U2UData-2: A Scalable Swarm UAVs Autonomous Flight Dataset for Long-horizon Tasks

Tongtong Feng
Department of Computer Science and Technology, Tsinghua University
Beijing, China
fengtongtong@tsinghua.edu.cn

Xin Wang*
Department of Computer Science and Technology, BNRist, Tsinghua University, Beijing, China
xin_wang@tsinghua.edu.cn

Feilin Han
Department of Film and Television Technology, Beijing Film Academy
Beijing, China
hanfeilin@bfa.edu.cn

Leping Zhang
Department of Film and Television Technology, Beijing Film Academy
Beijing, China
lepingzhang2002@gmail.com

Wenwu Zhu*
Department of Computer Science and Technology, BNRist, Tsinghua University, Beijing, China
wwzhu@tsinghua.edu.cn

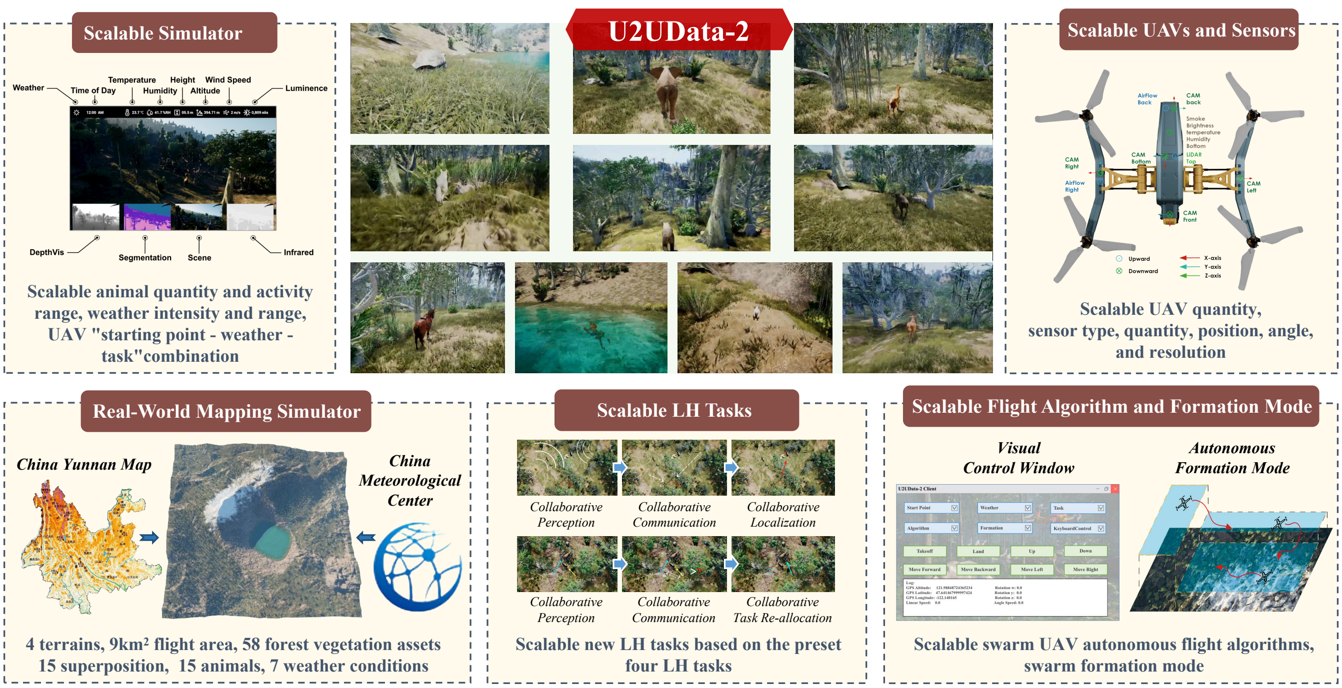


Figure 1: U2UData-2 collects a large-scale swarm UAV autonomous flight dataset for Long-Horizon (LH) tasks. U2UData-2 also provides a scalable data collection platform supporting the customization of simulators, UAVs, sensors, flight algorithms, formation modes, and LH tasks. Through its visual control window, U2UData-2 allows users to collect customized datasets through one-click deployment online and to verify algorithms by closed-loop simulation.

Abstract

Swarm UAV autonomous flight for Long-Horizon (LH) tasks is crucial for advancing the low-altitude economy. However, the maximum trajectory length in existing swarm UAV autonomous flight datasets is limited to 15 seconds per flight path, which fails to explore LH tasks. This paper presents U2UData-2, the first large-scale swarm UAV autonomous flight dataset for LH tasks and the first

*Corresponding Authors. BNRist is the abbreviation of Beijing National Research Center for Information Science and Technology.

MM '25, October 27-31, 2025, Dublin, Ireland

ACM ISBN 978-1-4503-XXXX-X/2018/06
https://doi.org/XXXXXXX.XXXXXXX

scalable data collection platform. The dataset is captured by 15 UAVs in autonomous collaborative flights for LH tasks, comprising 12 scenes (weather and terrain combination), 720 traces, 120 hours (each trace 10 minutes), 4.32M LiDAR frames, and 12.96M RGB frames. They also include brightness, temperature, humidity, smoke, and airflow values covering all flight routes. The data collection platform supports the customization of simulators, UAVs, sensors, flight algorithms, formation modes, and LH tasks. Through its visual control window, U2UData-2 allows users to collect customized datasets through one-click deployment online and to verify algorithms by closed-loop simulation. U2UData-2 can be found at https://fengtt42.github.io/U2UData-2/.

CCS Concepts

• **Computing methodologies → Cognitive robotics**; *Cooperation and coordination*; Robotic planning.

Keywords

Swarm UAVs; Autonomous Flight; Dataset; Scalable Data Collection Platform; Long-horizon Tasks

ACM Reference Format:

Tongtong Feng, Xin Wang, Feilin Han, Leping Zhang, and Wenwu Zhu. 2025. U2UData-2: A Scalable Swarm UAVs Autonomous Flight Dataset for Long-horizon Tasks. In *Proceedings of Proceedings of the 33nd ACM International Conference on Multimedia (MM '25).* ACM, New York, NY, USA, 7 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

Swarm Unmanned Aerial Vehicle (UAV) autonomous flight for Long-Horizon (LH) tasks is crucial for advancing the low-altitude economy, such as logistics distribution[1], patrol security[31], wildlife conservation[7], disaster rescue[25], and infrastructure inspection[19]. LH tasks [2, 8] are complex, multi-step tasks that require sustained planning, sequential decision-making, and extended execution over a prolonged period to achieve a final goal.

Current low-altitude economy research mainly focuses on single-UAV autonomous flight and has matured core capabilities[29, 32], including object detection, semantic segmentation, localization, obstacle avoidance, navigation, tracking, and stabilized flight control in controlled environments. However, they still suffer from many real-world challenges, for example: (1) Their perception remains fundamentally constrained by single-viewpoint occlusion and limited sensor range[11, 17], severely reducing situational awareness in dynamic open environments. (2) Onboard computational resources restrict real-time decision-making for dynamic obstacle negotiation and executing LH tasks[26, 28]. (3) Operational robustness is inherently fragile[24, 33], as hardware failures or unexpected obstacles often lead to task failure with no redundancy.

Swarm UAV autonomous flight[27] can solve the inherent limitations of single-UAV through collaborative perception, localization, communication, navigation, tracking, and task re-allocation. By leveraging UAV-to-UAV (U2U) technologies, swarm UAVs overcome single-viewpoint occlusion and sensor range limits through multi-view collaborative perception[30]. Furthermore, swarm UAV ensures operational robustness against failures or obstacles for accurate navigation[14] and dynamic tracking[16] by collaborative localization, communication, and task re-allocation, while mitigating computational constraints via shared processing and decentralized decision-making. Finally, swarm UAV autonomous flight can achieve robust, scalable, and adaptive task execution in complex and harsh environments unattainable by single-UAV systems.

Swarm UAV research strongly relies on the development of datasets. Existing swarm UAV datasets, as shown in Table 1, CoPerception-UAVs[10] and CoPerception-UAVs+[11] are based on open-source simulators such as AirSim[23] and CARLA[5] and consider only 1 terrain, 1 weather, and 1 to 2 sensor types; they collect datasets using fixed altitude and consistent or fixed formation mode. In real-world scenarios, compared to autonomous driving, autonomous flight has more freedom, faces more complex environments, and is more susceptible to the influence of temperature, humidity, and airflow due to its smaller size. Obviously, there will be a clear domain gap between existing synthetic data and real-world data. U2UData[7] is the first swarm UAVs autonomous flight dataset, which is collected by three UAVs flying autonomously in the U2USim[9], covering a 9 km^2 flight area, 4 terrains, 7 weather conditions, and 8 sensor types. Due to the emergence of U2UData, swarm UAV autonomous flight algorithms have begun to be studied.

However, the maximum trajectory length in existing swarm UAV autonomous flight datasets is limited to 15 seconds per flight path, which fails to explore LH tasks. U2UData only considers three UAVs tracking three animals, the length of each trajectory is only 15 seconds, and the dataset size is fixed and cannot be expanded; only focuses on basic collaborative perception and tracking tasks. Complex LH tasks for swarm UAV autonomous flight in dynamic open environments cannot be explored.

In this paper, we present U2UData-2, the first large-scale swarm UAV autonomous flight dataset for LH tasks and the first scalable data collection platform. On the one hand, the dataset is captured by 15 UAVs in autonomous collaborative flight for LH tasks, comprising 12 scenes (weather and terrain combination), 720 traces, 100+ hours (each trace 10 minutes), 4.32M LiDAR frames, 12.96M RGB, and 12.96M depth frames. They also include brightness, temperature, humidity, smoke, and airflow values covering all flight routes. On the other hand, as shown in Figure 1, the data collection platform supports the customization of simulators, UAVs, sensors, flight algorithms, formation modes, and LH tasks. Through its visual control window, U2UData-2 allows users to collect customized datasets through one-click deployment online and to verify algorithms by closed-loop simulation. Compared with U2UData, U2UData-2 includes more UAVs (3->15) and more longer data collection trajectory (15s->600s), which supports autonomous flight algorithms to more comprehensively explore LH tasks; U2UData-2 includes a scalable data collection platform, which can greatly alleviate the limitations of existing datasets on algorithm development.

Our contributions can be summarized as follows:

- **Dataset.** We collect the first large-scale swarm UAV autonomous flight dataset for LH tasks.
- **Scalable data collection platform.** We build the first scalable data collection platform for swarm UAV autonomous flight, which allows users to collect customized datasets through one-click deployment online and to verify algorithms by closed-loop simulation.

Table 1: A detailed comparison of swarm UAV datasets. - indicates that specific information is not provided. DF: Discipline formation mode, where swarm UAVs keep a consistent and relatively static array; FF: Fixed formation mode, where each UAV navigates independently with a fixed path; AF: Autonomous formation mode, where each UAV flies autonomously. ET-Length: Each Trajectory Length. U2USim⋆ represents the scalable U2USim.

Dataset	Year	Terrains	Weather	Sensors	Formation	Sample	Real Data	Tasks	Simulation	UAVs	ET-Lengh	Scalable
CoPerception-UAVs[10]	2022	1	1	1	DF, FF	4s	-	Basic	AirSim + Carla	5	-	N
CoPerception-UAVs+[11]	2023	1	1	2	DF, FF	4s	-	Basic	AirSim + Carla	10	-	N
U2UData[7]	2024	4	7	8	DF, FF, AF	0.03s	China	Basic	U2USim	3	15s	N
U2UData-2	2025	4	7	8	DF, FF, AF	0.03s	China	**LH**	**U2USim⋆**	**15**	**600s**	**Y**

2 Related Work

This section introduces the related work of swarm UAV simulators and datasets in detail.

Swarm UAV simlators. Existing swarm UAV simulators include FightGear[4], XPlan[21], Jmavsim[20], Gazebo[22], AirSim[23], RflySim[6], Isaac Sim[18], and U2USim[9]. Swarm UAV simulators need to more realistically simulate dynamic physical characteristics[18] (such as collision); sensors such as IMU, camera, GPS, LiDAR, temperature, humidity, and airflow due to their small size; and interaction with the ROS ecosystem. FightGear[4] is not open source. XPlan[21] and Jmavsim[20] can only interact with ROS. Gazebo[22], AirSim[23], and RflySim[6] can interact with ROS, simulate physical collision, and output visual sensor content. AirSim and RflySim can also implement weather control. However the information on these simulators is purely simulated, and the models trained on these simulators are difficult to run in the real world. Isaac Sim[18] and U2USim[9] add real environment data based on previous simulators. Isaac Sim can visually realize digital twins of the real world through GPU rendering, but it is difficult to provide modal information other than visual and LiDAR modalities. U2USim is the first real-world mapping swarm UAV simulator, taking Yunnan Province as the prototype, including 4 terrains, 7 weather conditions, and 8 sensor types. However, all parameters of U2USim are fixed: it only contains 3 types of animals, the number of animals is fixed, the intensity and range of weather are fixed, and the take-off point of the UAV is also fixed. If we want to test in another terrain, we need to fly to the target location for a long time before each test.

Swarm UAV datasets. Existing swarm UAV Datasets include CoPerception-UAVs[10], CoPerception-UAVs+[11], and U2UData[7]. Collecting datasets[3, 15] is crucial for advancing algorithms research. Public swarm UAV datasets have significantly accelerated progress in UAV flight technologies in recent years. Existing swarm UAV datasets, such as CoPerception-UAVs[10] and CoPerception-UAVs+[11], rely on open-source simulators like AirSim[23] and CARLA[5], featuring limited terrain, weather, and sensor types. These datasets collect data at fixed altitudes and in consistent or fixed formation modes. In contrast to autonomous driving, UAVs' autonomous flight presents greater freedom, encounters more complex environments, and is more susceptible to the influence of temperature, humidity, and airflow due to its smaller size. Hence, there exists a notable domain gap between existing synthetic data and real-world data, potentially limiting the generalization of models trained. U2UData[7] is the first large-scale cooperative perception dataset for swarm UAVs autonomous flight, which is collected by three UAVs flying autonomously in the U2USim[9], covering a 9 km^2 flight area, 4 terrains, 7 weather conditions, and 8 sensor types. U2UData manually selects 100 scenarios for each weather condition; U2UData collects 15 seconds of swarm UAV cooperative perception dataset for each scenario. U2UData samples the image frames at 30Hz and the LiDAR frames at 10Hz, comprising a total of 945K RGB frames, 945K depth frames, and 315K LiDAR frames. Due to the emergence of U2UData, swarm UAV autonomous flight algorithms have begun to be studied. However, since U2UData only considers three UAVs tracking three animals, the length of each trajectory is only 15 seconds, and the dataset size is fixed and cannot be expanded; only basic collaborative perception and tracking tasks can be designed. Complex LH tasks for swarm UAVs in dynamic open environments cannot be explored.

3 U2UData-2

U2UData-2 includes a large-scale swarm UAV autonomous flight dataset for LH tasks and a scalable data collection platform.

3.1 Scalable Data Collection Platform

The scalable data collection platform is based on U2USim[9], as shown in Figure 1 and Figure 2, supporting the customization of simulators, UAVs, sensors, flight algorithms, formation modes, and LH tasks. Through its visual control window, U2UData-2 allows users to collect customized datasets through one-click deployment online and to verify algorithms by closed-loop simulation.

Real-world mapping simulator. The platform is based on U2USim[9], a real-world mapping swarm UAV simulator.The platform uses Unreal Engine (UE) 5.2[1] to construct a scaled-down 3km*3km simulated environment map based on the map of Yunnan Province. The platform includes 4 types of terrain: mountains, hills, plains and basins. The elevation range is [56.6, 3000]m. Based on the vegetation and animal distribution in Yunnan, 58 types of original forest vegetation and 15 types of animal assets were constructed, and more than 15 superposition methods were used to combine vegetation assets, including epiphytic growth, diagonal staggered growth, and so on. Among them, the leaves of each plant will dynamically change with wind, rain, snow, and other weather conditions. The platform includes 7 weather conditions: sunny, rain, snow, sandstorm, wind, thunder, and fog at specific positions within the simulation environment. The platform uses the real meteorological data of Yunnan Province collected by the China Meteorological Center to map the simulation environment based on longitude and

[1]https://www.unrealengine.com/en-US/unreal-engine-5

Figure 2: Scalable data collection platform. The platform is based on a real-world mapping simulator, which can continuously collect new datasets through customized design of simulators, UAVs, sensors, flight algorithms, formation modes, and LH tasks.

latitude. Among them, temperature and humidity are scalars, and missing values are filled by the moving average method (interval 5m). Wind speed and wind direction are first decomposed into scalars along longitude and latitude, then missing values are filled by sliding average, and finally constructed by vector synthesis.

Scalable simulator. The simulator delivers extensive configurability through UE5.2, enabling dynamic adjustments to animal quantity and activity ranges, weather intensity and coverage, and UAV "starting point-weather-task" combinations. In the simulator startup interface, users can directly click the F11 key on the keyboard to make visual adjustments; input the animal quantity and the activity radius value; fine-tune the weather parameters using intuitive sliders, including intensity (e.g., rainfall severity, fog density) and spatial range. Six predefined UAV starting points are mapped to specific weather scenarios (rain, snow, sandstorm, thunder, fog, and sunny). Since wind is located throughout the map, there is no specific starting point setting. The starting point, weather, and task are added as options to the visual control window, and users can select from drop-down menus to implement custom "starting point-weather-task" combinations.

Scalable UAVs and sensors. The platform includes 8 sensor types: RGB, depth, LiDAR, brightness, temperature, humidity, smoke, and airflow. These sensors are installed on the multirotor to explore the simulator map and collect data at 0.03-second intervals, which can be customized using a JSON settings file ("setting.json"). In this JSON file, UAV quantity is customizable; the type, quantity, position, angle, and resolution of sensors are also customizable; users can edit the JSON file to customize their own multirotor by selecting practical sensors and designing the sensor parameters.

Scalable LH tasks. The platform provides four preset LH tasks: wildlife conservation employs adaptive animal tracking algorithms using real-time behavior prediction across variable terrains and vegetation density. Logistics distribution dynamically reroutes paths around simulated urban obstacles and weather disruptions while maintaining payload integrity. Patrol security implements anomaly detection through continuous environmental scanning, adapting surveillance patterns to emergent threats in real-time. Disaster rescue prioritizes survivor identification in volatile conditions (collapsing structures, spreading fires) via multi-sensor fusion and probabilistic hazard mapping. Each task integrates specialized perception-action loops that respond to unpredictable environmental changes without predefined waypoints, such as sudden weather shifts or moving obstacles. New LH tasks (e.g., precision agriculture) can be added by modifying the UE5.2 simulator source code.

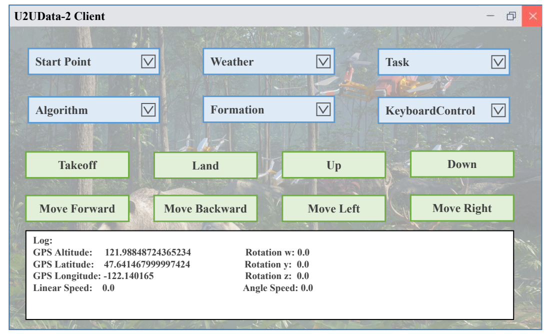


Figure 3: Visual control window. Users can collect customized datasets through one-click deployment online and verify algorithms by closed-loop simulation.

Scalable flight algorithm and formation mode. The platform supports four swarm UAV autonomous flight algorithms for four LH tasks: wildlife conservation, logistics distribution, patrol security, and disaster rescue. Those algorithms are built upon modularized components, including task planning, collaborative perception, localization, communication, navigation, tracking, and task re-allocation. Users can add or modify these algorithms via the open-source code of the visual control window, where modular code blocks allow drag-and-drop replacement or augmentation of existing logic. New autonomous flight algorithms for custom LH tasks can be integrated by directly modifying the provided Python/ROS 2 interfaces in the code repository. The platform implements three distinct swarm formation modes: Discipline formation mode maintains strict geometric coordination (e.g., linear/radial arrays) for high-precision collaborative tasks, with real-time position correction compensating for environmental disturbances. Fixed formation mode enables individual UAVs to follow predefined paths, critical for infrastructure inspection or convoy protection scenarios. Autonomous formation mode supports dynamic reconfiguration where UAVs independently adapt spacing and topology using real-time perception data, ideal for complex environments like wildlife conservation or disaster rescue. Users can select any swarm formation modes via the visual control window for task-specific optimization.

Visual control window. The platform provides a visual control window, as shown in Figure 3. Users can collect datasets through a one-click deployment of the customized UAV starting point, weather,

Table 2: A detailed comparison of the data size between U2UData-2 with existing swarm UAV datasets.

Datasets	RGB		Depth	LiDAR	New Sensors				
	RGB	Resolution			Airflow	Brightness	Temperature	Humidity	Smoke
CoPerception-UAVs[10]	131.9K	800*450	-	-	-	-	-	-	-
CoPerception-UAVs+[11]	52.76K	800*450	-	-	-	-	-	-	-
U2UData[7]	945K	1920*1080	945K	315K	1.89M	945K	945K	945K	945K
U2UData-2	12.96M	1920*1080	12.96M	4.32M	25.92M	12.96M	12.96M	12.96M	12.96M

Table 3: Scene settings. ESTN: The trajectory number of each scene.

Weather	Scenes	ESTN
Single-weather	Sunny, Rain, Snow, Sandstorm, Thunder, Fog	5
Cross-weather	Sunny->Rain, Sunny->Snow, Sunny->Fog, Sunny->Sandstorm, Rain->Thunder, Rain->Snow	3

LH tasks, swarm UAV autonomous flight algorithms, and swarm formation mode. Users also can verify swarm UAV autonomous flight algorithms by closed-loop simulation. For the platform basic capability test, users can first click the "keyboardControl" button, and then control the UAV by clicking the following buttons: "Take off", "Land", "Up", "Down", "Move Forward", "Move Backward", "Move Left", and "Move Right".

3.2 Dataset

U2UData-2 provides a large-scale swarm UAV autonomous flight dataset. The dataset is collected by 15 UAVs in autonomous formation mode for the LH task (wildlife conservation).

Sensor setting. The dataset bulids a comprehensive sensor suite including 5 RGBD cameras (1920x1080 resolution, 90° FOV, 30Hz sample rate), one 64-channel LiDAR (1 million points/second, 200m capturing range, ±3cm accuracy, -30° to 30° vertical FOV, -180° to 180° horizontal FOV, 10Hz sample rate), two airflow sensors measuring latitudinal and longitudinal wind speeds, and a GPS and IMU system providing odometry data. Complementary environmental sensors comprise one brightness sensor, one temperature sensor, one humidity gauge, and one smoke sensor. Navigation is enabled by integrated GPS and IMU systems providing odometry data. As shown in Figure 1, all UAVs are equipped with 5 RGBD cameras (front, back, left, right, and bottom), a 64-LiDAR sensor (top), 1 brightness, temperature, humidity, and smoke sensor (bottom), 2 airflow sensors (back and right), and GPS/IMU systems. This multi-sensor configuration supports real-time environmental interaction across dynamic scenarios from LiDAR-based terrain mapping in the dense forest to airflow-adaptive flight control during storms. The synchronized RGBD cameras enable high-fidelity object tracking essential for wildlife monitoring.

Scene setting. The simulator map first is divided into 6 areas. Except for wind, which is located throughout the map, other weathers are deployed in specific areas and have no intersection. For specific area locations, please watch the web page demonstration video. Since each specific area has different terrain, weather and terrain are strongly coupled. As shown in Table 3, we construct 12 scenes based on the most common weather combinations.

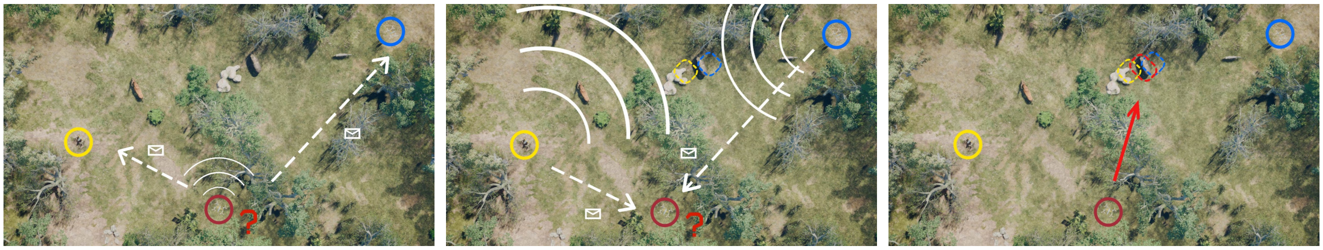
Dataset collection. As shown in Table 4, the dataset is collected by 15 UAVs in autonomous formation mode for the LH task (wildlife conservation), comprising 12 scenes, 3 or 5 trajectories of each scene, 600s length for each trajectory. As shown in Table 2, we collect a total of 12.96M RGB frames, 12.96M depth frames, 4.32M LiDAR frames, 25.92M airflow frames, 12.96M brightness frames, 12.96M temperature frames, 12.96M humidity frames, and 12.96M smoke frames. The total length of the entire dataset is 120 hours.

Table 4: Data collection settings between U2UData-2 with existing swarm UAV datasets. ESTN: The trajectory number of each scene. ET-Length: The length of each trajectory.

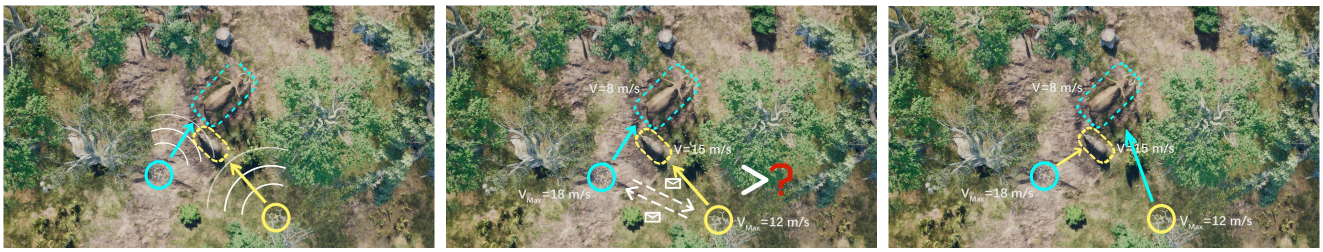
Datasets	UAVs	Scenes	ESTN	ET-Length
CoPerception-UAVs[13]	5	1	-	-
CoPerception-UAVs+[11]	10	1	-	-
U2UData[7]	3	7	100	15s
U2UData-2	15	12	3 or 5	600s

Table 5: A detailed comparison between U2UData and U2UData-2. Basic tasks: collaborative perception and tracking. LH tasks: wildlife conservation, logistics distribution, patrol security, disaster rescue based on collaborative perception, localization, communication, navigation, tracking, and task re-allocation. ⋆ represents the scalable. ✔ represents the newly added function of U2UData-2.

Comparison	**U2UData**	**U2UData-2**
Tasks	Basic tasks	LH tasks ⋆
Each Trajectory Length	15s	600s ⋆
ALL Trajectory Length	8.75h	120h ⋆
UAV Number	3	15 ⋆
Tracking Goal	3	15 ⋆
Sensor	8	8 ⋆
Flight Start Pointing	Fixed	Selected ⋆
Flight Algorithm	Fixed	Selected ⋆
Visual Control Window	No	✔
Data online Collection	No	✔
Algorithm Closed-loop	No	✔



(a) Collaborative communication (b) Collaborative perception (c) Collaborative localization



(d) Collaborative perception (e) Collaborative communication (f) Collaborative task re-allocation

Figure 4: The visualization of the U2UData-2 dataset. we select two swarm UAV collaboration clips and annotate them.

3D bounding boxes annotation. For annotating 3D bounding boxes on the gathered LiDAR data, we utilize SusTechPoint[12], a robust open-source labeling tool. There are a total of 15 object classes, we annotate its 3D bounding box with 7 degrees of freedom, encompassing its location (x, y, z) and rotation (expressed as quaternions: w, x, y, z). The location (x, y, z) corresponds to the center of the bounding box. These 3D bounding boxes are annotated separately based on the global coordinate system of each UAV. This approach enables the sensor data from each UAV to be treated independently as a single-agent detection task. We initialize the relative pose of the two UAVs for each frame using positional information provided by the GPS on both UAVs.

Data usage. We randomly divide the dataset into training sets, validation sets, and test sets according to the ratio of 0.7/0.15/0.15. It can greatly facilitate the credibility of algorithm performance compared to different papers.

3.3 U2UData-2 vs. U2UData

U2UData-2 significantly expands upon its predecessor (U2UData) by transitioning from basic collaborative perception and tracking tasks to scalable LH tasks based on multi-UAV collaborative perception, localization, communication, navigation, tracking, and task re-allocation, such as wildlife conservation, logistics distribution, patrol security, and disaster rescue. As shown in Table 5, key enhancements include 40× longer UAV trajectory of each scene (600s vs 15s) and 13.7× greater total trajectory duration (120h vs 8.75h), alongside 5× increases in UAV number (15 vs 3) and tracking targets number (15 vs 3). While retaining eight sensors per UAV, U2UData-2 introduces dynamic flight algorithm selection and customizable starting points. Crucially, it adds three core innovations: a visual control window for real-time monitoring, one-click online data collection, and closed-loop algorithm validation. Those functionalities are absent in the original U2UData. Most importantly, U2UData-2 establishes a scalable framework for swarm UAV autonomous flight in dynamic open environments, which can greatly alleviate the limitations of existing datasets on algorithm development.

3.4 Visualization

U2UData-2 dataset is the first large-scale swarm UAV autonomous flight dataset for the LH task (wildlife conservation). As shown in Figure 4, we select two swarm UAV collaboration clips and annotate them. The first clip ((a)-(c)) demonstrates that the target localization accuracy of a single UAV is limited due to obstacle obstruction and restricted field of view; swarm UAVs eliminate target localization errors through collaborative communication, perception, and localization. The second clip ((d)-(f)) illustrates that a single UAV makes it difficult to complete complex tasks independently due to its hardware limitations; swarm UAVs can improve the robustness of completing LH tasks through collaborative perception, communication, and task re-allocation. These visualizations highlight the dataset's ability to design algorithms for LH tasks.

4 Conclusion

Swarm UAV autonomous flight for LH tasks is crucial for advancing the low-altitude economy. In this paper, we present U2UData-2, the first large-scale swarm UAV autonomous flight dataset for LH tasks and the first scalable data collection platform. The dataset is captured by 15 UAVs in autonomous collaborative flight for LH tasks, comprising 12 scenes, 720 traces, and 100+ hours (each trace 10 minutes). The data collection platform supports the customization of simulators, UAVs, sensors, flight algorithms, formation modes, and LH tasks. Through a visual control window, U2UData-2 allows users to collect datasets through one-click deployment online and to verify algorithms by closed-loop simulation. We hope U2UData-2 can assist UAV algorithms in being deployed in the real world.

References

[1] Francesco Betti Sorbelli. 2024. UAV-based delivery systems: A systematic review, current trends, and research challenges. *Journal on Autonomous Transportation Systems* 1, 3 (2024), 1–40.

[2] Xiaoshuai Chen, Wei Chen, Dongmyoung Lee, Yukun Ge, Nicolás Rojas, and Petar Kormushev. 2024. A Backbone for Long-Horizon Robot Task Understanding. *IEEE Robotics and Automation Letters* 10 (2024), 2048–2055.

[3] Xi Chen, Yun Xiong, Siqi Wang, Haofen Wang, Tao Sheng, Yao Zhang, and Yu Ye. 2023. ReCo: A dataset for residential community layout planning. In *Proceedings of the 31st ACM International Conference on Multimedia*. 397–405.

[4] DigitalOcean. [n. d.]. FlightGear. https://www.flightgear.org/.

[5] Alexey Dosovitskiy, German Ros, Felipe Codevilla, Antonio Lopez, and Vladlen Koltun. 2017. CARLA: An open urban driving simulator. In *Conference on robot learning*. PMLR, 1–16.

[6] FEISILAB. [n. d.]. RflySim. https://rflysim.com/doc/zh/.

[7] Tongtong Feng, Xin Wang, Feilin Han, Leping Zhang, and Wenwu Zhu. 2024. U2UData: A Large-scale Cooperative Perception Dataset for Swarm UAVs Autonomous Flight. In *ACM Multimedia 2024*.

[8] Weihang Guo, Zachary K. Kingston, and Lydia E. Kavraki. 2024. CaStL: Constraints as Specifications through LLM Translation for Long-Horizon Task and Motion Planning. *ArXiv* (2024).

[9] Feilin Han, Leping Zhang, Xin Wang, Ke-Ao Zhao, Ying Zhong, Ziyi Su, Tongtong Feng, and Wenwu Zhu. 2024. U2USim - A UAV Telepresence Simulation Platform with Multi-agent Sensing and Dynamic Environment. In *Proceedings of the 32nd ACM International Conference on Multimedia*. 11258–11260.

[10] Yue Hu, Shaoheng Fang, Zixing Lei, Yiqi Zhong, and Siheng Chen. 2022. Where2comm: Communication-efficient collaborative perception via spatial confidence maps. *Advances in neural information processing systems* 35 (2022), 4874–4886.

[11] Yue Hu, Yifan Lu, Runsheng Xu, Weidi Xie, Siheng Chen, and Yanfeng Wang. 2023. Collaboration helps camera overtake lidar in 3d detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 9243–9252.

[12] E Li, Shuaijun Wang, Chengyang Li, Dachuan Li, Xiangbin Wu, and Qi Hao. 2020. SUSTech POINTS: A Portable 3D Point Cloud Interactive Annotation Platform System. In *2020 IEEE Intelligent Vehicles Symposium (IV)*. 1108–1115.

[13] Yiming Li, Dekun Ma, Ziyan An, Zixun Wang, Yiqi Zhong, Siheng Chen, and Chen Feng. 2022. V2X-Sim: Multi-agent collaborative perception dataset and benchmark for autonomous driving. *IEEE Robotics and Automation Letters* 7, 4 (2022), 10914–10921.

[14] Yiming Li, Shunli Ren, Pengxiang Wu, Siheng Chen, Chen Feng, and Wenjun Zhang. 2021. Learning distilled collaboration graph for multi-agent perception. *Advances in Neural Information Processing Systems* 34 (2021), 29541–29552.

[15] Hongpeng Lin, Ludan Ruan, Wenke Xia, Peiyu Liu, Jingyuan Wen, Yixin Xu, Di Hu, Ruihua Song, Wayne Xin Zhao, Qin Jin, et al. 2023. TikTalk: A Video-Based Dialogue Dataset for Multi-Modal Chitchat in Real World. In *Proceedings of the 31st ACM International Conference on Multimedia*. 1303–1313.

[16] Yen-Cheng Liu, Junjiao Tian, Nathaniel Glaser, and Zsolt Kira. 2020. When2com: Multi-agent perception via communication graph grouping. In *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition*. 4106–4115.

[17] Yifan Lu, Yue Hu, Yiqi Zhong, Dequan Wang, Siheng Chen, and Yanfeng Wang. 2024. An Extensible Framework for Open Heterogeneous Collaborative Perception. *The Twelfth International Conference on Learning Representations* (2024).

[18] NVIDIA. [n. d.]. Isaac Sim. https://developer.nvidia.com/isaac-sim.

[19] Yue Pan, Linfeng Li, Jianjun Qin, Jin-Jian Chen, and Paolo Gardoni. 2024. Unmanned aerial vehicle–human collaboration route planning for intelligent infrastructure inspection. *Computer-Aided Civil and Infrastructure Engineering* 39, 14 (2024), 2074–2104.

[20] PX4. [n. d.]. Jmavsim. https://docs.px4.io/main/en/sim_jmavsim/.

[21] Laminar Research. [n. d.]. XPlan. https://www.x-plane.com/.

[22] Open Robotics. [n. d.]. Gazebo. https://gazebosim.org/home.

[23] Shital Shah, Debadeepta Dey, Chris Lovett, and Ashish Kapoor. 2018. Airsim: High-fidelity visual and physical simulation for autonomous vehicles. In *Field and Service Robotics: Results of the 11th International Conference*. 621–635.

[24] Dengdi Sun, Leilei Cheng, Song Chen, Chenglong Li, Yun Xiao, and Bin Luo. 2023. UAV-Ground Visual Tracking: A Unified Dataset and Collaborative Learning Approach. *IEEE Transactions on Circuits and Systems for Video Technology* (2023).

[25] Geng Sun, Long He, Zemin Sun, Qingqing Wu, Shuang Liang, Jiahui Li, Dusit Niyato, and Victor CM Leung. 2024. Joint task offloading and resource allocation in aerial-terrestrial UAV networks with edge and fog computing for post-disaster rescue. *IEEE Transactions on Mobile Computing* 23, 9 (2024), 8582–8600.

[26] Luca Valente, Alessandro Nadalini, Asif Hussain Chiralil Veeran, Mattia Sinigaglia, Bruno Sá, Nils Wistoff, Yvan Tortorella, Simone Benatti, Rafail Psiakis, Ari Kulmala, et al. 2024. A heterogeneous risc-v based soc for secure nano-uav navigation. *IEEE Transactions on Circuits and Systems I: Regular Papers* 71, 5 (2024), 2266–2279.

[27] Tsun-Hsuan Wang, Sivabalan Manivasagam, Ming Liang, Bin Yang, Wenyuan Zeng, and Raquel Urtasun. 2020. V2vnet: Vehicle-to-vehicle communication for joint perception and prediction. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16*. 605–621.

[28] Sizhe Wei, Yuxi Wei, Yue Hu, Yifan Lu, Yiqi Zhong, Siheng Chen, and Ya Zhang. 2024. Asynchrony-Robust Collaborative Perception via Bird's Eye View Flow. *Advances in Neural Information Processing Systems* 36 (2024).

[29] Runsheng Xu, Zhengzhong Tu, Hao Xiang, Wei Shao, Bolei Zhou, and Jiaqi Ma. 2022. Cobevt: Cooperative bird's eye view semantic segmentation with sparse transformers. *arXiv preprint arXiv:2207.02202* (2022).

[30] Runsheng Xu, Hao Xiang, Zhengzhong Tu, Xin Xia, Ming-Hsuan Yang, and Jiaqi Ma. 2022. V2x-vit: Vehicle-to-everything cooperative perception with vision transformer. In *European conference on computer vision*. 107–124.

[31] Zihao Yuan, Fangfang Xie, and Tingwei Ji. 2024. Patrol Agent: An Autonomous UAV Framework for Urban Patrol Using on Board Vision Language Model and on Cloud Large Language Model. In *2024 6th International Conference on Robotics and Computer Vision (ICRCV)*. IEEE, 237–242.

[32] Kunyi Zhang, Tiankai Yang, Ziming Ding, Sheng Yang, Teng Ma, Mingyang Li, Chao Xu, and Fei Gao. 2022. The visual-inertial-dynamical multirotor dataset. In *2022 International Conference on Robotics and Automation (ICRA)*. 7635–7641.

[33] Binyu Zhao, Wei Zhang, and Zhaonian Zou. 2023. BM2CP: Efficient Collaborative Perception with LiDAR-Camera Modalities. *arXiv preprint arXiv:2310.14702* (2023).