Computing. *IEEE J. Sel. Top. Signal Process.* **2024**, *18*, 66–82. [\[CrossRef\]](#)
19. Jia, Z.; Sheng, M.; Li, J.; Han, Z. Toward Data Collection and Transmission in 6G Space–Air–Ground Integrated Networks: Cooperative HAP and LEO Satellite Schemes. *IEEE Internet Things J.* **2022**, *9*, 10516–10528. [\[CrossRef\]](#)
20. Arafat, M.Y.; Alam, M.M.; Moh, S. Vision-Based Navigation Techniques for Unmanned Aerial Vehicles: Review and Challenges. *Drones* **2023**, *7*, 89. [\[CrossRef\]](#)
21. Arafat, M.Y.; Pan, S. Urban Air Mobility Communications and Networking: Recent Advances, Techniques, and Challenges. *Drones* **2024**, *8*, 702. [\[CrossRef\]](#)
22. Arafat, M.Y.; Moh, S. JRCS: Joint Routing and Charging Strategy for Logistics Drones. *IEEE Internet Things J.* **2022**, *9*, 21751–21764. [\[CrossRef\]](#)
23. Poudel, S.; Moh, S. Priority-Aware Task Assignment and Path Planning for Efficient and Load-Balanced Multi-UAV Operation. *Veh. Commun.* **2023**, *42*, 100633. [\[CrossRef\]](#)
24. Xing, X.; Zhou, Z.; Li, Y.; Xiao, B.; Xun, Y. Multi-UAV Adaptive Cooperative Formation Trajectory Planning Based on An Improved MATD3 Algorithm of Deep Reinforcement Learning. *IEEE Trans. Veh. Technol.* **2024**, *73*, 12484–12499. [\[CrossRef\]](#)
25. Poudel, S.; Moh, S. Task Assignment Algorithms for Unmanned Aerial Vehicle Networks: A Comprehensive Survey. *Veh. Commun.* **2022**, *35*, 100469. [\[CrossRef\]](#)
26. Li, S.; Yu, Z.; Dong, M.; Ota, K.; Chen, H.; Zhang, N.; Yang, C. Two-Hop Packet Scheduling, Resource Allocation, and UAV Trajectory Design for Internet of Remote Things in Air–Ground Integrated Network. *IEEE Internet Things J.* **2024**, *11*, 26160–26172. [\[CrossRef\]](#)
27. Kim, N.; Na, W.; Lakew, D.S.; Dao, N.; Cho, S. DQN-Based Directional MAC Protocol in Wireless Ad Hoc Network in Internet of Things. *IEEE Internet Things J.* **2024**, *11*, 12918–12928. [\[CrossRef\]](#)
28. Raivi, A.M.; Moh, S. JDACO: Joint Data Aggregation and Computation Offloading in UAV-Enabled Internet of Things for Post-Disaster Scenarios. *IEEE Internet Things J.* **2024**, *11*, 16529–16544. [\[CrossRef\]](#)
29. Yu, H.; Leng, S.; Wu, F. Joint Cooperative Computation Offloading and Trajectory Optimization in Heterogeneous UAV-Swarm-Enabled Aerial Edge Computing Networks. *IEEE Internet Things J.* **2024**, *11*, 17700–17711. [\[CrossRef\]](#)

30. Wu, Q.; Wang, W.; Fan, P.; Fan, Q.; Zhu, H.; Letaief, K.B. Cooperative Edge Caching Based on Elastic Federated and Multi-Agent Deep Reinforcement Learning in Next-Generation Networks. *IEEE Trans. Netw. Serv. Manag.* **2024**, *21*, 4179–4196. [\[CrossRef\]](#)
31. Iqbal, S.; Sha, F. Actor-Attention-Critic for Multi-Agent Reinforcement Learning. In Proceedings of the 36th International Conference on Machine Learning, ICML 2019, Long Beach, CA, USA, 9–15 June 2019; pp. 5261–5270.
32. Wu, G.; Liu, Z.; Fan, M.; Wu, K. Joint Task Offloading and Resource Allocation in Multi-UAV Multi-Server Systems: An Attention-Based Deep Reinforcement Learning Approach. *IEEE Trans. Veh. Technol.* **2024**, *73*, 11964–11978. [\[CrossRef\]](#)
33. Chen, D.; Qi, Q.; Zhuang, Z.; Wang, J.; Liao, J.; Han, Z. Mean Field Deep Reinforcement Learning for Fair and Efficient UAV Control. *IEEE Internet Things J.* **2021**, *8*, 813–828. [\[CrossRef\]](#)
34. Zhang, C.; Li, Z.; He, C.; Wang, K.; Pan, C. Deep Reinforcement Learning Based Trajectory Design and Resource Allocation for UAV-Assisted Communications. *IEEE Commun. Lett.* **2023**, *27*, 2398–2402. [\[CrossRef\]](#)
35. Ding, Y.; Yang, Z.; Pham, Q.-V.; Zhang, Z.; Shikh-Bahaei, M. Distributed Machine Learning for UAV Swarms: Computing, Sensing, and Semantics. *IEEE Internet Things J.* **2023**, *11*, 7447–7473. [\[CrossRef\]](#)
36. Baidya, T.; Moh, S. Comprehensive Survey on Resource Allocation for Edge-Computing-Enabled Metaverse. *Comput. Sci. Rev.* **2024**, *54*, 100680. [\[CrossRef\]](#)
37. McEnroe, P.; Wang, S.; Liyanage, M. A Survey on the Convergence of Edge Computing and AI for UAVs: Opportunities and Challenges. *IEEE Internet Things J.* **2022**, *9*, 15435–15459. [\[CrossRef\]](#)
38. Zayat, A.; Hasabelnaby, M.A.; Obeed, M.; Chaaban, A. Transformer Masked Autoencoders for Next-Generation Wireless Communications: Architecture and Opportunities. *IEEE Commun. Mag.* **2024**, *62*, 88–94. [\[CrossRef\]](#)
39. Nabi, A.; Baidya, T.; Moh, S. Comprehensive Survey on Reinforcement Learning-Based Task Offloading Techniques in Aerial Edge Computing. *Internet Things* **2024**, *28*, 101342. [\[CrossRef\]](#)
40. Baidya, T.; Nabi, A.; Moh, S. Trajectory-Aware Offloading Decision in UAV-Aided Edge Computing: A Comprehensive Survey. *Sensors* **2024**, *24*, 1837. [\[CrossRef\]](#)
41. Raivi, A.M.; Moh, S. A Comprehensive Survey on Data Aggregation Techniques in UAV-Enabled Internet of Things. *Comput. Sci. Rev.* **2023**, *50*, 100599. [\[CrossRef\]](#)
42. Chen, J.; Xing, H.; Xiao, Z.; Xu, L.; Tao, T. A DRL Agent for Jointly Optimizing Computation Offloading and Resource Allocation in MEC. *IEEE Internet Things J.* **2021**, *8*, 17508–17524. [\[CrossRef\]](#)
43. Li, K.; Ni, W.; Dressler, F. LSTM-Characterized Deep Reinforcement Learning for Continuous Flight Control and Resource Allocation in UAV-Assisted Sensor Network. *IEEE Internet Things J.* **2022**, *9*, 4179–4189. [\[CrossRef\]](#)
44. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention Is All You Need. *Adv. Neural Inf. Process. Syst.* **2017**, *2017*, 5999–6009.
45. Hao, H.; Xu, C.; Zhang, W.; Yang, S.; Muntean, G.-M. Joint Task Offloading, Resource Allocation, and Trajectory Design for Multi-UAV Cooperative Edge Computing With Task Priority. *IEEE Trans. Mob. Comput.* **2024**, *23*, 8649–8663. [\[CrossRef\]](#)
46. Qiu, X.; Xu, L.; Wang, P.; Yang, Y.; Liao, Z. A Data-Driven Packet Routing Algorithm for an Unmanned Aerial Vehicle Swarm: A Multi-Agent Reinforcement Learning Approach. *IEEE Wirel. Commun. Lett.* **2022**, *11*, 2160–2164. [\[CrossRef\]](#)
47. Ouamri, M.A.; MacHter, Y.; Singh, D.; Alkama, D.; Li, X. Joint Energy Efficiency and Throughput Optimization for UAV-WPT Integrated Ground Network Using DDPG. *IEEE Commun. Lett.* **2023**, *27*, 3295–3299. [\[CrossRef\]](#)
48. Wang, Z.; Yao, H.; Mai, T.; Xiong, Z.; Wu, X.; Wu, D.; Guo, S. Learning to Routing in UAV Swarm Network: A Multi-Agent Reinforcement Learning Approach. *IEEE Trans. Veh. Technol.* **2023**, *72*, 6611–6624. [\[CrossRef\]](#)
49. Cao, P.; Lei, L.; Shen, G.; Cai, S.; Liu, X.; Liu, X.; Tian, S. UAV Swarm Cooperative Search Based on Scalable Multiagent Deep Reinforcement Learning with Digital Twin-Enabled Sim-to-Real Transfer. *IEEE Trans. Mob. Comput.* **2025**, 1–17. [\[CrossRef\]](#)
50. Pu, Z.; Wang, H.; Liu, Z.; Yi, J.; Wu, S. Attention Enhanced Reinforcement Learning for Multi Agent Cooperation. *IEEE Trans. Neural Netw. Learn. Syst.* **2023**, *34*, 8235–8249. [\[CrossRef\]](#) [\[PubMed\]](#)
51. Xiao, J.; Yuan, G.; He, J.; Fang, K.; Wang, Z. Graph Attention Mechanism Based Reinforcement Learning for Multi-Agent Flocking Control in Communication-Restricted Environment. *Inf. Sci.* **2023**, *620*, 142–157. [\[CrossRef\]](#)
52. Gao, Z.; Yang, L.; Dai, Y. Fast Adaptive Task Offloading and Resource Allocation in Large-Scale MEC Systems via Multi-Agent Graph Reinforcement Learning. *IEEE Internet Things J.* **2023**, *11*, 758–776. [\[CrossRef\]](#)
53. Wu, H.; Tian, L.; Tang, H.; Li, R.; Jiao, P. Graph Convolutional Reinforcement Learning-Guided Joint Trajectory Optimization and Task Offloading for Aerial Edge Computing. *IEEE Trans. Intell. Transp. Syst.* **2024**, 1–12. [\[CrossRef\]](#)
54. Park, S.; Kim, J.P.; Park, C.; Jung, S.; Kim, J. Quantum Multi-Agent Reinforcement Learning for Autonomous Mobility Cooperation. *IEEE Commun. Mag.* **2024**, *62*, 106–112. [\[CrossRef\]](#)
55. Park, C.; Yun, W.J.; Kim, J.P.; Rodrigues, T.K.; Park, S.; Jung, S.; Kim, J. Quantum Multiagent Actor-Critic Networks for Cooperative Mobile Access in Multi-UAV Systems. *IEEE Internet Things J.* **2023**, *10*, 20033–20048. [\[CrossRef\]](#)
56. Park, S.; Kim, G.S.; Han, Z.; Kim, J. Quantum Multi-Agent Reinforcement Learning Is All You Need: Coordinated Global Access in Integrated TN/NTN Cube-Satellite Networks. *IEEE Commun. Mag.* **2024**, *62*, 86–92. [\[CrossRef\]](#)
57. He, B.; Hofmann, T. Simplifying Transformer Blocks. *arXiv* **2023**, arXiv:2311.01906.

58. Jin, L.; Liu, L.; Wang, X.; Shang, M.; Wang, F.Y. Physical-Informed Neural Network for MPC-Based Trajectory Tracking of Vehicles With Noise Considered. *IEEE Trans. Intell. Veh.* **2024**, *9*, 4493–4503. [\[CrossRef\]](#)
59. Zhao, K.; Peng, L.; Tak, B. Joint DRL-Based UAV Trajectory Planning and TEG-Based Task Offloading. *IEEE Trans. Consum. Electron.* **2025**, *1*. [\[CrossRef\]](#)
60. Jiang, K.; Cai, X.; Cui, Z.; Li, A.; Ren, Y.; Yu, H.; Yang, H.; Fu, D.; Wen, L.; Cai, P. KoMA: Knowledge-Driven Multi-Agent Framework for Autonomous Driving With Large Language Models. *IEEE Trans. Intell. Veh.* **2024**, 1–15. [\[CrossRef\]](#)
61. Zhou, L.; Deng, X.; Wang, Z.; Zhang, X.; Dong, Y.; Hu, X.; Ning, Z.; Wei, J. Semantic Information Extraction and Multi-Agent Communication Optimization Based on Generative Pre-Trained Transformer. *IEEE Trans. Cogn. Commun. Netw.* **2024**, *1*. [\[CrossRef\]](#)
62. Hoa, N.T.; Dai, D.V.; Lan, L.H.; Luong, N.C.; Van Le, D.; Niyato, D. Deep Reinforcement Learning for Multi-Hop Offloading in UAV-Assisted Edge Computing. *IEEE Trans. Veh. Technol.* **2023**, *642*, 16917–16922. [\[CrossRef\]](#)
63. Lan, Z.; Liu, L.; Fan, B.; Lv, Y.; Ren, Y.; Cui, Z. Traj-LLM: A New Exploration for Empowering Trajectory Prediction With Pre-Trained Large Language Models. *IEEE Trans. Intell. Veh.* **2024**, 1–14. [\[CrossRef\]](#)
64. Guo, Q.; Tang, F.; Kato, N. Routing for Space-Air-Ground Integrated Network With GAN-Powered Deep Reinforcement Learning. *IEEE Trans. Cogn. Commun. Netw.* **2024**, *1*. [\[CrossRef\]](#)
65. Zhou, W.; Li, Y.; Yang, Y.; Wang, H.; Hospedales, T.M. Online Meta-Critic Learning for Off-Policy Actor-Critic Methods. *Adv. Neural Inf. Process. Syst.* **2020**, *2020*, 17662–17673.
66. Xiao, Z.; Han, Z.; Nallanathan, A.; Dobre, O.A.; Clerckx, B.; Choi, J.; He, C.; Tong, W. Antenna Array Enabled Space/Air/Ground Communications and Networking for 6G. *IEEE J. Sel. Areas Commun.* **2022**, *40*, 2773–2804. [\[CrossRef\]](#)
67. Alam, M.M.; Moh, S. Survey on Neighbor Discovery and Beam Alignment in MmWave-Enabled UAV Swarm Networks. In Proceedings of the 11th International Conference on Smart Media and Applications, Saipan, NM, USA, 14–17 December 2022.
68. Tariq, S.; Arfeto, B.E.; Khalid, U.; Kim, S.; Duong, T.Q.; Shin, H. Deep Quantum-Transformer Networks for Multimodal Beam Prediction in ISAC Systems. *IEEE Internet Things J.* **2024**, *11*, 29387–29401. [\[CrossRef\]](#)
69. Jiang, H.; Cui, M.; Ng, D.W.K.; Dai, L. Accurate Channel Prediction Based on Transformer: Making Mobility Negligible. *IEEE J. Sel. Areas Commun.* **2022**, *40*, 2717–2732. [\[CrossRef\]](#)
70. Zhang, J.; Xu, W.; Wang, F.; Liu, Y.; Pan, M. Codebook Design and Beam Training for Conformal Array-Enabled UAV MmWave Communications With the Airframe Shadowing. *IEEE Wirel. Commun. Lett.* **2023**, *12*, 2248–2252. [\[CrossRef\]](#)
71. Zhang, J.; Xu, W.; Gao, H.; Pan, M.; Han, Z.; Zhang, P. Codebook-Based Beam Tracking for Conformal Array-Enabled UAV MmWave Networks. *IEEE Internet Things J.* **2021**, *8*, 244–261. [\[CrossRef\]](#)
72. Xiong, K.; Wang, R.; Leng, S.; Huang, C.; Yuen, C. RIS-Empowered Topology Control for Decentralized Federated Learning in Urban Air Mobility. *IEEE Internet Things J.* **2024**, *11*, 40757–40770. [\[CrossRef\]](#)
73. Karaçora, Y.; Chaccour, C.; Sezgin, A.; Saad, W. Event-Based Beam Tracking With Dynamic Beamwidth Adaptation in Terahertz (THz) Communications. *IEEE Trans. Commun.* **2023**, *71*, 6195–6210. [\[CrossRef\]](#)
74. Elbir, A.M.; Mishra, K.V.; Vorobyov, S.A.; Heath, R.W. Twenty-Five Years of Advances in Beamforming: From Convex and Nonconvex Optimization to Learning Techniques. *IEEE Signal Process. Mag.* **2022**, *40*, 118–131. [\[CrossRef\]](#)
75. Wang, W.; Fei, Z.; Guo, J.; Durrani, S.; Yanikomeroglu, H. Outage Performance of Multi-Tier UAV Communication with Random Beam Misalignment. *IEEE Internet Things J.* **2023**, *11*, 4163–4178. [\[CrossRef\]](#)
76. Deng, N.; Shi, X.; Sheng, M.; Liu, J.; Wei, H.; Zhao, N.; Niyato, D. Enhancing Millimeter Wave Cellular Networks via UAV-Borne Aerial IRS Swarms. *IEEE Trans. Commun.* **2024**, *72*, 524–538. [\[CrossRef\]](#)
77. Liu, J.; Zhao, X.; Qin, P.; Du, F.; Chen, Z.; Zhou, H.; Li, J. Joint UAV 3D Trajectory Design and Resource Scheduling for Space-Air-Ground Integrated Power IoRT: A Deep Reinforcement Learning Approach. *IEEE Trans. Netw. Sci. Eng.* **2024**, *11*, 2632–2646. [\[CrossRef\]](#)
78. Qin, P.; Fu, Y.; Xie, Y.; Wu, K.; Zhang, X.; Zhao, X. Multi-Agent Learning-Based Optimal Task Offloading and UAV Trajectory Planning for AGIN-Power IoT. *IEEE Trans. Commun.* **2023**, *71*, 4005–4017. [\[CrossRef\]](#)
79. Chen, J.; Wu, Z. Dynamic Computation Offloading with Energy Harvesting Devices: A Graph-Based Deep Reinforcement Learning Approach. *IEEE Commun. Lett.* **2021**, *25*, 2968–2972. [\[CrossRef\]](#)
80. Li, B.; Yang, R.; Liu, L.; Wang, J.; Zhang, N.; Dong, M. Robust Computation Offloading and Trajectory Optimization for Multi-UAV-Assisted MEC: A Multiagent DRL Approach. *IEEE Internet Things J.* **2024**, *11*, 4775–4786. [\[CrossRef\]](#)
81. Hoang, L.T.; Nguyen, C.T.; Pham, A.T. Deep Reinforcement Learning-Based Online Resource Management for UAV-Assisted Edge Computing With Dual Connectivity. *IEEE/ACM Trans. Netw.* **2023**, *31*, 2761–2776. [\[CrossRef\]](#)
82. Liu, C.; Zhang, G.; Guo, W.; He, R. Kalman Prediction-Based Neighbor Discovery and Its Effect on Routing Protocol in Vehicular Ad Hoc Networks. *IEEE Trans. Intell. Transp. Syst.* **2020**, *21*, 159–169. [\[CrossRef\]](#)
83. Gao, H.; Jia, C.; Xu, W.; Yuen, C.; Feng, Z.; Lu, Y. Machine Learning-Empowered Beam Management for MmWave-NOMA in Multi-UAVs Networks. *IEEE Trans. Veh. Technol.* **2022**, *71*, 8487–8502. [\[CrossRef\]](#)

84. Chang, B.; Tang, W.; Yan, X.; Tong, X.; Chen, Z. Integrated Scheduling of Sensing, Communication, and Control for MmWave/THz Communications in Cellular Connected UAV Networks. *IEEE J. Sel. Areas Commun.* **2022**, *40*, 2103–2113. [\[CrossRef\]](#)
85. Yan, L.; Fang, X.; Fang, Y.; Hao, L.; Xue, Q.; Xu, C. KF-LSTM Based Beam Tracking for UAV-Assisted MmWave HSR Wireless Networks. *IEEE Trans. Veh. Technol.* **2022**, *71*, 10796–10807. [\[CrossRef\]](#)
86. Buchnik, I.; Steger, D.; Revach, G.; van Sloun, R.J.G.; Routtenberg, T.; Shlezinger, N. Learned Kalman Filtering in Latent Space with High-Dimensional Data. In Proceedings of the ICASSP 2023—2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Rhodes, Greece, 4–10 June 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 1–5.
87. Lu, X.; Yang, W.; Yan, S.; Li, Z.; Ng, D.W.K. Covertness and Timeliness of Data Collection in UAV-Aided Wireless-Powered IoT. *IEEE Internet Things J.* **2022**, *9*, 12573–12587. [\[CrossRef\]](#)
88. Liu, Y.; Dai, H.N.; Wang, Q.; Imran, M.; Guizani, N. Wireless Powering Internet of Things with UAVs: Challenges and Opportunities. *IEEE Netw.* **2022**, *36*, 146–152. [\[CrossRef\]](#)
89. Jeganathan, A.; Dhayabaran, B.; Perera, T.D.P.; Jayakody, D.N.K.; Muthuchidambaranathan, P. Harvest-on-Sky: An AoI-Driven UAV-Assisted Wireless Communication System. *IEEE Internet Things Mag.* **2022**, *5*, 142–146. [\[CrossRef\]](#)
90. Qi, C.; Wang, J.; Lyu, L.; Tan, L.; Zhang, J.; Li, G.Y. Key Issues in Wireless Transmission for NTN-Assisted Internet of Things. *IEEE Internet Things Mag.* **2024**, *7*, 40–46. [\[CrossRef\]](#)
91. Zhang, S.; Liu, W.; Ansari, N. Joint Wireless Charging and Data Collection for UAV-Enabled Internet of Things Network. *IEEE Internet Things J.* **2022**, *9*, 23852–23859. [\[CrossRef\]](#)
92. Zheng, K.; Luo, R.; Liu, X.; Qiu, J.; Liu, J. Distributed DDPG-Based Resource Allocation for Age of Information Minimization in Mobile Wireless-Powered Internet of Things. *IEEE Internet Things J.* **2024**, *11*, 29102–29115. [\[CrossRef\]](#)

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