Enhancing Reliability in Multimodal UAV Communication Based on Opportunistic Task Space

Chao Ren, Chao Gong, Difei Cao, Yingqi Li, Haijun Zhang, *Fellow, IEEE*, and Arumugam Nallanathan, *Fellow, IEEE*

***Abstract*—An opportunistic task space-based model is proposed to enhance reliability in the execution of unmanned aerial vehicle (UAV) tasks. The task information consists of predefined basic information units, and multimodal communication is realized by sharing and transferring modality information against the interference from task-irrelevant information. The multimodal task information alignment and task-oriented multimodal non-intersecting mapping are also used to align and enhance the reliability of task-specific information from different modalities. Numerical results are provided to demonstrate the effectiveness of task space searching and its impact on task reliability.**

***Index Terms*—UAV network, multimodal communication, task execution, multimodal tasks.**

I. Introduction

UNMANNED aerial vehicles (UAVs) are pivotal in reconnaissance, monitoring, disaster rescue, photography, electronic countermeasures, and route planning, enabling them to perform tasks in complex environments effectively. UAVs acquire information and receive signals through communication and sensation during their active motion, thereby significantly enhancing information acquisition [1]. The reliability of multimodal information by UAVs, such as wireless signals, video, images, and sound, plays a crucial role in determining the success of their tasks. This reliability is essential for precise UAV control, secure communication, and target tracking, particularly in real-time conditions. However, during task execution, UAVs may encounter interference from other UAVs within a specific range. Moreover, complex wireless environments further complicate task execution, which decreases the reliability of information reception from each modality. Therefore, it is imperative to find a solution that improves the reliability of multimodal information reception in UAVs while considering environmental resistance and interference from other UAVs.

Multimodal communication involves the simultaneous utilization of diverse transmission modalities to transmit information [2], [3]. UAV task execution also experiences remarkable growth across numerous fields, promising improved task reliability [4], [5]. Leveraging multimodal communication can effectively enhance communication capabilities [6]. However, there is a lack of applications that integrate multimodal communication with collaborative UAV task execution to enhance reliability [7]. Additionally, foundational approaches to UAV reliability lack sufficient multimodal communication methods. The key advantage of employing multimodal communication is its ability to enhance the accuracy, depth, and potential of multiple information transmissions, thereby ensuring robust task information acquisition for UAVs [8].

This letter proposes an opportunistic task space-based method to enhance the reliability of multimodal UAV communication and task execution. The method achieves reliable task information by aligning and interacting multimodal information within the UAVs. The contributions of this letter are: (1) **Opportunistic Task with Basic Task Information** - orthogonal and common task information concepts are utilized when processing diverse raw task data to investigate the relationship between task success rates and the reliability of multimodal data reception; (2) **Multimodal Task Space** - the task space model is proposed to enhance the accuracy of UAV task execution by defining specific task coordinates, while the task deterministic space model reduces dimensionality by eliminating common information.

II. Multimodal Opportunistic Task Space

In Fig. 1, a UAV U_k with position coordinates $(d_x,\ d_y,\ d_z)$ and velocity vectors $(V_x,\ V_y,\ V_z)$ is on a task *Ta*. Other UAV U_j $(U_j \in \{U_\alpha, U_\beta, \ldots, U_\delta\})$'s position coordinate $\overrightarrow{d}_j$ is $(\vec{d}_{j_x}, \vec{d}_{j_y}, \vec{d}_{j_z})$. In advanced information age, UAVs can be precisely computer-controlled to execute a task with all the predefined basic information units $(I_{Ta} = \{i_1, i_2, \ldots, i_{N_{Ta}}\})$. Therefore, the success of task execution hinges on obtaining reliable multimodal information I_{M_*}, encompassing various sources such as communication transmissions (communication modality M_{com}), sensing (sensation modality M_{sen}), and motion functions (motion modality M_{mot}), including surveillance, radar, position, and velocity.

Manuscript received 18 September 2023; revised 13 October 2023; accepted 17 October 2023. Date of publication 19 October 2023; date of current version 9 February 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 62201034, Grant U22B2003, and Grant 62341103; in part by the Beijing Municipal Natural Science Foundation under Grant L212004-03; in part by the Beijing Natural Science Foundation under Grant L212004; and in part by the Xiaomi Fund of Young Scholar. The associate editor coordinating the review of this article and approving it for publication was K. Yang. *(Corresponding author: Haijun Zhang.)*

Chao Ren, Chao Gong, Difei Cao, Yingqi Li, and Haijun Zhang are with the Beijing Engineering and Technology Research Center for Convergence Networks and Ubiquitous Services, University of Science and Technology Beijing, Beijing 100083, China (e-mail: chaoren@ustb.edu.cn; 42027152@xs.ustb.edu.cn; caodifei@hotmail.com; erin_qi@foxmail.com; haijunzhang@ieee.org).

Arumugam Nallanathan is with the Communication Systems Research Group, Queen Mary University of London, E1 4FS London, U.K. (e-mail: a.nallanathan@qmul.ac.uk).

Digital Object Identifier 10.1109/LWC.2023.3326130

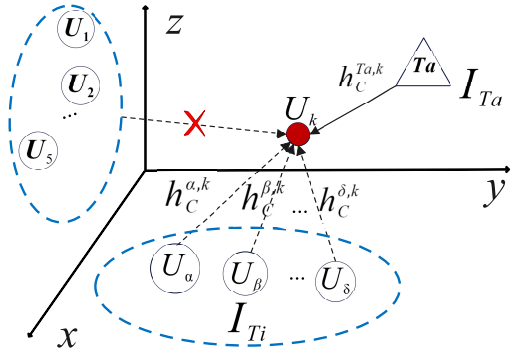

Fig. 1. Basic scenario for UAV receiving task information. The solid and the dashed arrows represent the task signal and interference signal, respectively.

In this way, the unreliability is due to the negative effect of multimodal reception, which is unified as task-irrelevant information I_{Ti} affected by environmental noise (denoted as N), interference from other UAVs j (each with transmitting power $P_I(j)$) and background sensed information. We assume that all the received information $H(I)$ is composed of task-specific information I_{Ta} and task-irrelevant information I_{Ti}, i.e., as $H(I) = I_{Ta} \cup I_{Ti}$, $(I_{Ta} \cap I_{Ti} = \emptyset)$.

For multimodal reception, taking M_{com} as an example, the task-specific signal is from source S_{Ta} with transmitting power P_{Ta} undergoes channel $h_C^{Ta,K}$, while task-irrelevant information is from other UAVs and noise within the influence range impacts UAV k through channel $h_C^{\alpha,k}$, $h_C^{\beta,k}$, $h_C^{\chi,k}$, ... $h_C^{\delta,k}$. Due to path fading, UAVs that are outside the influence range, such as U_1, U_2, U_3, U_4, and U_5, can not provide task-irrelevant information or task-specific information to U_k.

To enhance UAV task execution reliability, we introduce the concepts of opportunistic task, multimodal task information space to characterize the mutual influence among multimodal task-specific information.

A. Opportunistic Task With Basic Task Information

In a task *Ta*, if all the required correct task information (denoted as I_{Ta} and composed of multiple basic information units $i_1, i_2, \ldots, i_{N_{Ta}}$) is covered within a specified time frame, the task success probability (TSP) is 1; if some of the task information is mission/incorrect due to unreliable multimodal reception, the TSP is less than 1. We refer to *Ta* as an **opportunistic task**.

However, the raw information from multimodal reception is highly heterogeneous and does not directly provide the required units i_k directly. Therefore, we propose a task information extraction and multimodal task information alignment process to preprocess this information. It is also essential to achieve complementarity in the multimodal information by adhering to a unified format to mitigate heterogeneity. Hence, the alignment of information in multimodal tasks serves a dual purpose: enhancing task efficiency and standardization, while also improving the reliability of task execution at the informational level (see Section III-C).

- *Basic task information.* We assume that the task's basic information is fixed, standardized, and independent of multimodal reception.

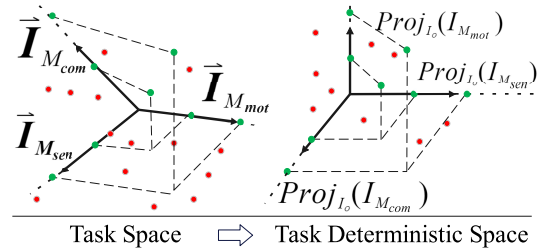

Fig. 2. Multimodal task information space. The green and red points represent the certain task information and the other task information, respectively.

- *Task information extraction*. For the original information obtained from multimodal reception, a deep learning network is utilized to extract semantic elements such as time, space, entity, behavior, event, and intention.
- *Multimodal task information alignment*. The resulting metadata is then mapped to the basic units i_k of the task, creating a unified representation across multiple modalities.

The basic information units of I_{Ta}, denoted as $i_1, i_2, \ldots, i_{N_{Ta}}$, can belong to one or more modalities. For instance, an information unit $i_k \in I_{M_{com}} \cap I_{M_{sen}}$ can exist in both the communication modality $(I_{M_{com}})$ and the sensation modality $(I_{M_{sen}})$, while another information unit $i_j \in I_{M_{mot}}$ $(i_j \notin I_{M_{com}} \cap I_{M_{sen}})$ may only exist in the motion modality.

Considering set properties, an information unit i_l belongs to I_O, which is defined as $I_O = I_{Ta} \setminus \bigcap(\forall I_{M_*})$, where M_* represents any two modalities. Then, i_l can only be obtained from one modality, and we refer to the set I_O consisting of all i_l as the **orthogonal information** set. On the other hand, the remaining information units satisfy $i_j \in \bigcap(I_{M_1}, I_{M_2}, \ldots, I_{M_p})$, where $I_{M_1}, I_{M_2}, \ldots, I_{M_p}$ represent any modalities. Then, i_j can be obtained from multiple modalities, and we refer to the set $I_{C;I_{M_1},I_{M_2}\cdots I_{M_p}}$ consisting of all i_j as the **common information** set of the modality group $I_{M_1}, I_{M_2}, \ldots, I_{M_p}$. It is important to note that the **collectivity property** $I_O \cap I_{C,*} = \emptyset$ and $I_{C;*} \subseteq I_{Ta}/\ I_O$ holds.

For successful task execution, the orthogonal information should be correct in each modality, while the common information should be correct in any of the modalities it belongs to. In the following example, the basic information units is $I_{Ta} = \{i_1, i_2, i_3, i_4, i_5\}$, and the basic information in each modality is provided in Equ. (1) (with $I_O = \{i_1, i_3, i_4\}$). It is evident that relying solely on the orthogonal information is insufficient for task execution. Utilizing the common information from the communication, sensing, and motion modalities is necessary to obtain i_2 and i_5.

$$\begin{aligned} I_{m_{com}} &= \{\ i_1,\ i_2,\ \ \ \ \ \ \ \ i_5\ \} \\ I_{m_{sen}} &= \{\ \ \ \ \ \ i_2,\ i_3,\ \ \ \ i_5\ \} \\ I_{m_{mot}} &= \{\ \ \ \ \ \ \ \ \ \ \ \ \ i_4,\ i_5\ \} \end{aligned} \tag{1}$$

B. Multimodal Task Space

1) Task Space: The task execution requires basic information units from a finite number of modalities. Thus, We can explore subsets of information units from these modalities, forming three vectors $\vec{I}_{M_*} = (i_j, \ldots, i_k)$, where $i_j, \ldots, i_k$

represent combinations of information units. For a UAV, these three modality vectors ($\overrightarrow{I}_{M_{com}}$, $\overrightarrow{I}_{M_{sen}}$, and $\overrightarrow{I}_{M_{mot}}$) combine in a tensor space (see Fig. 2). The coordinates within this space, ($\overrightarrow{I}_{M_{com}}, \overrightarrow{I}_{M_{sen}}, \overrightarrow{I}_{M_{mot}}$), correspond to all possible UAV tasks, including I_{Ta} and others. We refer to this as the multimodal task information space or simply the **task space**.

Conceptually, the task space treats all UAV tasks as distinct points composed of multimodal orthogonal information and common information, shaping a task-oriented multimodal mapping space. As the UAV's basic information aligns with $i_k \in I_{Ta} = I_{M_{com}} \cup I_{M_{sen}} \cup I_{M_{mot}}$, the subset I_O on each modality axis within the task *Ta* uniquely determines the task. Therefore, as long as the UAV can match the received coordinate points corresponding to different modality information sets with the points in the task space, it has high possibility to complete the task successfully.

Within the task space, we define the **projection operation** as follows: finding the projection of I_B onto I_A, where the operation rule is $Proj_{I_A}(I_B) = I_A \cap I_B$; when $Proj_{I_A}(I_B) = \emptyset$, we declare the two as mutually orthogonal. Notably, projection of orthogonal information onto the three modality vectors represents the intersection of all possible subsets of information related to the modality, expressed as $Proj_{I_O}(I_{M_*}) = I_O \cap I_{M_*}$, where $* \in \{com, sen, mot\}$. These projections are mutually orthogonal.

2) Task Deterministic Space: Based on orthogonality, the basic information obtained from $I_x = Proj_{I_O}(I_{M_{com}})$, $I_y = Proj_{I_O}(I_{M_{sen}})$, and $I_z = Proj_{I_O}(I_{M_{mot}})$ forms a space. In this space, all points are determined solely by multimodal orthogonal information. This space can be seen as a reduction of the dimensionality of the original task space by eliminating common information, resulting in a reduction of ΔU_{I_C} points. We refer to this space as the multimodal task information deterministic space or simply the **task deterministic space**.

Conceptually, the task deterministic space represents the extraction of orthogonal information from the basic information of all UAV tasks, independent of different modalities. It forms a task-oriented multimodal non-intersecting mapping space, where each basic information can only be obtained from a single modality. Although the dimensionality is reduced in this space by eliminating common information, it exposes the lowest reliability of task execution.

III. Processing for Enhanced Reliability

A. Multimodal Reception of the UAV

The signal to interference plus noise ratio (SINR) in the multimodal receptions of U_k can be unified as

$$\gamma_{M_*} = \frac{P^{M_{mot}}(Proj_{I_{Ta}}(I_{M_*}))}{P^{M_{mot}}(Proj_{I_{Ti}}(I_{M_*}))}, * \in \{com, sen\} \tag{2}$$

where $P^{M_{mot}}(\cdot)$ means extracting the power of $\cdot$'s carrier/medium with the dynamic influence of UAV's real-time flying motion. Specifically, with source's coordinate $d_S = (d_{S,x}, d_{S,y}, d_{S,z})$ and transmitting power P_S, we can easily derive the SINR in communication modality with

$$\begin{aligned} &P^{M_{mot}}(Proj_{I_{Ta}}(I_{M_*})) \\ &= P_{I_S}|h_c^{Ta,k}|^2 \times \Big\| \Big(d_{U,x} + V_{U,x}t - x_{I_{Ta}}\Big)^2 \\ &\quad + \Big(d_{U,y} + V_{U,y}t - y_{I_{Ta}}\Big)^2 + \Big(d_{U,z} + V_{U,z}t - z_{I_{Ta}}\Big)^2 \Big\|_2^{-\gamma}, \end{aligned} \tag{3}$$

$$\begin{aligned} &P^{M_{mot}}(Proj_{I_{Ti}}(I_{M_*})) \\ &= P_N + \sum_{j \in USet_I}^{n} P_I(j)\Big|h_c^{j,K}\Big|^2 \times \Big\| \big(d_{U,x} + d_{V,x}t - d_{j_x}\big)^2 \\ &\quad + \Big(d_{U,y} + d_{V,y}t - d_{j_y}\Big)^2 + \big(d_{U,z} + d_{V,z}t - d_{j_z}\big)^2 \Big\|_2^{-\gamma}, \end{aligned} \tag{4}$$

where $USet_I = \{U_\alpha, U_\beta, \ldots, U_\delta\}$, P_N is the noise power, $d_{U,t}(t \in \{x, y, z\})$ is the corresponding value on each axis of the position coordinates of the UAV U_k, $d_{U,t}(t \in \{x, y, z\})$is the corresponding value in each axis direction of the velocity vector of the UAV U_k, and *t* is the time it takes U_k to transfer its position in task space and γ is the path loss index. Note that the motion modality may enhance the SINR of multimodal reception by dynamic motions. If $SINR_C^k \geq SINR_C^{th}$, UAV U_k is able to obtain all the correct task information I_{Comm} from communication modality.

B. Reliability Based on Task Space

Considering the unreliability caused by multimodal reception error, the actual task information of UAV may have error, i.e., $\widetilde{I}_{Ta} = \widetilde{I}_{M_{com}} \cup \widetilde{I}_{M_{sen}} \cup \widetilde{I}_{M_{mot}} \neq I_{Ta}$, where each $\widetilde{I}_{M_*}$ may vary within a specific range in the coordinate axis (a wider range implies lower reliability). In $\widetilde{I}_{M_*} = (\widetilde{i_j}, \ldots, \widetilde{i_k})$, $\widetilde{i_j}, \ldots, \widetilde{i_k}$ represent any possible combination of basic information units for modality $*$. As a result, the fluctuation in each axis changes the distance between $\widetilde{I}_{Ta}$ and I_{Ta}, which can be represented as $\|\widetilde{I}_{Ta} - I_{Ta}\|_{TS}$. A smaller distance means $\widetilde{I}_{Ta}$ has higher probability of being I_{Ta}.

Interestingly, although the task information may be biased in the multimodal reception stage, the corresponding correct task information can still be found by utilizing the distance defined in the task space. Then, the task execution based on unreliable multimodal reception can be determined by minimizing the task space searching distance, given by

$$\min_{U_k} D_{TSS} = \min_{U_k} \|\widetilde{I}_{Ta} - I_{Ta}^{U_k}\|_{TS}, \tag{5}$$

where U_k means any correct points in the task space. We refer this method as **task space searching** (TSS), which starts from the actual points with fluctuations to search for the correct points with the minimum distance, with a complexity of $O(U_k)$.

C. Multimodal Task Information Alignment

1) Multimodal Task Information Classification: If the raw information Ri_k within the raw information layer $\mathbb{M}$ obtained from multimodal sources can be classified into two types: semantic-type information and data-type information.

Type 1 Semantic-type information: Due to semantic nature, the raw information is processed through a deep learning network to extract temporal information $t \in S_T$, spatial information $s \in S_S$, and entity behavior information vector $\overrightarrow{a} = \{Actions, FromEntity, ToEntity\} \in \overrightarrow{S_A}$ (comprising

the actions of task *Ta* along with their initiators and targets). These components collectively form the entity information layer $\mathbb{S} = \{S_T, S_S, \overrightarrow{S_A}\}$, where each component represents vectors of time, space, and action (generated through graph-based learning representing the actions in the behavior set).

By assessing the match between the information attributes $\{t, s, \overrightarrow{a}\}$ of $\mathbb{M}_i$, and a given threshold, it can be determined whether the raw information aligns with the corresponding entity in the entity layer $\mathbb{S}_i$. This mapping from the multimodal layer to the entity information layer is denoted as $\mathbb{M}_i \rightarrow \mathbb{S}_i$. Subsequently, using the semantic elements of these entities, the metadata is mapped to the basic information units of the task, resulting in a cross-modal unified representation of results. In other words, information obtained from different modalities eventually aligns to become basic task-specific information $i_k \in I_{Ta}$ after discarding task-irrelevant information that does not meet the requirements of task execution.

Type 2 Data-type information: Due to the nature of data, multimodal task information alignment primarily involves unit conversions for the same state (e.g., dBW and Watt), without any semantic conflicts during the representation mapping.

2) Reliability Enhancement via Multimodal Information: After multimodal task information alignment, inconsistencies may exist in specific values, which be addressed in the reliability enhancement phase.

Step 1 Ranking: Based on the SINR of multimodal reception, the most reliable modality is selected.

Step 2 Multimodal Combination for Common Information: In our previous paper [9], common information can be enhanced by multimodal combination (e.g., maximal ratio combining or selective combing) with enhanced diversity gain $d_m = p$, where p is the number of modalities in the intersection $I_{C;I_{M_1},I_{M_2}\cdots I_{M_p}}$.

Step 3 TSS in Task Deterministic Space: This step focuses on orthogonal information, which constitutes the task deterministic space (Section II-B2). The task points can be determined by TSS (Section III-B) with a reduced searching complexity $O(U_k - \Delta U_{I_C})$.

To determine the reliability (TSP) in this step, assume the minimum and maximum distances between correct points in the task deterministic space's axis are D_{min} and D_{max}. If $D_{TSS} \leq \frac{1}{2}D_{min}$, the TSP is 1; if $D_{TSS} > \frac{1}{2}D_{max}$, the TSP is 0. The error probability increases proportionally with distance, and we assume the multimodal reception error happens uniformly, which yields $TSP = 1 - \frac{D_{TSS} - \frac{1}{2}D_{min}}{D_{max} - \frac{1}{2}D_{min}}$.

IV. Simulations

In Fig. 3, upon removing the shared information in modality (a), the dimension in (b) corresponds to the task determination space, shows a reduction in the number of correct points from 1000 to 100. Consequently, the TSS complexity of is correspondingly reduced. Fig. 4 shows the distances (denoted by TSS) and success probabilities (denoted by TSP) between the actual point A5 and all other correct points S0 to S10. The closest standard point is S4, with a distance of 0.25 and a TSP of 100%. Similarly, employing this method for searching

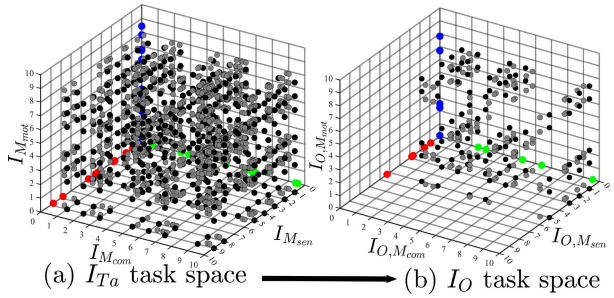

Fig. 3. Orthogonal task space without TSS. The green, red and blue points represent the certain task information. The grey and black points represent the certain actual task coordinate and the certain standard actual respectively.

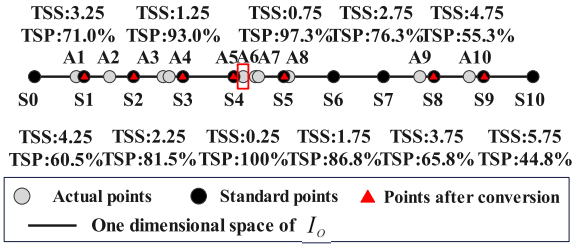

Fig. 4. Orthogonal task axis with *TSS*.

with TSS among other points A0 to A9 can easily yield the optimal task matching based on Equ. (5).

V. Conclusion

This letter has proposed multimodal UAV communication based on opportunistic task space, which decompose multimodal reception into task-specific and task-irrelevant information units, and use multimodal information alignment and task space concept to enhance the reliability.

References

[1] H. Halevi, I. Bergel, and Y. Noam, "Asymptotic performance of TDOA estimation using satellites," *IEEE Trans. Signal Process.*, vol. 70, pp. 2349–2361, 2022.

[2] C. Ren, L. Liu, and H. Zhang, "Multimodal interference compatible passive UAV network based on location-aware flexibility," *IEEE Wireless Commun. Lett.*, vol. 12, no. 4, pp. 640–643, Apr. 2023.

[3] Z. Zhao, Y. Zhang, C. Li, Y. Xiao, and J. Tang, "Thermal UAV image super-resolution guided by multiple visible cues," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5000314.

[4] C. Zhang et al., "WebUAV-3M: A benchmark for unveiling the power of million-scale deep UAV Tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 7, pp. 9186–9205, Jul. 2023.

[5] Y. Kong et al., "UAV LiDAR data-based lane-level road network generation for urban scene HD Maps," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, 2023, Art. no. 6501805.

[6] S. Shao, H. Li, Y. Zhao, and X. Wu, "A new method for multi-UAV cooperative mission planning under fault," *IEEE Access*, vol. 11, pp. 52653–52667, 2023.

[7] X. Wang, M. Yi, J. Liu, Y. Zhang, M. Wang, and B. Bai, "Cooperative data collection with multiple UAVs for information freshness in the Internet of Things," *IEEE Trans. Commun.*, vol. 71, no. 5, pp. 2740–2755, May 2023.

[8] X. Li, T. Zhang, S. Wang, G. Zhu, R. Wang, and T.-H. Chang, "Large-scale bandwidth and power optimization for multi-modal edge intelligence autonomous driving," *IEEE Wireless Commun. Lett.*, vol. 12, no. 6, pp. 1096–1100, Jun. 2023.

[9] C. Ren, C. Gong, and L. Liu, "Task-oriented multi-modal communication based on cloud-edge-UAV collaboration," *IEEE Internet Things J.*, early access, Jul. 14, 2023, doi: 10.1109/JIOT.2023.3295650.