



REVIEW

A Bibliometric Analysis of Deep Reinforcement Learning in UAV Path Planning

Qiwu Wu¹, Tao Yang^{2,*}, Yunchen Su², Lingzhi Jiang³ and Tao Tong²

¹Key Laboratory of Counter-Terrorism Command & Information Engineering, Engineering University of PAP, Xi'an, China

²School of Equipment Management and Support, Engineering University of PAP, Xi'an, China

³School of Information Engineering, Engineering University of PAP, Xi'an, China

*Corresponding Author: Tao Yang. Email: yangtao_pap@163.com or 2246423010@stu.xjtu.edu.cn

Received: 22 March 2026; Accepted: 15 May 2026; Published: 23 July 2026

ABSTRACT: Deep reinforcement learning (DRL) has become an important method in Unmanned Aerial Vehicle(UAV) path planning, but the field still lacks a dedicated bibliometric review that summarizes its publication patterns, intellectual structure, and thematic evolution. This study analyzes 1402 Web of Science publications from 2010 to 2025 using CiteSpace, VOSviewer, and the Bibliometrix R package. Three main findings are reported. *First*, the bibliometric evidence suggests a four-phase evolution of the field—foundational exploration (2015–2016), continuous-control breakthrough (2017–2019), multi-agent collaborative coordination (2020–2022), and complex-scenario integration (2023–2025)—as reflected in publication trends, keyword bursts, and co-citation clusters. *Second*, co-citation and keyword analyses indicate a gradual shift from geometric navigation toward the joint consideration of communication, energy, and mission objectives, a pattern also reflected in the prominence of Internet of Things (IoT) and vehicular-technology journals. *Third*, burst and clustering results highlight several active research directions, including multi-objective cooperative decision-making, hierarchical planning architectures that combine global perception with local control, and communication–energy co-design. Rather than offering prescriptive conclusions, these results provide descriptive bibliometric evidence that may help researchers understand the development and emerging priorities of DRL-based UAV path planning.

KEYWORDS: Deep reinforcement learning; UAV path planning; bibliometric analysis; co-citation analysis; keyword burst detection

1 Introduction

Unmanned aerial vehicle (UAV) path planning has evolved from manual waypoint charting into a computationally intensive discipline, driven by advances in satellite positioning, onboard sensing, and automated flight control [1,2]. As mission complexity grew—encompassing large-scale combinatorial search, real-time energy allocation, and dynamic re-optimization under uncertainty—artificial intelligence techniques became indispensable [3–5].

Over the past decade, deep reinforcement learning (DRL) has become a major method in this field, with annual publication growth exceeding 44%. However, existing reviews still leave an important analytical gap. Prior surveys mainly fall into three categories:

(1) *Comprehensive reviews* construct multi-dimensional algorithm taxonomies and synthesize cross-domain challenges, but remain anchored in qualitative narrative—they catalog what exists without quantifying how the field's intellectual structure has shifted over time [6,7].

(2) *Bio-inspired algorithm reviews* dissect computational complexity and benchmark hybrid meta-heuristics, yet confine their lens to intra-algorithmic competition. This narrow focus obscures the broader trend that DRL is increasingly combined with communication theory, edge computing, and swarm intelligence [8,9].

(3) *Scenario-specific reviews* deliver depth within individual application areas such as remote sensing and precision agriculture, but fragment the overall picture. By separating insights by scenario, they miss the cross-domain shift from geometric shortest-path computation toward joint optimization of communication, energy, and mission-level objectives [10].

These limitations have a common feature: all three categories rely mainly on qualitative synthesis and expert judgment, without a reproducible, data-driven account of how the field has evolved. At present, there is still no large-scale body of research on reinforcement learning-based UAV path planning.

This paper addresses that gap. We conduct a bibliometric analysis of 1402 publications from the Web of Science Core Collection (2010–2025) using CiteSpace, VOSviewer, and the Bibliometrix R package. The main contributions are as follows:

(1) We identify a four-phase evolutionary trajectory spanning foundational exploration (2015–2016), continuous-control breakthrough (2017–2019), multi-agent collaborative coordination (2020–2022), and complex-scenario integration (2023–2025), and summarize the key technical changes and remaining bottlenecks in each phase.

(2) We construct a quantitative profile of publication trends, co-citation clusters, keyword co-occurrence networks, and burst patterns, providing a reproducible analytical baseline for the field.

(3) We provide bibliometric support for three research directions: multi-objective cooperative decision-making, hierarchical perception-planning-control architectures, and communication–energy co-design.

The remainder of this paper is organized as follows. [Section 3](#) details the data sources and retrieval strategy. [Section 4](#) profiles the macro-level publication landscape. [Section 5](#) maps collaborative networks across authors, institutions, countries, and disciplines. [Section 6](#) analyzes co-citation relationships and burst dynamics. [Section 7](#) investigates keyword clustering and temporal emergence patterns. [Section 8](#) synthesizes these bibliometric findings into a four-phase evolutionary framework that decomposes the discipline’s evolution according to the depth of DRL integration. [Section 9](#) concludes with limitations and future directions.

2 Related Work

Existing studies most closely related to the present work can be broadly divided into two streams: narrative or systematic reviews on UAV path planning, reinforcement learning, and deep reinforcement learning, and bibliometric studies on UAV research as a whole, UAV path planning as a whole, or adjacent themes related to reinforcement-learning-based UAV path planning.

In the non-bibliometric stream, prior reviews have mainly examined UAV path planning as a broad problem and summarized graph-search methods, sampling-based methods, metaheuristics, evolutionary computation, and learning-driven approaches, often comparing them in terms of path length, computational complexity, obstacle avoidance capability, energy constraints, and real-time feasibility. Representative examples include the survey by Aggarwal and Kumar [10], the systematic literature review by Husnain et al. [11], the analytical review by Luo et al. [12], and the remote-sensing-oriented review by Debnath et al. [13]. These studies provide an important broad framework for understanding UAV path planning. However, their primary unit of analysis remains UAV path planning in general, where reinforcement learning is usually

treated as only one algorithmic branch among many rather than as a rapidly expanding and increasingly independent research domain with its own evolutionary logic.

More closely related to the present study are reviews on reinforcement learning or deep reinforcement learning for UAV navigation. Azar et al. reviewed drone deep reinforcement learning across guidance, navigation, and control tasks and pointed out that most existing approaches are still trained in simulated environments, with real-world deployment constrained by limited generalization, safety concerns, and environmental complexity [14]. AlMahamid and Grolinger further provided a systematic review of reinforcement learning for autonomous UAV navigation, covering control, obstacle avoidance, path planning, flocking, and related navigation frameworks, and organizing the literature according to environment type, algorithm properties, and simulation platforms [15]. The importance of this line of work lies in recognizing the central role of RL and DRL in UAV autonomy. Nevertheless, its unit of analysis is still the broader notion of autonomous UAV navigation, in which path planning appears only as one subtask alongside attitude control, trajectory tracking, flocking, and target pursuit. As a result, these studies provide limited insight into the internal knowledge structure and thematic migration of reinforcement-learning-based UAV path planning itself.

A third line of closely related reviews is scenario-specific or multi-UAV RL-oriented research. Puente-Castro et al. reviewed artificial intelligence for UAV swarm path planning and grouped existing approaches into reinforcement learning, evolutionary computation, swarm intelligence, and graph neural networks [16]. Bai et al. surveyed reinforcement-learning-based approaches for autonomous multi-UAV wireless networks, focusing on task allocation, trajectory planning, resource management, and communication efficiency [17]. Amodu et al. reviewed deep reinforcement learning in UAV-assisted IoT data collection and showed that DRL is often used not only for path planning, but also for velocity and altitude control, power allocation, scheduling, and charging optimization [18]. These studies help clarify multi-agent coordination, communication coupling, and task-constrained optimization. However, they are limited to specific application scenarios, and path planning is usually embedded in a broader system-level optimization framework, which weakens the treatment of DRL-based UAV path planning as an independent research problem.

From a bibliometric perspective, directly relevant studies remain very limited. Based on the current literature search, the most direct related work is the Springer proceedings chapter by Borrego-Dominguez et al., *Reinforcement Learning for Drone Path Planning: A Bibliometric Analysis* [19]. Using Scopus as the data source, that study analyzed 98 publications from 2018 to August 2024 and employed VOSviewer and Litmaps to describe annual publication trends, leading authors, funding sponsors, country contributions, subject areas, and publication types. Its main contribution is to show that reinforcement learning for drone path planning has become identifiable as a relatively independent research set. However, the study also has clear limitations: it is a conference-book chapter rather than a standalone journal article; its corpus is relatively small and temporally limited; and its analysis remains largely descriptive, without systematically examining co-citation structure, burst references, keyword evolution, disciplinary interaction, or stage-wise knowledge migration.

More mature bibliometric studies are available only at adjacent levels. For example, Wu et al. analyzed UAV path planning as a whole from 2000 to 2024 and identified broad trends such as multi-agent coordination, obstacle avoidance, intelligent planning in urban environments, and communication–trajectory co-optimization [20]. Guzsvinecz and Szűcs examined UAV research more broadly in terms of publication growth, thematic evolution, and collaboration patterns [21]. Although these studies provide valuable macro-level background, their units of analysis are general UAV path planning or general UAV research, where reinforcement learning appears only as one emerging topic rather than as an independently reconstructed knowledge domain.

Taken together, the existing literature still reveals a common gap: there is a lack of a study that treats deep-reinforcement-learning-based UAV path planning as an independent knowledge domain and systematically reconstructs its evolution. Existing reviews are either too broad, embedding DRL within larger UAV path planning or UAV navigation frameworks, or too narrow, focusing only on swarms, remote sensing, IoT, or wireless communication scenarios. Existing bibliometric studies are either limited in sample size and analytical depth or too broad in scope to reveal when DRL emerged as a dominant paradigm in UAV path planning, how the field migrated from geometric routing and local obstacle avoidance toward multi-agent coordination, communication coupling, and energy-aware joint optimization, and which key references and disciplinary linkages shaped its current knowledge base and research frontiers. The present study is intended to fill this gap.

3 Data Source

The data utilized in this study were retrieved from the Web of Science Core Collection (WoSCC), a globally recognized and authoritative citation database. The credibility of this data stems from the rigorous evaluation and selection process that journals must undergo for inclusion in WoSCC, which adheres to strict criteria to index only the most influential and high-quality publications across various disciplines. This selective nature ensures the authority and academic value of the sourced literature, effectively minimizing interference from non-core publications. Consequently, analytical results derived from this dataset can more accurately reflect the genuine frontiers and foundational knowledge of the research domain.

Furthermore, WoSCC provides the highly standardized data structure essential for sophisticated bibliometric analysis. Beyond fundamental bibliographic details, it offers comprehensive metadata, including complete reference lists, citation counts, author affiliations, and institutional addresses. This comprehensiveness is critical for conducting in-depth analyses such as co-citation studies, research collaboration network mapping, and knowledge flow tracing—capabilities where many alternative databases are limited.

The database also offers extensive historical coverage, with a substantial corpus of core literature dating back to 1900, thereby facilitating robust longitudinal trend analysis. Its constituent sub-databases, notably the Science Citation Index Expanded (SCIE) and the Social Sciences Citation Index (SSCI), collectively encompass publications from the natural sciences, social sciences, arts, and humanities. This broad interdisciplinary coverage provides a solid foundation for cross-disciplinary research analysis [22].

The literature search encompassed document types including “article”, “review”, “editorial”, and “proceedings paper”, covering the period from January 2010 to December 2025. The core search query was constructed using key terms such as “UAV”, “path planning”, “deep reinforcement learning”, and “deep learning”. To ensure comprehensive retrieval and account for terminological variations in the academic literature, a set of synonymous and semantically related terms was incorporated to broaden the search scope. The derived search expression was as follows:

TS = (((“Drone” OR “Drones” OR “Unmanned Aerial Vehicle” OR “Unmanned Aerial Vehicles” OR “Unmanned Aircraft System” OR “Unmanned Aircraft Systems” OR “Remotely Piloted Aircraft” OR “Quadcopter” OR “Quadcopters” OR “Multicopter” OR “Multicopters” OR “Hexacopter” OR “Hexacopters” OR “Octocopter” OR “Octocopters” OR “Tricopter” OR “Tricopters” OR “VTOL UAV” OR “VTOL UAVs”) AND (“Deep Learning” OR “Deep Reinforcement Learning” OR “Reinforcement Learning”) AND (“Path Planning” OR “Path Optimization” OR “Route Planning” OR “Route Optimization” OR “Path Minimization” OR “Trajectory Optimization” OR “Motion Planning” OR “Navigation Optimization”)))

The initial search yielded a total of 2319 publications. Following a two-stage screening process—automated filtering of duplicates and non-research document types, followed by manual title–abstract–full-text review to exclude studies not substantively addressing DRL-based UAV path planning—a final corpus of 1402 papers was retained for analysis. The complete screening workflow, including exclusion criteria and the number of records removed at each stage, is depicted in Fig. 1.

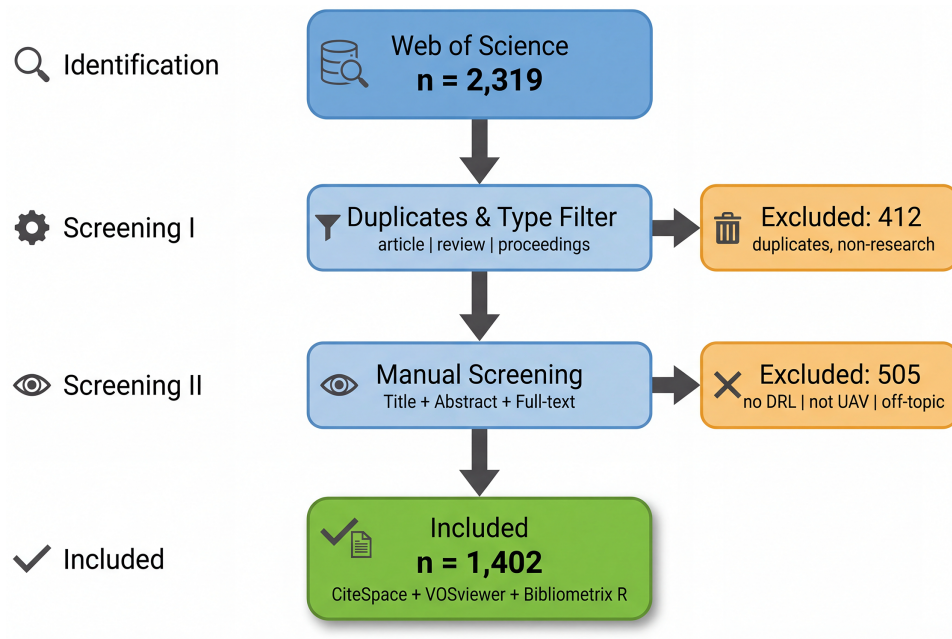


Figure 1: Literature screening flowchart. Note: Studies were screened by manually examining the title and abstract of each paper to assess its relevance to the present study.

The broader term “reinforcement learning” was retained in the search query to capture publications that apply RL methods subsequently enhanced with deep networks, as well as comparative studies; traditional RL papers without deep learning components were excluded during manual screening.

To improve methodological transparency and reproducibility, the main software settings and analytical parameters used in this bibliometric analysis are summarized in Table 1. These settings include software versions, time slicing, node-selection criteria, threshold values, visualization ranges, and source-specific analytical configurations.

Table 1: Reproducible configuration.

Analysis in This Manuscript	Software	Parameter Settings
Publication trend, source distribution, and descriptive statistics	Bibliometrix R package	Plain-text files exported from the Web of Science Core Collection were imported to calculate annual publication counts, source distribution, country/region output, institutional output, author output, and cumulative descriptive indicators.

(Continued)

Table 1 (continued)

Analysis in This Manuscript	Software	Parameter Settings
Institutional collaboration network	VOSviewer 1.6.20	Minimum number of documents of an organization = 5.
Country/region collaboration network	VOSviewer 1.6.20	Country/region cooperation relationships were visualized as a collaboration network; node size represents publication volume and link thickness represents collaboration intensity.
Disciplinary domain analysis and dual-map overlay	CiteSpace 6.4.1	Dual-map overlay parameters: $N = 40$, $N\% = 10\%$, and $k = 25$.
Reference co-citation network	CiteSpace 6.4.1	Timespan: 2014–2025; slice length = 1; node-selection criterion: g -index with $k = 25$.
Co-citation burst analysis	CiteSpace 6.4.1	Timespan: 2010–2025; slice length = 1; node-selection criterion: g -index with $k = 25$; burst references were detected using the burst-detection function in CiteSpace.
Keyword co-occurrence and clustering	VOSviewer 1.6.20	Minimum keyword occurrence threshold = 20; 56 major keywords were retained; 481 low-frequency terms were filtered out.
Temporal keyword overlay visualization	VOSviewer 1.6.20	Temporal keyword analysis focused on keyword appearances after 2022; overlay visualization was generated with Top $N = 10$.
Keyword burst intensity	CiteSpace 6.4.1	Keyword burst intensity was calculated for the 2010–2025 period using the burst-detection function in CiteSpace.
Subject-area timeline analysis	CiteSpace 6.4.1	Timespan: 2010–2025; slice length = 1; node-selection criterion: g -index with $k = 25$.

The bibliometric analysis was conducted utilizing a complementary suite of three analytical tools: CiteSpace (version 6.4.1) for co-citation network construction and burst detection, VOSviewer (version 1.6.20) for keyword co-occurrence clustering and overlay visualization, and the Bibliometrix R package for publication trend profiling and descriptive statistics. The integrated analytical framework linking data collection, tool deployment, analysis dimensions, and output synthesis is summarized in [Fig. 2](#).

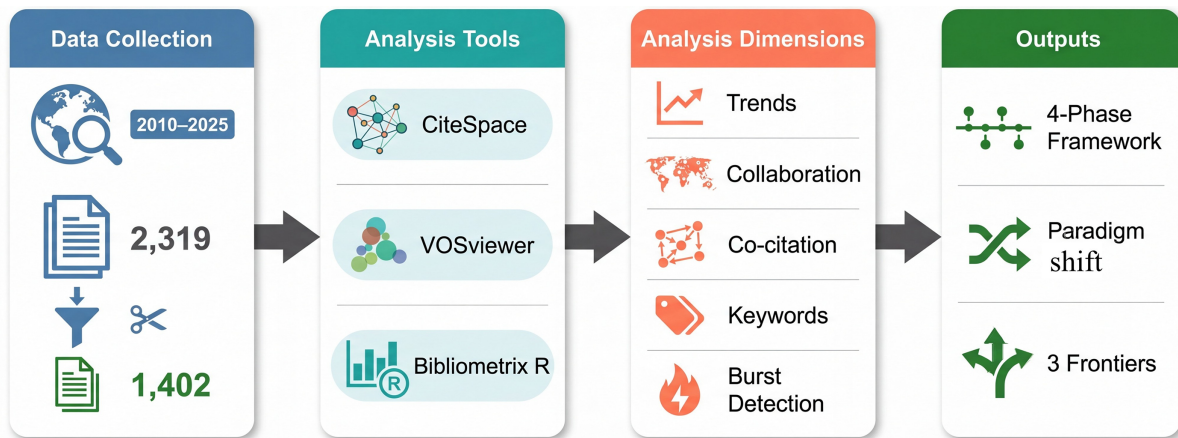


Figure 2: Research methodology framework.

4 Overview of the Overall Situation

This section presents the overall publication profile of DRL-based UAV path planning. It examines publication growth, source distribution, and research output to show whether the topic has developed into a stable research area. These results provide the background for the later analyses of collaboration, intellectual structure, and thematic evolution.

4.1 Trend Analysis

We processed the dataset using the Bibliometrix R package to extract annual publication counts, journal distributions, and cumulative volume since 2010.

Fig. 3 shows an overall upward trend in annual publications, with a visible increase around 2021. This pattern suggests that DRL-based UAV path planning received growing scholarly attention during this period. However, the observed increase should be interpreted as a bibliometric signal of rising publication activity rather than as direct evidence that the field has fully consolidated as a mature research frontier.

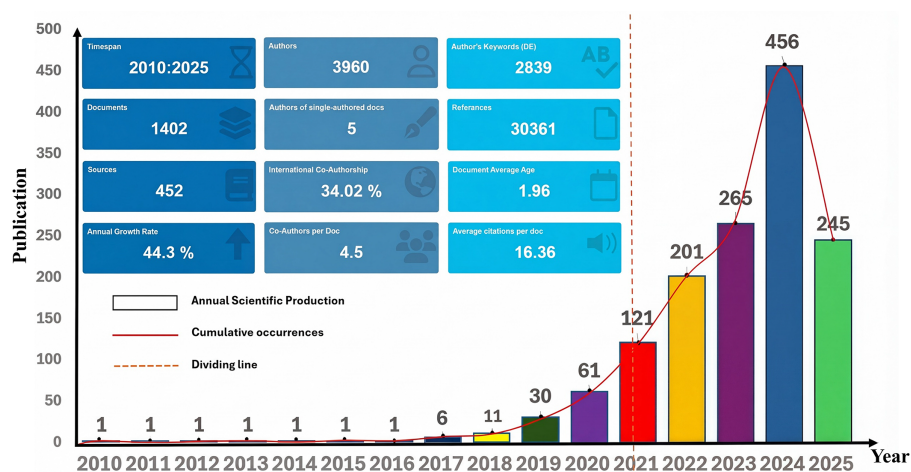


Figure 3: Annual publication volume and overall status overview.

The Bibliometrix overview yields several telling statistics: 1402 articles across 452 publication sources, contributed by 4532 authors, with a compound annual growth rate of 44.3%. International collaborations account for 34.02% of total output—a figure that underscores the inherently transnational character of this field.

Table 2 ranks the top 10 journals by publication volume (ISO 4 abbreviations). Collectively, these venues contributed 461 articles (32.8% of the corpus) and amassed 4385 citations. The two leading outlets—IEEE Internet of Things Journal (102 articles, 819 citations) and IEEE Transactions on Vehicular Technology (86 articles, 869 citations)—are IoT and vehicular-technology journals rather than traditional robotics or aerospace venues. This pattern suggests that UAV path planning research has shifted toward communication-oriented and networked-system communities, consistent with the co-citation pattern discussed in Section 6.1.

Table 2: Basic information on the top 10 journals by publication volume.

Journal	Number	JCR	H-Index	Citation Frequency
IEEE Internet of Things Journal	102	Q1	29	819
IEEE Transactions on Vehicular Technology	86	Q1	31	869
Drones	67	Q1	11	341
IEEE Access	54	Q2	21	595
IEEE Transactions on Wireless Communications	39	Q1	19	514
Sensors	27	Q2	9	380
IEEE Transactions on Mobile Computing	25	Q1	14	411
Journal of Electronics & Information Technology	22	–	3	55
IEEE Transactions on Intelligent Transportation Systems	20	Q1	9	299
Computer Networks	19	Q2	6	102

This shift may reflect several non-exclusive factors. One possible explanation is *problem-formulation convergence*: as UAV path planning increasingly incorporates communication quality of service (QoS), spectrum allocation, and edge computing as co-optimization objectives, the resulting mathematical formulations—including mixed-integer nonlinear programs, Lyapunov-drift-plus-penalty frameworks, and stochastic-geometry models—fit more naturally within the conventions of wireless communication journals than those of traditional robotics venues [23,24]. Another explanation is *reviewer-expertise alignment*: communication-oriented journals often have reviewer communities with stronger expertise in channel modeling, resource allocation, and information-theoretic analysis, which are increasingly relevant to UAV trajectory–communication co-design studies. A third possibility is *citation-network reinforcement*: once influential foundational papers established communication journals as major publication venues for UAV-related optimization problems, later authors may have continued to submit to the same outlets because of visibility, familiarity, and readership [23,25,26]. These interpretations should be treated as hypotheses suggested by the bibliometric pattern rather than as causal mechanisms established by the present analysis.

4.2 Analysis of Document Output

Fig. 4 disaggregates annual publication counts across the top five journals, countries/regions, institutions, and authors.

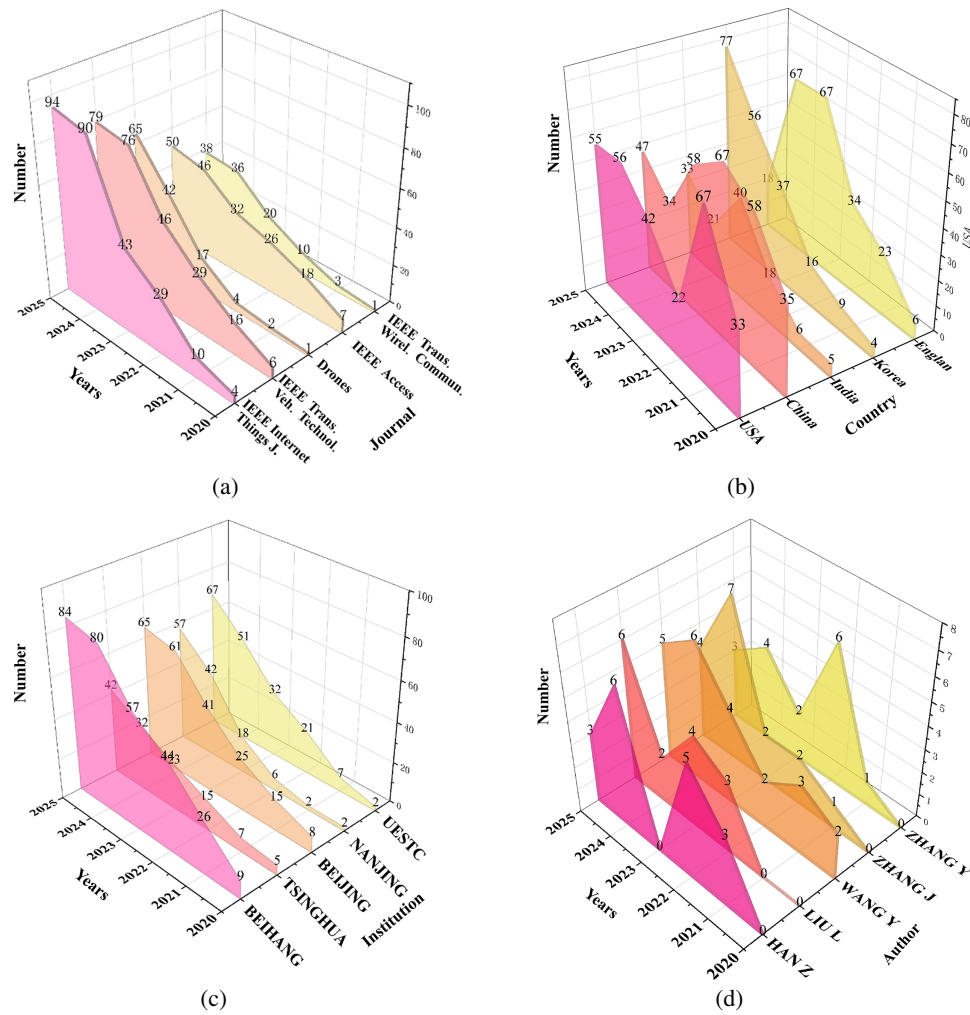


Figure 4: Publication comparison. (a) Journal; (b) Country/Region; (c) Institution; (d) Author.

The five leading journals—IEEE Internet of Things Journal, IEEE Transactions on Vehicular Technology, Drones, IEEE Access, and IEEE Transactions on Wireless Communications—maintain comparable output volumes and have each sustained a consistent upward trajectory since 2020, reflecting broad and durable scholarly momentum (Fig. 4a).

At the national level (Fig. 4b), China accounts for the largest number of publications in the dataset, followed by the USA, India, South Korea, and the United Kingdom. The annual count for China appears lower after 2023, but this observation should be interpreted cautiously. Several possible explanations may be considered, including indexing delay in recent years, a temporary redistribution of research attention toward adjacent AI topics, or partial methodological stabilization in commonly used DRL frameworks. These explanations remain hypotheses suggested by the observed publication pattern rather than conclusions established by the present data. Further verification would require longer-term tracking and complementary analysis of additional databases, including Chinese-language sources such as CNKI.

Among institutions (Fig. 4c), Beihang University leads in total volume, consistent with its established aerospace engineering strengths. Yet Tsinghua University and Peking University sustain robust output, indicating a competitive, multi-polar research landscape rather than single-institution dominance.

Author-level data (Fig. 4d) reveal that HAN Z achieved notably higher output in 2022, suggesting a concentrated research push during that year. The subsequent rise in output from other authors signals a broadening of the active researcher base—a hallmark of a field transitioning from niche specialization to mainstream prominence.

5 Collaborative Relationships and Disciplinary Research

This section examines how the field is organized socially and intellectually. It focuses on collaboration networks and disciplinary interactions to show how researchers, institutions, and subject areas connect and shape the research agenda. The analysis helps clarify the cooperative structure and interdisciplinary basis of DRL-based UAV path planning.

5.1 Analysis of National Contributions and Collaboration Networks

Fig. 5 maps the international collaboration network, where node size encodes national publication volume and line thickness captures collaborative intensity.

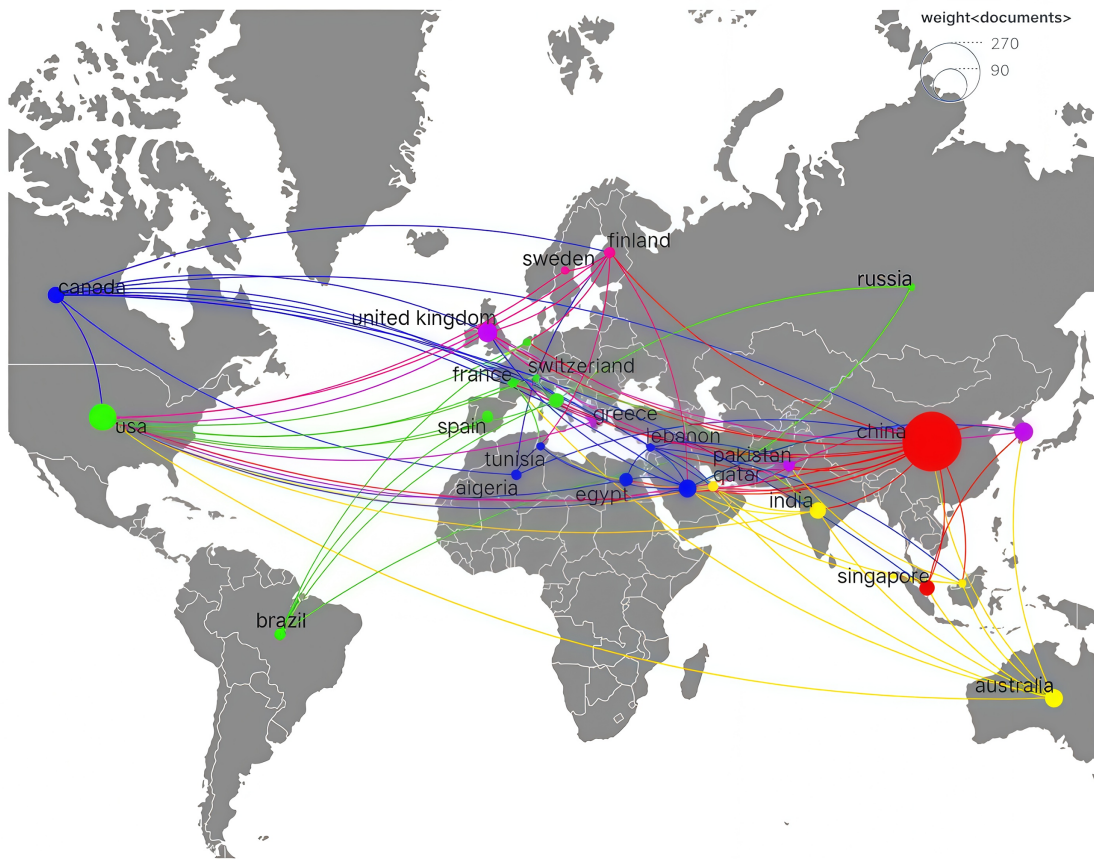


Figure 5: Country/region cooperation map.

China (≥ 270), the United States (≥ 60), and the United Kingdom (≥ 40) appear as major nodes in the network. The collaboration pattern covers multiple countries, but the publication distribution is uneven in this dataset. Because publication volume, collaboration frequency, and intellectual influence are not

equivalent indicators, the observed concentration should be treated as a descriptive network pattern rather than as evidence of balanced or unbalanced intellectual participation.

5.2 Analysis of Institutional Collaboration Networks

Fig. 6 maps institutional collaboration, revealing a densely interconnected core of Chinese universities surrounded by an international periphery.

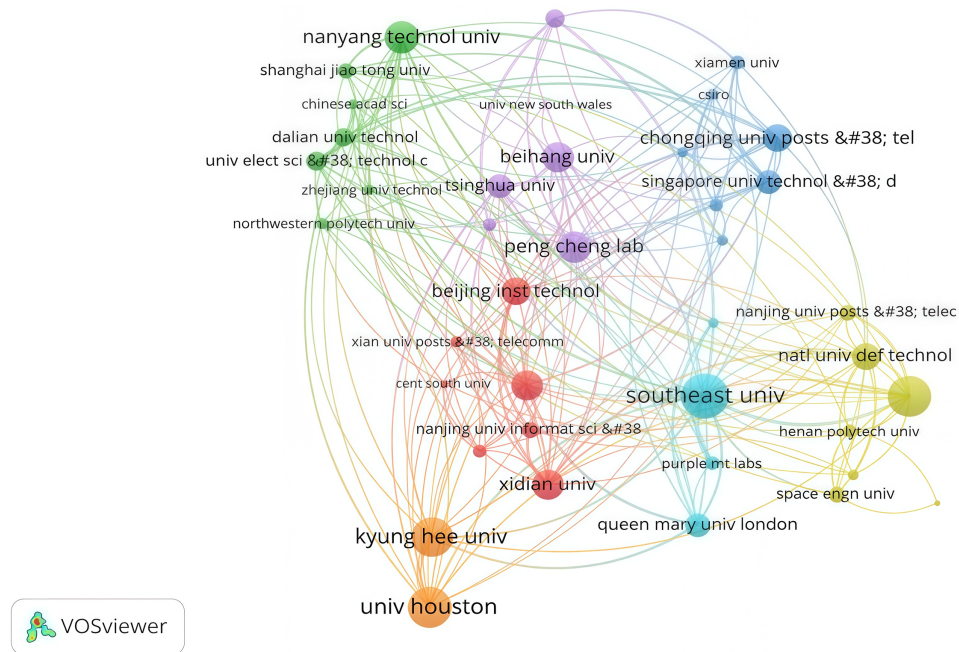


Figure 6: Institutional collaboration network. Note: The image is generated by VOSviewer. The lines connecting nodes indicate cooperative relationships between institutions, with node radius proportional to the strength of collaboration.

Three institutions—Tsinghua University, Shanghai Jiao Tong University, and the Chinese Academy of Sciences—occupy the network’s central positions, maintaining extensive domestic and international linkages concentrated in engineering and information technology. A distinct electronics-and-communications cluster coalesces around the University of Electronic Science and Technology of China and Xi’an University of Posts and Telecommunications, with thick interconnecting lines signifying deep industry-academia-research integration.

Geographical proximity amplifies collaboration: institutions in Nanjing, including the Purple Mountain Laboratory, form a tightly knit regional subnetwork. Internationally, Singapore University of Technology and Design and Queen Mary University of London serve as critical bridge nodes—fewer direct connections, but disproportionate influence in linking the Chinese core to Asia-Pacific, European, and American research communities.

Overall, domestic institutions form dense subnetworks through disciplinary complementarity and regional concentration, while international nodes extend the network outward. This pattern suggests that future progress may depend more on strengthening international links than on further densifying the domestic core.

5.3 Disciplinary Domain Analysis

Fig. 7 presents a dual-map overlay of journal citations. Citing journals (left) map current research fronts; cited journals (right) delineate the foundational knowledge base. Connecting lines trace interdisciplinary citation flows, with thickness encoding relationship strength.

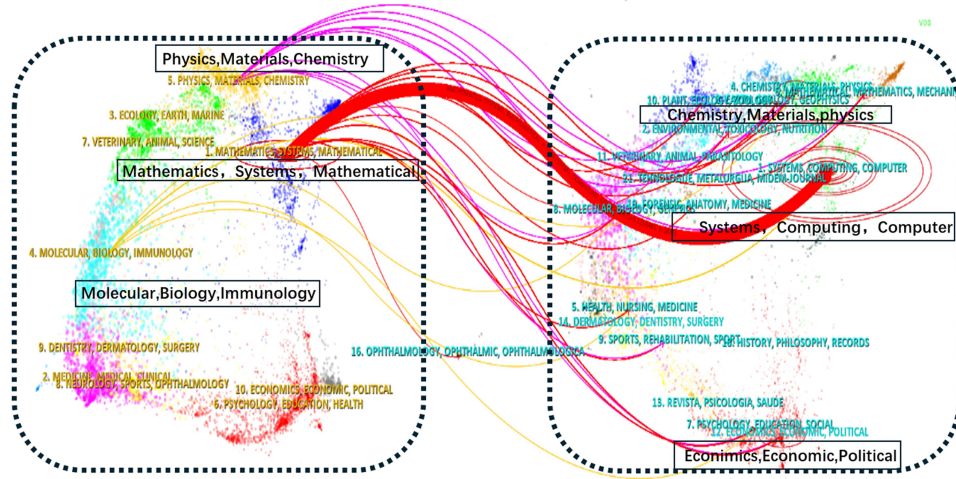


Figure 7: Dual-map overlay. Note: other disciplinary areas in this figure appear consistently as background elements, while the fields connected by the lines represent those covered in this study.

The dominant flow (red, $z = 5.13$, $f = 1811$) links current research fronts to foundational disciplines. Here f denotes raw citation frequency and z its z -score (i.e., the number of standard deviations above the mean), enabling cross-dataset comparison on a common scale.

The overlay indicates that Systems Science, Computer Science, and Mathematics form the main knowledge base of UAV path planning, while Materials Science and Economics contribute through areas such as lightweight structural design and cost-benefit optimization. Interdisciplinary citation flows also connect these core domains with applied fields such as precision agriculture and logistics optimization [27,28], indicating a continued integration of algorithmic methods with domain-specific applications.

CiteSpace's centrality ranking (Table 3) quantifies each discipline's bridging role within the research network. Three fields dominate: ENGINEERING (0.44), COMPUTER SCIENCE (0.34), and TRANSPORTATION SCIENCE (0.25). This triad confirms that UAV path planning sits at the intersection of hardware systems, algorithmic intelligence, and mobility logistics—rather than belonging to any single disciplinary silo. The steep centrality drop-off beyond the top three (AUTOMATION at 0.18, INFORMATION SYSTEMS at 0.14) further suggests that cross-disciplinary integration remains concentrated among a small number of bridging fields, leaving substantial room for emerging disciplines to strengthen their connective role.

Table 3: Centrality in various fields.

R	Subject Areas	Counts	Centrality
1	ENGINEERING	790	0.44
2	COMPUTER SCIENCE	363	0.34
3	TRANSPORTATION SCIENCE	394	0.25
4	AUTOMATION	116	0.18
5	INFORMATION SYSTEMS	115	0.14
6	REMOTE SENSING	95	0.10
7	TELECOMMUNICATIONS	48	0.10
8	OPERATIONS RESEARCH	25	0.09
9	CONTROL SYSTEMS	115	0.06
10	ENERGY	45	0.06

Note: Data sourced from CiteSpace.

6 Co-Citation Relationships

This section examines the intellectual structure of the field through co-citation relationships among references. Co-citation analysis helps identify foundational studies, thematic clusters, and knowledge connections that cannot be captured by publication counts alone. It therefore shows how the field's core knowledge base has formed and how its focus has changed over time.

6.1 Research on Cluster Co-Citation Relationships

Fig. 8, generated using CiteSpace, visualizes the co-citation network of references and their thematic clustering. Node size encodes citation frequency; connecting lines denote co-citation relationships. The network weaves together control theory, reinforcement learning, wireless communication, and IoT technology, confirming the field's pronounced interdisciplinary character [29–32].

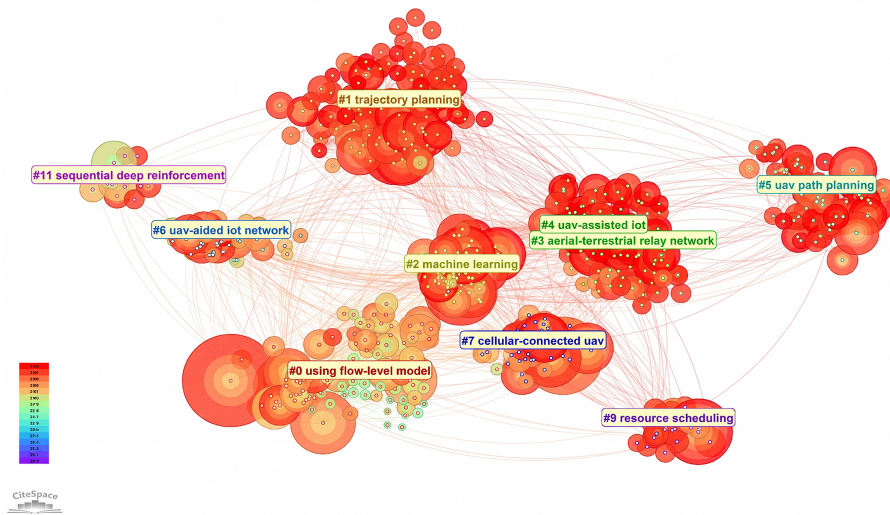


Figure 8: Co-citation network. Note: Node size is proportional to the number of times the node has been co-cited, and the intensity of the cluster colors represents the average vintage.

We organize the twelve identified clusters into three interpretive threads—core themes, methodological drivers, and application-driven expansion—to expose the structural logic governing this network.

Core themes. Clusters #1 (Trajectory Planning) and #5 (UAV Path Planning) form the core of the network. Their deep coloration indicates both foundational importance and sustained recent activity, showing that trajectory optimization remains central to the field. Cluster #9 (Resource Scheduling) is closely coupled with these two clusters: flight-path quality affects scheduling efficiency, while resource allocation constrains the search space for feasible paths. This relationship reflects a shift from “independent optimization” to “joint design”. A hierarchical DRL framework developed by researchers at King Abdullah University of Science & Technology illustrates this coupling by treating trajectory planning and power allocation as lower-level actions and content placement as the higher-level option space, with improved learning performance and lower energy consumption.

Methodological drivers. Clusters #11 (Sequential Deep Reinforcement), #2 (Machine Learning), and #6 (UAV-Aided IoT Network) represent major methodological drivers in the field. They address four recurring challenges: exploration in unknown environments, dynamic obstacle avoidance, joint communication–energy optimization, and multi-UAV collaborative planning. Their lighter coloration indicates that these research fronts are relatively recent and still developing [18,33].

Application-driven expansion. Five clusters—#7 (Cellular-Connected UAV), #6 (UAV-Aided IoT Network), #4 (UAV-Assisted IoT), #3 (Aerial-Terrestrial Relay Network), and #0 (Flow-Level Model)—collectively encode a paradigm shift. Path planning is no longer confined to geometric shortest-path computation; instead, flight trajectories are increasingly orchestrated by the imperative to maximize communication task performance [34]. Within this new paradigm, each cluster assigns UAVs a distinct operational role: mobile base stations (#4), data collectors (#6), aerial networking units (#7), and relay nodes (#3). Cluster #0 underpins all of these by providing the flow-level modeling tool that translates communication-centric objectives into tractable path optimization formulations [35]. Researchers from the California Institute of Technology and Seoul National University have demonstrated the practical consequence of this convergence, employing a quantum-inspired algorithm to jointly optimize relay deployment and path planning in extreme environments with a significant reduction in computational time [36].

Taken together, these three threads reveal that UAV path planning has evolved from isolated point-to-point navigation into a deeply integrated research domain spanning networking, edge computing, communication theory, and multi-agent coordination.

6.2 Co-Citation Burst Network Analysis

Fig. 9 visualizes the co-citation burst network, and Table 4 ranks the ten most influential publications by burst strength. Two points are notable. First, Yong Zeng contributed three of the top ten papers, all published in IEEE Transactions on Wireless Communications, showing the importance of this journal in the field’s early development. Second, the highest-burst work, by Mnih et al. [37], established a key algorithmic foundation for DRL by:

- (1) demonstrating end-to-end reinforcement learning from raw sensory inputs to control commands;
- (2) stabilizing deep-network–RL integration through experience replay and a separate target network; and
- (3) achieving superhuman performance across 49 Atari 2600 games with a single architecture, establishing the viability of general-purpose learning systems.

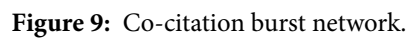
**Table 4:** Basic information on the 10 most influential papers.(Continued)

Table 4 (continued)

R	T	C	LS	N
7 [26]	Energy-Efficient UAV Communication with Trajectory Optimization	115	247	Yong Zeng
8 [24]	A Tutorial on UAVs for Wireless Networks: Applications, Challenges, and Open Problems	114	205	Mohammad Mozaffari
9 [42]	Energy-Efficient UAV Control for Effective and Fair Communication Coverage: A Deep Reinforcement Learning Approach	106	234	Chi Harold Liu
10 [43]	Multi-Agent Deep Reinforcement Learning Based Trajectory Planning for Multi-UAV Assisted Mobile Edge Computing	91	216	Liang Wang

Note: R = Rank; T = Title; C = Citation count; LS = Link strength; N = First author name.

7 Keywords

This section analyzes keywords because they directly reflect the topics, methods, and application priorities emphasized by the research community. Compared with citation-based indicators, keyword analysis more clearly shows how researchers frame their work and how thematic attention changes over time. It helps identify the main research streams, emerging topics, and the shift from isolated path-planning problems toward integrated system-level optimization.

7.1 Keyword Co-Occurrence Network Analysis

We constructed a co-occurrence network using VOSviewer with a threshold of $k = 100$, retaining 56 major keywords and filtering out 481 low-frequency terms. Node size encodes occurrence frequency; connecting lines capture co-occurrence relationships.

The 56 keywords cluster into six thematic groups (Fig. 10a–f):

- (a) Red Cluster: Optimization; (b) Light Blue Cluster: Energy Efficiency; (c) Green Cluster: Framework;
- (d) Dark Blue Cluster: Wireless Sensor Networks; (e) Purple Cluster: Model; (f) Orange Cluster: System

Together, these clusters outline the conceptual structure of the field and its main research streams.

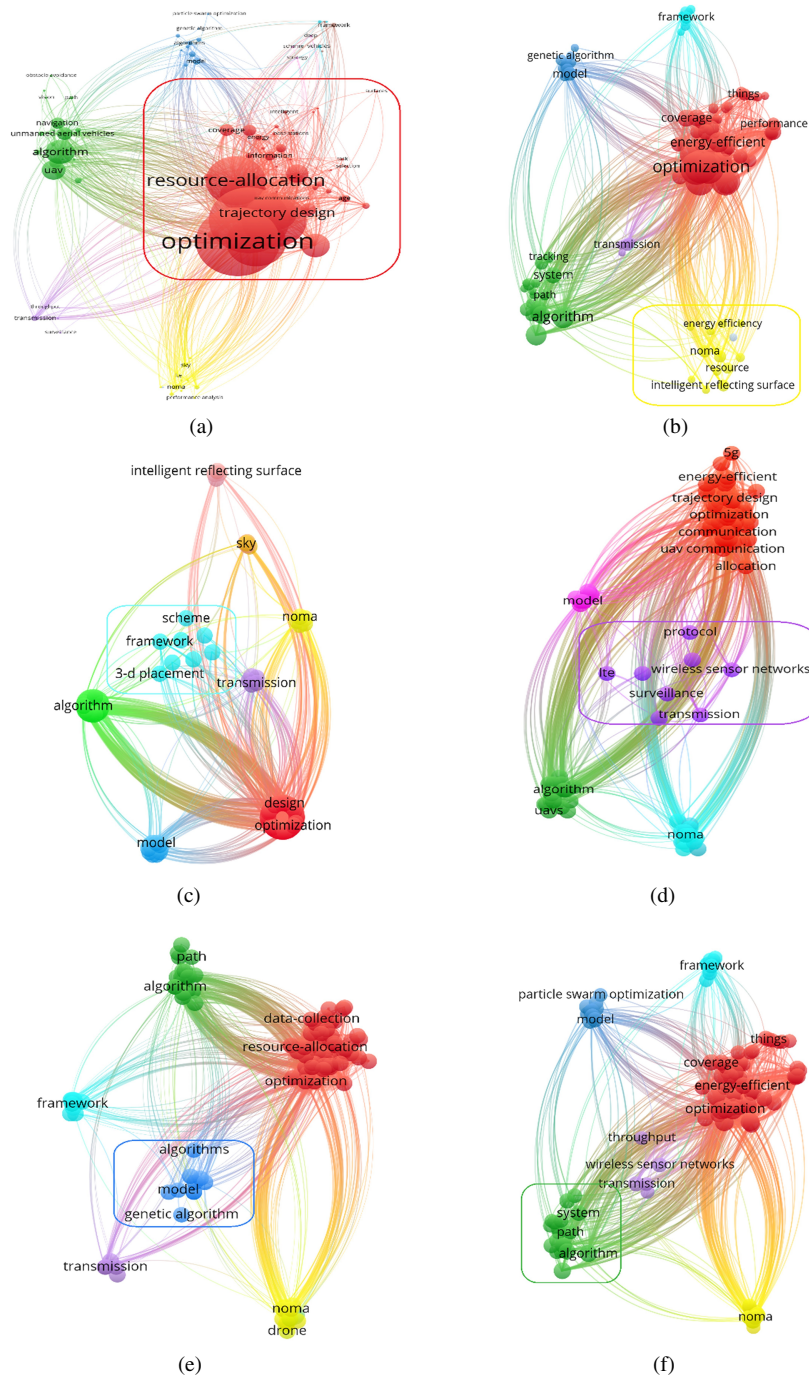


Figure 10: Keyword clustering network: (a) optimization; (b) energy efficiency; (c) framework; (d) wireless sensor networks; (e) model; (f) system.

Table 5 classifies the keywords by semantic meaning while retaining their occurrence frequencies and total link strengths, thus providing a clearer view of the thematic structure of UAV path-planning research. The results show that “optimization” remains the overarching objective of the field, whereas terms such as “autonomous-aerial-vehicles,” “unmanned aerial vehicles,” “trajectory,” “trajectory design,” and “path planning” define its core research object and technical focus. Together, these keywords indicate that UAV

path planning is fundamentally concerned with optimizing autonomous flight trajectories under complex operational constraints [44–46].

Table 5: Semantic classification of high-link-strength keywords.

Category	Semantic Focus	Keywords (Occurrences; Total Link Strength)
1	Overall objective and problem orientation	Optimization (355; ≥ 3000); design (212; 1500–1999); maximization (31; 250–499)
2	UAV platforms and autonomous flight carriers	Unmanned aerial vehicles (687; 2001–2499); autonomous-aerial-vehicles (257; 2500–2999); unmanned aerial vehicle (186; 1000–1499); drones (38; 250–499)
3	Trajectory planning and navigation control	Trajectory (214; 2500–2999); trajectory design (171; 1000–1499); path planning (203; 1000–1499); trajectory planning (77; 500–999); navigation (64; 250–499); altitude real-time systems (32; 250–499); vehicle dynamics (39; 250–499); coverage (41; 250–499)
4	Intelligent algorithms and learning-based decision-making	Deep-reinforcement-learning (342; 1500–1999); reinforcement-learning (306; 1500–1999); deep learning (94; 500–999); heuristic algorithms (67; 500–999); algorithm (78; 250–499); machine learning (34; 250–499); q-learning (38; 250–499); training (50; 250–499)
5	Communication, networking, and IoT-supported environments	Communication (174; 1500–1999); networks (159; 1000–1499); internet of things (90; 1000–1499); wireless communication (67; 500–999); mobile edge computing (67; 500–999); Internet (96; 500–999); wireless networks (45; 250–499); wireless sensor networks (40; 250–499); sensors (44; 250–499); servers (36; 250–499); intelligent reflecting surface (27; 250–499); noma (43; 250–499); edge computing (31; 250–499); three-dimensional displays (32; 250–499)
6	Resource allocation, energy management, and computation offloading	Resource-allocation (157; 1500–1999); resource management (118; 1000–1499); allocation (76; 500–999); resource allocation (55; 500–999); energy consumption (80; 500–999); energy-efficient (61; 500–999); energy efficiency (60; 500–999); throughput (47; 500–999); power allocation (29; 250–499); interference (25; 250–499); fair communication (22; 250–499); task offloading (34; 250–499); placement (30; 250–499); age (24; 250–499)
7	Tasks, data acquisition, and application-oriented support	Task analysis (99; 1000–1499); data collection (68; 500–999); deployment (60; 500–999); data-collection (36; 250–499); information (31; 250–499); systems (42; 250–499); array signal processing (26; 250–499)

The semantic categories further reveal the methodological and application-oriented expansion of the field. Keywords such as “deep-reinforcement-learning,” “reinforcement-learning,” “deep learning,” and “heuristic algorithms” show that intelligent and adaptive algorithms have become major approaches for solving path-planning problems. Meanwhile, terms including “communication,” “networks,” “internet of things,” “resource-allocation,” “energy consumption,” and “energy efficiency” suggest that UAV path planning has evolved from a standalone navigation task into a comprehensive optimization problem involving communication support, resource coordination, and energy-aware operation [44–46].

In addition, lower-frequency keywords, such as “fair communication,” “power allocation,” “task offloading,” and “intelligent reflecting surface,” reflect emerging interests in integrated communication-computation-navigation frameworks. Although these topics remain less prominent, they indicate a clear trend toward multi-dimensional system optimization. Therefore, the semantic classification in Table 5 not only identifies the major themes of UAV path-planning research but also highlights its ongoing transition from trajectory generation to coordinated, intelligent, and constraint-aware system design [42,46,47].

A temporal keyword analysis using VOSviewer (Fig. 11, Table 6) tracks thematic changes from 2022 to 2025. In the present dataset, 2022 contains a relatively broad set of emergent terms spanning communication, deployment, and energy efficiency; 2023 shows visible concentration around optimization, resource allocation, and trajectory design; and by 2024, only four keywords meet the emergence threshold—“robust” (16), “UAV” (83), “data collection” (24), and “MEC” (25). This apparent narrowing may suggest a movement toward a smaller set of visible research concerns, such as robustness, data-driven mission planning, and edge-computing integration. Nevertheless, because the 2024–2025 records remain incompletely indexed, this pattern should be treated as an observation from the current dataset rather than as definitive evidence of thematic convergence.

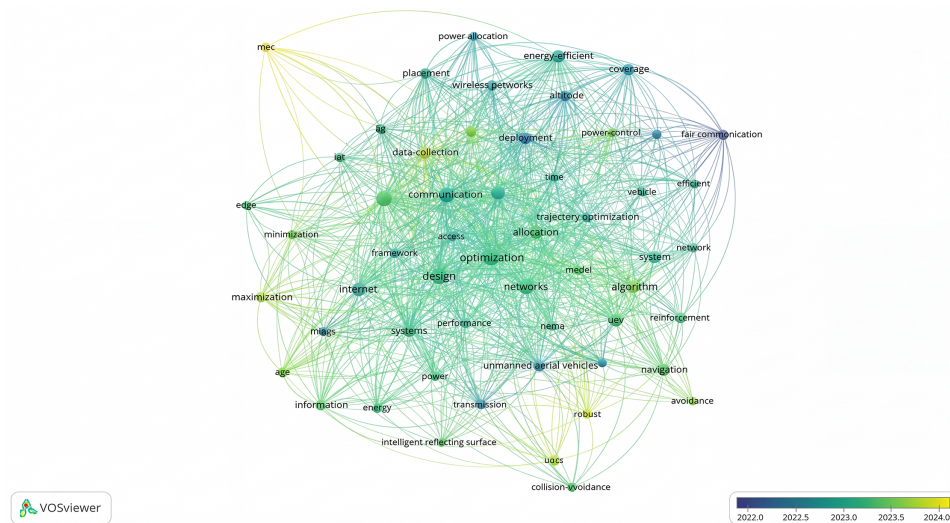


Figure 11: Keyword real-time network. Note: The overall timeframe for keyword appearances is restricted to after 2022. The more yellow the color, the later the appearance occurred.

Table 6: Keyword time distribution.

Years	Keywords (Occurrences)
2022	Fair communication(20)access(13)5g(23)age(24)algorithm(97) algorithms(14)allocation(76)altitude(32)avoidance(15) collision-avoidance(15)communication(170)coverage(38) data-collection(36)deployment(60)design(212)edge(18)efficient(16)energy(20) energy-efficient(59)framework(15)information(31) intelligent reflecting surface(14)internet(94) iot(18)management(21)maximization(31)mec(14)minimization(15) trajectory optimization(32)transmission(18)uav(83) unmanned aerial vehicles(42)vehicle(14)wireless networks(27)wireless(11) sensor networks(12)
2023	Noma(18)optimization(222)performance(17)placement(30)power(17) power allocation(16)power-control(20)reinforcement(19) resource(12)resource- allocation(157)search(12)sky(16)system(33)model(22)navigation(37)network(18) networks(158)systems(42)things(15) throughput maximization(12)time(15) tracking(12)trajectory design(105)
2024	Robust(16)data collection(24)mec(25)

7.2 Emergence Intensity

Fig. 12 charts keyword burst intensity from 2010 to 2025, where higher burst values indicate more rapid surges in research attention. Three evolutionary phases emerge:

Top 10 Keywords with the Strongest Citation Bursts

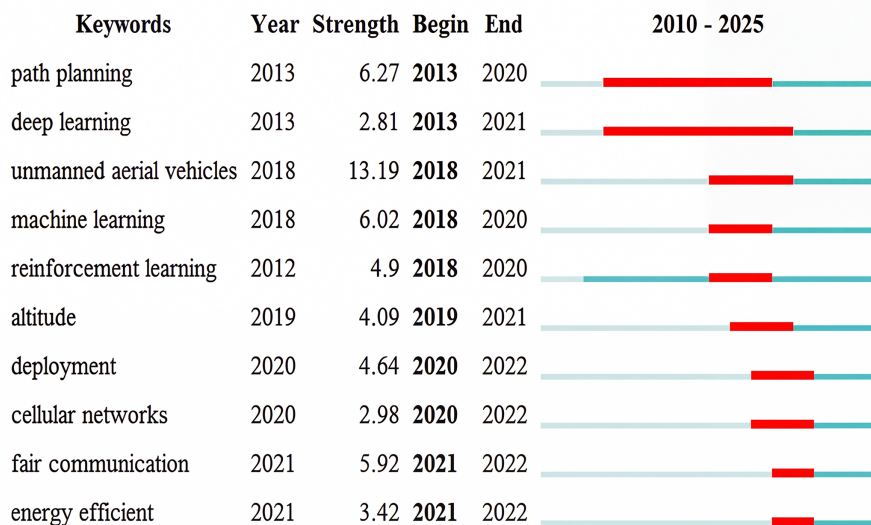


Figure 12: Keyword burst intensity. Note: The start and end points of the red line represent the beginning and end times of the keyword. The longer the red line, the longer the keyword was prominent.

Algorithmic Foundations (Early Phase). Foundational terms—“path planning” and “deep learning”—burst first, marking the initial penetration of AI into the domain [48].

Shift to Practical Application. “Unmanned aerial vehicles” exhibited the strongest burst intensity of any keyword, signaling a decisive pivot from algorithmic theory toward system-level implementation. The concurrent sustained prominence of “machine learning” and “reinforcement learning” confirms that intelligent algorithms had consolidated their role as mainstream solutions [49–51].

Specialization and System Integration. Keywords such as “deployment,” “cellular networks,” and “fair communication” emerged, indicating that the research paradigm expanded beyond pure algorithm optimization to encompass practical constraints including communication quality and energy consumption [52–54].

Fig. 13 visualizes keyword co-occurrence clusters on a timeline from 2012 to 2023, showing how the main research hotspots have evolved. The pattern moves from early foundational studies to closer integration with newer technologies. We highlight the main clusters below.

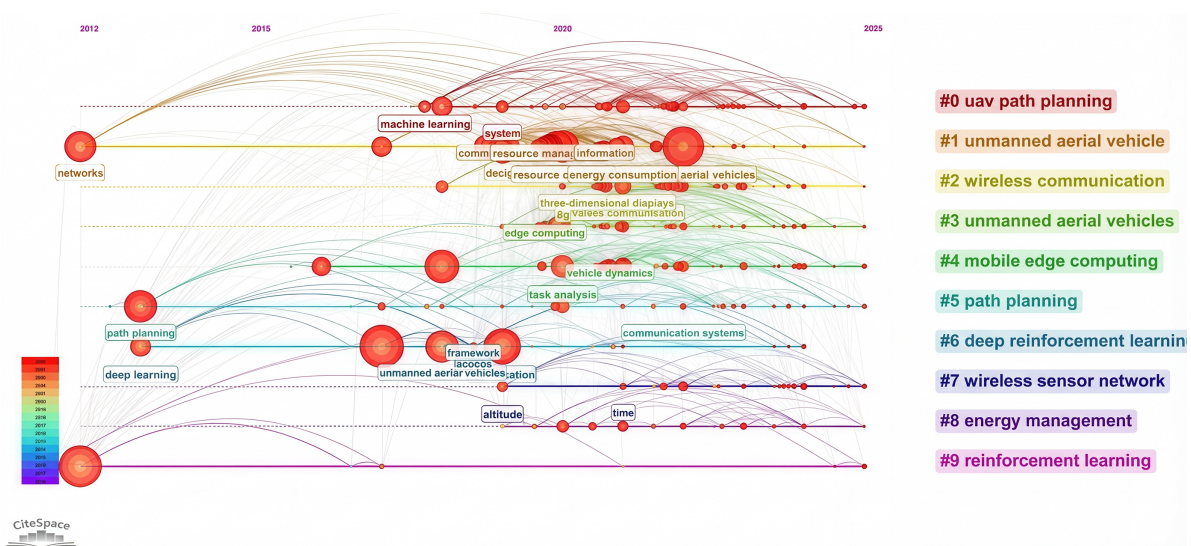


Figure 13: Subject area timeline.

Clusters #1 & #3: Unmanned Aerial Vehicle. These are the largest and most central clusters in the network, indicating that UAV design, system architecture, flight control, and dynamic vehicle modeling remain the common basis for later technical developments [55,56].

Clusters #2 & #7: Wireless Communication/Wireless Sensor Network. Communication and networking constitute the foundational infrastructure of UAV operations. Strong connections between “communication systems” and “communication management” nodes confirm that reliable, low-latency data links are prerequisites for coordinated multi-UAV control [57].

Cluster #4: Mobile Edge Computing. This cluster reflects a common strategy for addressing UAV computational and energy constraints. By offloading intensive tasks to nearby edge nodes, MEC can reduce processing latency and onboard energy consumption [58–60].

Cluster #5: Path Planning. The primary research objective of the entire field. This cluster synthesizes all enabling technologies—communication protocols, edge computing, deep learning—toward generating safer, more efficient, and more intelligent trajectories under complex operational constraints [61].

Clusters #6 & #9: Deep Reinforcement Learning. Their strong co-occurrence with Cluster #5 shows the growing importance of DRL in path planning for complex scenarios. The learn-by-interaction framework is especially useful in dynamic and uncertain conditions where explicit modeling is difficult [62,63].

Cluster #8: Energy Management. Energy management exhibits strong interdisciplinary linkages with MEC, path planning, and wireless communication, underscoring its role not as an isolated concern but as a cross-cutting constraint that permeates every dimension of UAV system design [58,64,65].

8 Evolutionary Framework: A Bibliometric Synthesis

This section synthesizes the preceding bibliometric findings into a developmental framework. Publication trends, collaboration patterns, co-citation structures, and keyword dynamics each capture only one dimension of the field, while an evolutionary framework helps connect them. The goal here is to clarify the main developmental stages of DRL-based UAV path planning, the factors behind each transition, and possible future directions.

Drawing on the publication trends (Section 4), collaborative networks (Section 5), co-citation topology (Section 6), and keyword dynamics (Section 7) presented above, this section synthesizes the bibliometric evidence into a four-phase evolutionary framework for DRL-based UAV path planning.

As illustrated in Fig. 14, this study partitions the evolution of DRL-based UAV path planning into four distinct phases, delineated by the maturity and depth of deep learning integration with planning tasks. This periodization should be understood as an evidence-based interpretive framework rather than a statistically unique segmentation. The proposed phase boundaries are anchored by milestone algorithmic breakthroughs. Specifically, the Deep Q-Network (DQN) [37] and Asynchronous Advantage Actor-Critic (A3C) [38] correspond to Phase 1; Deep Deterministic Policy Gradient (DDPG) [66], Proximal Policy Optimization (PPO) [41], Twin Delayed DDPG (TD3), and Soft Actor-Critic (SAC) [67] correspond to Phase 2; QMIX [68] and Multi-Agent Deep Deterministic Policy Gradient (MADDPG) [69] correspond to Phase 3; and Model-Agnostic Meta-Learning (MAML) [70] corresponds to Phase 4. These boundaries are broadly consistent with the observable bibliometric patterns presented in Sections 4–7, including publication-volume surges, keyword burst onsets, and co-citation cluster emergence.

The four-phase scheme is therefore intended as an interpretive synthesis supported by convergent bibliometric evidence, rather than as the only possible segmentation of the field.

Although the retrieval window spans 2010–2025, the four-phase periodization begins in 2015. The bibliometric patterns support this boundary: between 2010 and 2014, annual publication counts remained in single digits (fewer than 10 papers per year), keyword burst analysis detected no emergent terms, and no identifiable co-citation clusters formed. The field did not reach a critical mass of DRL-specific research until the publication of DQN [37] in 2015, which marked the first clear increase in both publication volume and thematic coherence. The 2010–2014 interval is therefore treated as a pre-DRL incubation period rather than as a separate evolutionary phase.

8.1 Phase 1: Foundational Exploration (2015–2016)

The earliest phase coupled single-agent reinforcement learning (RL) with shallow neural approximators for path planning in static or semi-static environments, laying the conceptual foundation for later deep reinforcement learning (DRL) systems [71,72]. Bibliometrically, this period contributed fewer than 30 cumulative publications, representing less than 2% of the total corpus—yet it established the conceptual seeds from which the entire field would grow. The publication of the Deep Q-Network (DQN) [37] in 2015 demonstrated that end-to-end deep reinforcement learning from raw sensory inputs was feasible, while

the Asynchronous Advantage Actor-Critic (A3C) algorithm [38] in 2016 introduced asynchronous multi-threaded training that substantially lowered the computational barrier to entry. The keyword burst analysis presented in Section 7.2 confirmed that “path planning” and “deep learning” emerged as the first burst terms in this period, marking the initial penetration of AI into the domain. Researchers paired conventional RL algorithms—Q-learning and State-Action-Reward-State-Action(SARSA)—with Multi-Layer Perceptrons (MLP), operating in simple two-dimensional grid worlds. Early systems relied on manually defined state features (e.g., obstacle distance and target azimuth), feeding them into MLPs to generate discrete control commands for rudimentary obstacle avoidance [71,72]. These approaches, however, suffered from three critical limitations: validation confined to small-scale laboratory grids, poor generalization due to hand-engineered features, and insufficient real-time performance stemming from the limited representational capacity of shallow architectures.

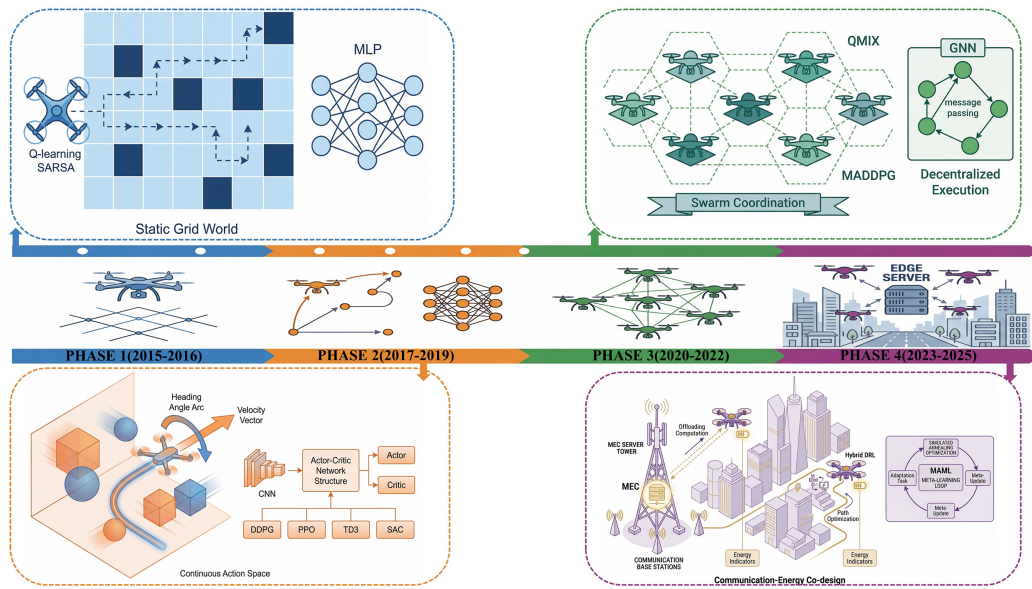


Figure 14: Evolutionary stages diagram. Note: This diagram serves only as a simplified illustration of each developmental stage. For specific details, please refer to the main text.

8.2 Phase 2: Continuous-Control Breakthrough (2017–2019)

This phase marked the broad introduction of deep reinforcement learning (DRL) into UAV path planning. The trend analysis in Section 4.1 shows that annual publications rose from approximately 20 in 2017 to over 80 by 2019, indicating that DRL had moved beyond a niche topic. The adoption of Convolutional Neural Networks (CNN) enabled end-to-end learning from raw sensor inputs and reduced reliance on manual feature engineering. Actor–Critic methods were especially important in this phase. Although Deep Deterministic Policy Gradient (DDPG) [66] was introduced in 2015, its influence on UAV path-planning research became more visible during 2017–2019 as the field moved toward continuous action spaces. Proximal Policy Optimization (PPO) [41] provided a stable on-policy baseline in 2017; Twin Delayed DDPG (TD3) addressed overestimation bias in 2018; and Soft Actor-Critic (SAC) [67] introduced maximum-entropy exploration in the same year. Together, these algorithms enabled continuous control commands such as velocity and heading angle, overcoming the discrete action-space limitation of earlier RL methods [72]. Experience replay and target networks also improved training stability in dynamic environments. This

phase coincides with the strongest keyword burst for “unmanned aerial vehicles” identified in [Section 7.2](#), indicating a shift from algorithm-oriented studies toward more practical system integration.

In multi-UAV coordination, the MADDPG algorithm enabled collaborative trajectory planning [69], supporting resilient “honeycomb-like” coverage networks that keep users within communication range of the nearest UAV. This framework further introduced dynamic computational offloading—intelligently distributing tasks to under-utilized drones during flight to balance processing latency against energy consumption [73].

8.3 Phase 3: Multi-Agent Collaborative Coordination (2020–2022)

This phase saw the rise of graph neural networks (GNN) in wide-area multi-UAV networking. As multi-UAV swarms became more common, GNN-enhanced reinforcement learning was increasingly used for swarm coordination and spatial relationship modeling [74,75], reflecting a move from single-agent optimization to swarm-level coordination. Value-decomposition methods such as QMIX [68] further advanced cooperative multi-agent learning by factorizing a joint action-value function into per-agent utilities under a monotonicity constraint, enabling decentralized execution with centralized training. The co-citation analysis in [Section 6.1](#) is consistent with this change: Clusters #11 (Sequential Deep Reinforcement) and #2 (Machine Learning) emerged during this period as important methodological drivers. Quantitatively, this phase accounts for much of the corpus: annual output exceeded 200 papers in 2021 and 350 in 2022, with a 44.3% compound annual growth rate. The keyword expansion in 2022—36 newly emergent terms spanning communication, deployment, energy efficiency, and Internet of Things (IoT) integration ([Table 6](#))—shows that this phase had the broadest thematic expansion in the dataset.

GNN represent drone swarms as dynamic graph structures—nodes as individual drones, edges as communication links—capturing evolving collaborative interactions as the swarm topology changes. The resulting fully distributed architecture allows each UAV to exchange information exclusively with neighboring nodes, eliminating reliance on a central controller and significantly improving adaptability in environments with unstable communication infrastructure [75].

Spatiotemporal graph neural networks (STGNN) further deepened this capability by concurrently capturing spatial dependencies and temporal state evolution through integrated convolutional operations, substantially improving path prediction accuracy in dynamic environments.

8.4 Phase 4: Complex-Scenario Integration (2023–2025)

This phase is characterized by closer integration between DRL, traditional optimization methods, and emerging technologies in large-scale and dynamic path-planning problems. *The 2024–2025 publication data were still incompletely indexed at the time of analysis, so the keyword and co-citation statistics for this sub-period should be interpreted as indicative rather than definitive.* The keyword time distribution ([Table 6](#)) supports this phase: emergent terms declined from 36 in 2022 to 22 in 2023 and to only 4 in 2024—“robust” (16), “UAV” (83), “data collection” (24), and “mobile edge computing (MEC)” (25). This pattern is consistent with a shift from broad exploratory growth toward a smaller set of focused problems, although incomplete indexing may also contribute. Co-citation Clusters #4 (UAV-Assisted IoT) and #7 (Cellular-Connected UAV) reached peak intensity during this period, indicating that communication-centric and edge-computing objectives have become central optimization targets. Two representative approaches characterize this phase. First, the combination of MAML [70] with DRL supports faster adaptation in previously unknown terrains [76]. Second, the SA-NNO-DRL hybrid framework combines Simulated Annealing with Nearest-Neighbor Optimized DRL and achieves strong path-planning performance in large-scale multi-UAV reconnaissance missions under tight time constraints [77].

9 Discussion

9.1 Main Findings

The bibliometric evidence presented in [Sections 4–7](#) indicates that DRL-based UAV path planning has developed from a relatively narrow technical topic into a distinct and rapidly expanding research domain. The publication surge after 2021, the concentration of influential outputs in IoT- and vehicular-technology-oriented journals, and the stable collaborative core centered on China collectively suggest that the field has moved beyond isolated algorithmic exploration and entered a stage of sustained interdisciplinary growth. More importantly, the combined co-citation and keyword evidence shows that the central research problem is no longer limited to geometric route generation. Instead, path planning is increasingly formulated as a system-level optimization problem in which communication quality, energy expenditure, resource allocation, and mission effectiveness are jointly considered. Within this transformation, hybrid DRL algorithms have become particularly important because they combine the adaptive decision-making capacity of DRL with complementary optimization, heuristic search, model-based control, or domain-specific planning mechanisms, thereby improving scalability and applicability in complex UAV operating environments.

From an evolutionary perspective, the field has followed a clear four-stage trajectory, progressing from shallow-RL exploration in simple environments to continuous-control DRL, then to multi-agent collaborative planning, and most recently to complex-scenario integration through hybrid DRL algorithms and hybrid learning-and-optimization frameworks. This progression implies that the scientific focus of the field has shifted from finding feasible flight paths to enabling autonomous aerial systems to make effective decisions under coupled sensing, communication, computation, and energy constraints. In this sense, the contribution of DRL lies not merely in improving path quality, but in redefining path planning as a core coordination layer within intelligent UAV systems. Hybrid DRL algorithms further reinforce this role by bridging data-driven policy learning with explicit optimization or prior knowledge, making them suitable for scenarios where pure DRL may face difficulties in convergence, safety assurance, constraint handling, or generalization. This application-oriented and cross-layer character also helps differentiate the present domain from broader UAV path planning research, because many of its most active problem settings are explicitly organized around mobile edge computing (MEC), Age of Information (AoI), data collection, smart-city services, and related integrated system objectives rather than flight geometry alone.

9.2 Current Application Hotspots

The current application hotspots of this field are strongly shaped by communication-oriented and mission-oriented requirements. In particular, representative hotspots increasingly cluster around MEC-enabled aerial services, AoI-aware information delivery, UAV data collection, and smart-city-oriented application scenarios, which further distinguishes this research stream from conventional path planning studies focused primarily on navigation feasibility or geometric optimization. One major hotspot is UAV-assisted IoT data collection and wireless sensing, where trajectory planning is tightly coupled with coverage, information freshness, throughput, and energy efficiency. In such studies, the flight path is no longer treated as an isolated navigation output, but as an integral component of data-acquisition and service-delivery strategies [[29,30,32–34](#)].

A second hotspot is cellular-connected UAV networking and aerial relay communication. Both the co-citation clusters and burst keywords indicate that communication reliability, fair transmission, power allocation, and connectivity maintenance have become central planning objectives rather than secondary constraints. This trend is reflected in representative studies on fair communication, connectivity-aware mobility, RIS-assisted trajectory control, and secure or relay-based transmission design [[42,45–47,53](#)].

A third hotspot is mobile edge computing and task offloading. In this line of research, trajectory planning is deeply integrated with server association, computation offloading, caching, and deployment decisions, making UAV mobility a carrier of communication and computing resources rather than a purely kinematic variable [43,58–60,73].

Another important hotspot is cooperative multi-UAV mission execution in complex environments, including target tracking, area coverage, reconnaissance, inspection, and dynamic obstacle avoidance. In these scenarios, DRL is valued primarily for its capacity to support online adaptation, decentralized coordination, and the balancing of local responses with global mission objectives [27,54,63]. Taken together, these hotspots confirm that the most active applications in the field are no longer defined by flight geometry alone, but by the deep integration of navigation with sensing, communication, computation, and multi-agent mission organization, often in service-oriented and smart-city-related operational settings.

9.3 Likely Future Hotspots and Emerging Application Directions

The keyword timeline provides a tentative basis for discussing possible future developments. After a broad thematic expansion in 2022, the emergent keyword set appears more concentrated in 2023 and 2024, with terms such as “robust”, “data collection”, and “mobile edge computing” remaining visible in the current dataset. Because recent records are still affected by incomplete indexing, this pattern should be understood as a provisional bibliometric observation rather than as firm evidence of field-wide convergence. On this basis, three possible future research directions can be cautiously outlined.

First, robust, safe, and generalizable decision-making is likely to receive more attention. As DRL methods move closer to real-world deployment, robustness under uncertainty, constraint satisfaction, and sim-to-real transfer may become more important than small gains on simulated reward metrics. Second, hierarchical and distributed collaboration is likely to expand, especially in large-scale multi-UAV systems. Future planning architectures may integrate global task allocation, intermediate resource coordination, and local closed-loop control, while drawing more heavily on federated learning, multi-agent reinforcement learning, and meta-learning to improve scalability and adaptability [33,46,76]. Third, communication-computation-energy co-design is likely to deepen. The field is already shifting from QoS-aware trajectory planning toward joint optimization of AoI, task offloading, edge intelligence, semantic communication, and reconfigurable wireless environments [32,45,57–60].

In terms of emerging applications, several directions stand out. One is infrastructure inspection, such as large-scale monitoring of wind farms, power corridors, and urban facilities, where path planning must jointly consider perception quality, endurance, and online adaptation [54]. A second is integrated air-ground-sea networking and cross-domain data collection, including SAGIN-related and underwater-IoT-related scenarios, which extend UAV path planning from isolated aerial autonomy to broader networked missions [32,34]. A third is mission-oriented confrontation and security applications, including reconnaissance, pursuit-evasion, and adversarial decision-making, all of which require planning policies with stronger adaptability, coordination, and robustness [55,77–81]. Overall, future work is likely to involve tighter integration among environment perception, communication infrastructure, onboard intelligence, and mission execution. This trend should still be interpreted cautiously, because the 2024–2025 literature remains incompletely indexed and the WoSCC-only dataset may underrepresent some very recent or non-English contributions.

9.4 Limitations

Several limitations should be acknowledged. First, the bibliometric corpus is inherently sensitive to query design. Although an expanded set of synonymous terms and a manual screening procedure were adopted, alternative keyword combinations or different database selections could still produce partially

different retrieval results. In particular, using TS as the search field may introduce some irrelevant records; despite efforts to remove such records during manual screening, inadvertent inclusions or omissions may still remain. Second, the title–abstract–full-text screening process necessarily involved researcher judgment, which introduces a degree of subjectivity, even though explicit exclusion criteria and the screening workflow shown in Fig. 1 were applied. Third, some analytical outcomes are affected by software settings and parameter choices in CiteSpace, VOSviewer, and Bibliometrix, including threshold selection, clustering granularity, and time slicing. Fourth, citation-based indicators are subject to citation-age bias, since older publications have had more time to accumulate citations and co-citations than more recent studies. Fifth, burst-detection results for keywords should be interpreted with caution, because some burst terms do not convey complete or self-contained semantic meaning when viewed in isolation and therefore require contextual interpretation in relation to the source documents and surrounding thematic structure. Sixth, self-citation and closely connected author- or journal-level citation practices may also influence local network prominence. Seventh, restricting the analysis to the WoSCC database may underrepresent relevant studies indexed exclusively in other databases, such as Scopus, IEEE Xplore, CNKI, or Wanfang. Taken together, these limitations do not invalidate the main structural patterns identified in this study, but they do suggest that the findings should be interpreted as robust at the structural level rather than as exact measurements fully independent of dataset construction or analytical settings.

10 Conclusion

This study conducted a bibliometric analysis of DRL-based UAV path planning using 1402 publications from the Web of Science Core Collection from 2010 to 2025. By combining CiteSpace, VOSviewer, and Bibliometrix, the study reconstructed the publication landscape, collaboration patterns, co-citation structure, keyword evolution, and developmental trajectory of the field. The results show that DRL-based UAV path planning has evolved from early algorithmic exploration into a rapidly expanding interdisciplinary research domain.

The main contribution of this study is to identify a four-phase evolutionary framework: foundational exploration, continuous-control breakthrough, multi-agent collaborative coordination, and complex-scenario integration. Across these phases, the research focus has gradually shifted from geometric path generation and local obstacle avoidance toward system-level decision-making that jointly considers communication, computation, energy consumption, resource allocation, and mission effectiveness. This shift indicates that UAV path planning is increasingly becoming a coordination layer within intelligent aerial systems rather than a standalone navigation problem.

Overall, the findings provide a structured and reproducible overview of how DRL-based UAV path planning has developed and where it is likely to move next. Future research is expected to place greater emphasis on robust and safe decision-making, scalable multi-UAV collaboration, communication–computation–energy co-design, and the integration of DRL with optimization, domain knowledge, and real-world deployment constraints.

Acknowledgement: Not applicable.

Funding Statement: This research was funded by the National Defense Science and Technology Innovation project, grant number ZZKY20253129.

Author Contributions: The authors confirm contribution to the paper as follows: Conceptualization, Tao Yang and Qiwu Wu; methodology, Qiwu Wu; software, Tao Yang; validation, Tao Tong, Tao Yang and Qiwu Wu; formal analysis, investigation, Yunchen Su; resources, Qiwu Wu; data curation, Qiwu Wu; writing—original draft preparation, Tao Yang;

writing—review and editing, Qiwu Wu; visualization, Tao Yang; supervision, Lingzhi Jiang; project administration, Qiwu Wu; funding acquisition, Qiwu Wu. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The raw data of WOSCC can be obtained through the clickable link: <https://github.com/yt32131234/WOSCC-DATA.git>.

Ethics Approval: Not applicable.

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

References

1. Ibraheem Mohammed Y, Hassan R, Hasan MK, Islam S, Saleh Abbas H, Asghar Khan M, et al. Revolutionizing FANETs with reinforcement learning: optimized data forwarding and real-time adaptability. *IEEE Open J Commun Soc.* 2025;6:4295–310.
2. Zhang M, Pang B, Yan C, Feroskhan M, Lv C. Real-time avoidance of obstacles and emergent geo-fences for urban air mobility using deep reinforcement learning. *IEEE Trans Intell Transp Syst.* 2025;26(11):20091–108.
3. Huang Y, Wang Y, Li Z, Zhang H, Zhang C. A hierarchical multi robot coverage strategy for large maps with reinforcement learning and dense segmented siamese network. *IEEE Robot Autom Lett.* 2025;10(1):444–51. doi:10.1109/lra.2024.3502067.
4. Li C, Dong W, He L, Cai M, Li Y. A hierarchical reinforcement learning method for intelligent decision-making in joint operations of sea–air unmanned systems. *Drones.* 2025;9(9):1–24. doi:10.3390/drones9090596.
5. Peng W, Wang D, Yin Y, Cheng T. Multi-agent deep reinforcement learning-based truck-drone collaborative routing with dynamic emergency response. *Transp Res Part E Logist Transp Rev.* 2025;195(4):1–29. doi:10.1016/j.tre.2025.103974.
6. Meng W, Zhang X, Zhou L, Guo H, Hu X. Advances in UAV path planning: a comprehensive review of methods, challenges, and future directions. *Drones.* 2025;9:376. doi:10.3390/drones9050376.
7. Kumar P, Pal K, Govil MC. Comprehensive review of path planning techniques for unmanned aerial vehicles (UAVs). *ACM Comput Surv.* 2025;58(3):1–38. doi:10.1145/3737280.
8. Dradoum A, Khelassi A, Lachehab F. Intelligent path planning algorithms for UAVs: classification, complexity analysis, hybrid ablation insights, and future directions. *Adv Mech Eng.* 2025;17:1–25. doi:10.1177/16878132251355020.
9. Arshid K, Krayani A, Marcenaro L, Gomez DM, Regazzoni C. Toward autonomous UAV swarm navigation: a review of trajectory design paradigms. *Sensors.* 2025;25:5877. doi:10.3390/s25185877.
10. Aggarwal S, Kumar N. Path planning techniques for unmanned aerial vehicles: a review, solutions, and challenges. *Comput Commun.* 2020;149:270–99. doi:10.1016/j.comcom.2019.10.014.
11. Husnain AU, Mokhtar N, Mohamed Shah N, Dahari M, Iwahashi M. A systematic literature review (SLR) on autonomous path planning of unmanned aerial vehicles. *Drones.* 2023;7(2):118. doi:10.3390/drones7020118.
12. Luo J, Tian Y, Wang Z. Research on unmanned aerial vehicle path planning. *Drones.* 2024;8(2):51. doi:10.3390/drones8020051.
13. Debnath D, Vanegas F, Sandino J, Hawary AF, Gonzalez F. A review of UAV path-planning algorithms and obstacle avoidance methods for remote sensing applications. *Remote Sens.* 2024;16:624.
14. Azar AT, Koubaa A, Mohamed NA, Ibrahim HA, Ibrahim ZF, Hameed IA, et al. Drone deep reinforcement learning: a review. *Electronics.* 2021;10(9):999. doi:10.3390/electronics10090999.
15. AlMahamid F, Grolinger K. Autonomous unmanned aerial vehicle navigation using reinforcement learning: a systematic review. *Eng Appl Artif Intell.* 2022;115(3):105321. doi:10.1016/j.engappai.2022.105321.
16. Puente-Castro A, Rivero D, Pazos A, Fernandez-Blanco E. A review of artificial intelligence applied to path planning in UAV swarms. *Neural Comput Appl.* 2022;34:153–70. doi:10.1007/s00521-021-06569-4.
17. Bai Y, Zhao H, Zhang X, Chen J, Yang L. Toward autonomous multi-UAV wireless network: a survey of reinforcement learning-based approaches. *IEEE Commun Surv Tutor.* 2023;25:3038–67. doi:10.1109/COMST.2023.3323344.

18. Amodu OA, Althumali H, Mohd Hanapi Z, Jarray C, Raja Mahmood RA, Adam MS, et al. A comprehensive survey of deep reinforcement learning in UAV-assisted IoT data collection. *Veh Commun.* 2025;55(2):100949. doi:10.1016/j.vehcom.2025.100949.
19. Borrego-Dominguez S, Jimenez L, Garcia-Guerrero EE, Esqueda-Elizondo JJ, Ortega-Corral C, Tamayo-Pérez UJ, et al. Reinforcement learning for drone path planning: a bibliometric analysis. In: *Complex systems and their applications*. Cham, Switzerland: Springer; 2025. p. 115–38.
20. Wu Q, Su Y, Tan W, Jiang L, Tong T. UAV path planning trends from 2000 to 2024: a bibliometric analysis and visualization. *Drones.* 2025;9:128. doi:10.3390/drones9020128.
21. Guzsvinecz T, Szűcs J. A bibliometric analysis of the trends in UAV research using the bibliometrix R-tool. *Appl Sci.* 2025;15(21):11305. doi:10.3390/app152111305.
22. Yang K, Meho LI. Citation analysis: a comparison of Google Scholar, Scopus, and Web of Science. *Proc Am Soc Inf Sci Technol.* 2006;43(1):1–15. doi:10.1002/meet.14504301185.
23. Zeng Y, Zhang R, Lim TJ. Wireless communications with unmanned aerial vehicles: opportunities and challenges. *IEEE Commun Mag.* 2016;54(5):36–42. doi:10.1109/mcom.2016.7470933.
24. Mozaffari M, Saad W, Bennis M, Nam YH, Debbah M. A tutorial on UAVs for wireless networks: applications, challenges, and open problems. *IEEE Commun Surv Tutor.* 2019;21:2334–60. doi:10.1109/COMST.2019.2902862.
25. Zeng Y, Xu J, Zhang R. Energy minimization for wirelcommunication with rotary-wing UAV. *IEEE Trans Wirel Commun.* 2019;18(4):2329–45. doi:10.1109/twc.2019.2902559.
26. Zeng Y, Zhang R. Energy-efficient UAV communication with trajectory optimization. *IEEE Trans Wirel Commun.* 2017;16(6):3747–60. doi:10.1109/twc.2017.2688328.
27. Huang J, Li H, Chen C, Liu Y, Zhang X. An improved deep reinforcement learning-based UAV area coverage algorithm for an unknown dynamic environment. *Appl Sci.* 2025;15:20.
28. Maw AA, Tyan M, Nguyen TA, Lee JW. iADA*-RL: anytime graph-based path planning with deep reinforcement learning for an autonomous UAV. *Appl Sci.* 2021;11(9):18. doi:10.3390/app11093948.
29. Talukdar N, Barman DC, Hazra A, Roy A, Mazumdar N. A dual map-based cooperative DDQN approach for trajectory planning in a UAV-enabled IoT network. *Phys Commun.* 2025;72(5):1–13. doi:10.1016/j.phycom.2025.102732.
30. Tan WX, Ding T, Liu L. Intelligent UAV deployment for energy-efficient IoT data collection. *IEEE Internet Things J.* 2025;12(17):34890–9. doi:10.1109/jiot.2025.3586685.
31. Thiyagarajan R, Kiran Kumar B, Singh AK, Karthigaikumar P. MAGCNN-based backcom multiple beamforming for 6G unmanned aerial vehicles in IoT networks. *IETE J Res.* 2025;71(10):3251–62. doi:10.1080/03772063.2025.2512050.
32. Zhang G, Wei X, Tan X, Han Z, Zhang G. AoI minimization based on deep reinforcement learning and matching game for IoT information collection in SAGIN. *IEEE Trans Commun.* 2025;73(8):5950–64. doi:10.1109/tcomm.2025.3525566.
33. Zhang X, Xing H, Shen Y, Xu J, Cui S. Age of information minimization in UAV-enabled IoT networks via federated reinforcement learning. *IEEE Trans Wirel Commun.* 2025;24:7923–39.
34. Fu XW, Kang SQ. Deep reinforcement learning-based collaborative data collection in UAV-assisted underwater IoT. *IEEE Sens J.* 2025;25(1):1611–26. doi:10.1109/jsen.2024.3493454.
35. Diallo EM, Chai R, Adam ABM, Bayessa GA, Liang C, Chen Q. Long-term energy consumption minimization based on UAV joint content fetching and trajectory design. *Sensors.* 2025;25(3):27. doi:10.3390/s25030898.
36. Zhou Q, Wu J, Zhu M, Zhou Y, Xiao F, Zhang Y. LLM-QL: a LLM-enhanced Q-learning approach for scheduling multiple parallel drones. *IEEE Trans Knowl Data Eng.* 2025;37:5393–406. doi:10.1109/TKDE.2025.3579386.
37. Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG, et al. Human-level control through deep reinforcement learning. *Nature.* 2015;518:529–33. doi:10.1038/nature14236.
38. Mnih V, Badia AP, Mirza M, Graves A, Lillicrap T, Harley T, et al. Asynchronous methods for deep reinforcement learning. In: *Proceedings of the 33rd International Conference on International Conference on Machine Learning*; 2016 Jun 19–24; New York, NY, USA. p. 1928–37.

39. Al-Hourani A, Kandeepan S, Lardner S. Optimal LAP altitude for maximum coverage. *IEEE Wirel Commun Lett.* 2014;3(6):569–72. doi:10.1109/lwc.2014.2342736.
40. Wu Q, Zeng Y, Zhang R. Joint trajectory and communication design for multi-UAV enabled wirel networks. *IEEE Trans Wirel Commun.* 2018;17(3):2109–21. doi:10.1109/twc.2017.2789293.
41. Schulman J, Wolski F, Dhariwal P, Radford A, Klimov O. Proximal policy optimization algorithms. *arXiv:1707.06347.* 2017.
42. Liu CH, Chen Z, Tang J, Xu J, Piao C. Energy-efficient UAV control for effective and fair communication coverage: a deep reinforcement learning approach. *IEEE J Sel Areas Commun.* 2018;36(9):2059–70. doi:10.1109/jsac.2018.2864373.
43. Wang L, Wang K, Pan C, Xu W, Aslam N, Hanzo L. Multi-agent deep reinforcement learning-based trajectory planning for multi-UAV assisted mobile edge computing. *IEEE Trans Cogn Commun Netw.* 2021;7(1):73–84. doi:10.1109/tccn.2020.3027695.
44. Lu YF, Xie FF, Ji TW. Human-in-the-loop reinforcement learning for dynamic soaring: a trajectory planning and control integrated system. *Eng Appl Artif Intell.* 2025;157:1–19.
45. Sun T, Yin S, Deng L, Richard Yu F. Reinforcement-learning-based trajectory design and phase-shift control in UAV-mounted-RIS communications. *IEEE Trans Mach Learn Commun Netw.* 2025;3(1):163–75. doi:10.1109/tmlcn.2024.3502576.
46. Tarekegn GB, Tesfaw BA, Juang RT, Saha D, Tarekegn RB, Lin HP, et al. Trajectory control and fair communications for multi-UAV networks: a federated multi-agent deep reinforcement learning approach. *IEEE Trans Wirel Commun.* 2025;24:7598–611.
47. Lei H, Meng D, Ran H, Park KH, Pan G, Alouini MS. Multi-UAV trajectory design for fair and secure communication. *IEEE Trans Cogn Commun Netw.* 2025;11(3):1966–80. doi:10.1109/tccn.2024.3487142.
48. Zhang C, Li X, He C, Li X, Lin D. Trajectory optimization for UAV-enabled relaying with reinforcement learning. *Digit Commun Netw.* 2025;11(1):200–9. doi:10.1016/j.dcan.2023.07.006.
49. Kamineni JN, Burela RG, Gupta A, Rao GVP, Balaji B. Data-driven mechanical and aerospace systems: transforming with machine learning algorithms and nascent paradigms. *Multiscale Multidiscip Model Exp Des.* 2025;8(9):1–24. doi:10.1007/s41939-025-00966-1.
50. Zhang J, Xian Y, Zhu X, Deng H. A hybrid deep learning model for UAV path planning in dynamic environments. *IEEE Access.* 2025;13(10):67459–75. doi:10.1109/access.2025.3557394.
51. Xiao B, Wan L, Han X, Xi Z, Ding C, Li Q. Addressing local minima in path planning for drones with reinforcement learning-based vortex artificial potential fields. *Machines.* 2025;13(7):27. doi:10.3390/machines13070600.
52. Afifi G, Gadallah Y. Autonomous real-time smoothness control for reliable DDQN-based UAV navigation using cellular networks. *IEEE Access.* 2025;13:22011–28. doi:10.1109/access.2025.3531931.
53. Yang WJ, Li B. Connectivity-aware UAV mobility in cellular networks: DRL path planning and predictive handover. *Ad Hoc Netw.* 2025;179:1–14.
54. Zhang XY, Yu H, Zheng X, Wang H, Mu C, Guo P. Hybrid deep reinforcement learning for UAV inspection in large-scale wind farms: deployment and routing optimization. *IEEE Trans Ind Inform.* 2025;21(11):9011–21. doi:10.1109/tii.2025.3593966.
55. Shi G, Cao Y, Wang D, Yang Q, Zhang J, Shi Z. Towards intelligent unmanned adversarial games: a reinforcement learning framework with the PHP-ROW method. *Drones.* 2025;9:24.
56. Wang ZP, Ng SX, El-Hajjar M. A 3D spatial information compression based deep reinforcement learning technique for UAV path planning in cluttered environments. *IEEE Open J Veh Technol.* 2025;6:647–61. doi:10.1109/ojvt.2025.3540174.
57. Yang YB, Tan TH, Liu L. Optimization research on UAV semantic communication system based on SVD-MADRL. *Drone Syst Appl.* 2025;13:1–18.
58. Wang Z, Zhang W, Yang D, Wu F, Xu Y, Xiao L. Trajectory design of cellular-connected UAV patrol and mobile edge computing system: a deep reinforcement learning approach. *Comput Netw.* 2025;268(12):1–11. doi:10.1016/j.comnet.2025.111384.

59. Zakaryia SA, Mead MA, Nabil T, Hussein MK. Task offloading for multi-UAV asset edge computing with deep reinforcement learning. *Clust Comput.* 2025;28(7):1–17. doi:10.1007/s10586-025-05382-1.
60. Zhang J, Zhang G, Wang X, Zhao X, Yuan P, Jin H. UAV-assisted task offloading in edge computing. *IEEE Internet Things J.* 2025;12(5):5559–74. doi:10.1109/jiot.2024.3488210.
61. Li WK, Xiong Y, Xiong Q. Reinforcement learning-guided particle swarm optimization for multi-objective unmanned aerial vehicle path planning. *Symmetry.* 2025;17(8):21. doi:10.3390/sym17081292.
62. Zhu Y, Tan Y, Chen Y, Chen L, Lee KY. UAV path planning based on random obstacle training and linear soft update of DRL in dense urban environment. *Energies.* 2024;17:19.
63. Rocha LGS, Caldas KAQ, Terra MH, Ramos F, Vivaldini KCT. Dynamic Q-planning for online UAV path planning in unknown and complex environments. *Int J Intell Robot Appl.* 2025;9(4):1654–74. doi:10.1007/s41315-025-00457-z.
64. Xu Y, An J, Zhou C, Xu H, Han Z. Dynamic energy management for UAV-enabled VR systems: a tile-based collaboration approach. *IEEE Trans Veh Technol.* 2024;73(11):17668–83. doi:10.1109/tvt.2024.3430311.
65. Tang Q, Wen S, He S, Yang K. Multi-UAV-assisted offloading for joint optimization of energy consumption and latency in mobile edge computing. *IEEE Syst J.* 2024;18(2):1414–25. doi:10.1109/jsyst.2024.3395845.
66. Lillicrap TP, Hunt JJ, Pritzel A, Heess N, Erez T, Tassa Y, et al. Continuous control with deep reinforcement learning. *arXiv:1509.02971.* 2015.
67. Haarnoja T, Zhou A, Abbeel P, Levine S. Soft actor-critic: off-policy maximum entropy deep reinforcement learning with a stochastic actor. In: *Proceedings of the 35th International Conference on Machine Learning (ICML); 2018 Jul 10–15; Stockholm, Sweden.* p. 1861–70.
68. Rashid T, Samvelyan M, Schroeder de Witt C, Farquhar G, Foerster J, Whiteson S. QMIX: monotonic value function factorisation for deep multi-agent reinforcement learning. In: *Proceedings of the 35th International Conference on Machine Learning (ICML); 2018 Jul 10–15; Stockholm, Sweden.* p. 4295–304.
69. Lowe R, Wu Y, Tamar A, Harb J, Abbeel P, Mordatch I. Multi-agent actor-critic for mixed cooperative-competitive environments. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems; 2017 Dec 4–9; Long Beach, CA, USA.* p. 6382–93.
70. Finn C, Abbeel P, Levine S. Model-agnostic meta-learning for fast adaptation of deep networks. In: *Proceedings of the 34th International Conference on Machine Learning; 2017 Aug 6–11; Sydney, Australia.* p. 1126–35.
71. Lin LJ. Reinforcement learning for robots using neural networks [dissertation]. Pittsburgh, PA, USA: Carnegie Mellon University; 1992.
72. Sutton RS, Barto AG. Reinforcement learning: an introduction. Cambridge, MA, USA: MIT Press; 1998.
73. Lin N, Han X, Hawbani A, Sun Y, Guan Y, Zhao L. Deep reinforcement learning-based dual-timescale service caching and computation offloading for multi-UAV assisted MEC systems. *IEEE Trans Netw Serv Manag.* 2025;22(1):605–17. doi:10.1109/tnsm.2024.3468312.
74. Jiang J, Dun C, Huang T, Lu Z. Graph convolutional reinforcement learning. *arXiv:1810.09202.* 2018.
75. Du Y, Qi N, Li X, Xiao M, Boulogeorgos AAA, Tsiftsis TA, et al. Distributed Multi-UAV trajectory planning for downlink transmission: a GNN-enhanced DRL approach. *IEEE Wirel Commun Lett.* 2024;13(12):3578–82. doi:10.1109/lwc.2024.3480152.
76. Archana T, Aravind T, Malini AH, Kalaivani CT. Energy-aware adaptive obstacle avoidance based on meta-reinforcement learning with segmentation for UAV trajectory planning. *Int J Aeronaut Space Sci.* 2025;27:598–618.
77. Fan M, Liu H, Wu G, Gunawan A, Sartoretti G. Multi-UAV reconnaissance mission planning via deep reinforcement learning with simulated annealing. *Swarm Evol Comput.* 2025;93:1–15.
78. Zong KX, Zhang J. Unmanned aerial vehicle for multi-target tracking using pursuit-evasion game model and deep reinforcement learning. *Dyn Games Appl.* 2025;16(2):722–51. doi:10.1007/s13235-025-00645-3.
79. Theile M, Bayerlein H, Nai R, Gesbert D, Caccamo M. UAV path planning using global and local map information with deep reinforcement learning. In: *Proceedings of the 2021 20th International Conference on Advanced Robotics (ICAR); 2021 Dec 6–10; Ljubljana, Slovenia.* p. 539–46.

80. Lyu C, Li M, Ou J. UAV hybrid path planning based on hierarchical deep reinforcement learning. *J Beijing Univ Aeronaut Astronaut*. 2025;51:3451–9. (In Chinese).
81. Kutpanova ZA, Kadhim MR, Zheng X, Zhakiyev N. Multi-UAV path planning for multiple emergency payloads delivery in natural disaster scenarios. *J Electron Sci Technol*. 2025;23(2):100303. doi:10.1016/j.jnlest.2025.100303.