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

Tongtong Feng[1], **Xin Wang**[1*], **Feilin Han**[2], **Leping Zhang**[2], **Wenwu Zhu**[1*]

[1]Department of Computer Science and Technology, BNRist, Tsinghua University, China
[2]School of Intelligent Imagery Engineering, Beijing Film Academy, China
{fengtongtong, xin_wang, wwzhu}@tsinghua.edu.cn, {hanfeilin, zhangleping}@bfa.edu.cn

Abstract

Swarm UAV autonomous flight for Embodied Long-Horizon (ELH) tasks is crucial for advancing the low-altitude economy. However, existing methods focus only on specific basic tasks due to dataset limitations, failing in real-world deployment for ELH tasks. ELH tasks are not mere concatenations of basic tasks, requiring handling long-term dependencies, maintaining embodied persistent states, and adapting to dynamic goal shifts. This paper presents *U2UData+*, the first large-scale swarm UAV autonomous flight dataset for ELH tasks and the first scalable swarm UAV data online collection and algorithm closed-loop verification platform. The dataset is captured by 15 UAVs in autonomous collaborative flights for ELH tasks, comprising 12 scenes, 720 traces, 120 hours, 600 seconds per trajectory, 4.32M LiDAR frames, and 12.96M RGB frames. This dataset also includes brightness, temperature, humidity, smoke, and airflow values covering all flight routes. The platform supports the customization of simulators, UAVs, sensors, flight algorithms, formation modes, and ELH tasks. Through a visual control window, this platform allows users to collect customized datasets through one-click deployment online and to verify algorithms by closed-loop simulation. U2UData+ also introduces an ELH task for wildlife conservation and provides comprehensive benchmarks with 9 SOTA models.

Dataset — https://fengtt42.github.io/U2UData-2/

Introduction

Swarm Unmanned Aerial Vehicle (UAV) autonomous flight (Wang et al. 2020) 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 (Xu et al. 2022b). Furthermore, swarm UAV ensures operational robustness against failures or obstacles for accurate navigation (Li et al. 2021) and dynamic tracking (Liu et al. 2020) 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 autonomous flight for Embodied Long-Horizon (ELH) tasks is crucial for advancing the low-altitude economy. ELH tasks (Feng et al. 2025b) are complex, multi-step tasks that require sustained embodied planning, sequential decision-making, and extended execution over a prolonged period to achieve a final goal. The practical applications of swarm UAV are almost all ELH tasks, such as logistics distribution (Betti Sorbelli 2024), wildlife conservation (Feng et al. 2024b), disaster rescue (Sun et al. 2024), and infrastructure inspection (Pan et al. 2024).

However, existing methods focus only on specific basic tasks due to dataset limitations, failing in real-world deployment for ELH tasks. Existing swarm UAV flight datasets, as shown in Table 1, CoPerception-UAVs (Hu et al. 2022) and CoPerception-UAVs+ (Hu et al. 2023) are based on open-source simulators such as AirSim (Shah et al. 2018) and CARLA (Dosovitskiy et al. 2017) 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 (Liu et al. 2024), 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 (Feng et al. 2024b) is the first swarm UAVs autonomous flight dataset, which is collected by 3 UAVs flying autonomously in the U2USim (Han et al. 2024), covering 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. But U2UData is hard to use for exploring ELH tasks: 1) the length of each trajectory in U2UData is only 15 seconds and only focuses on basic collaborative perception and tracking tasks; 2) the dataset size, tasks, and settings are preset and fixed and cannot be expanded. ELH tasks are not mere concatenations of basic tasks, requiring handling long-term dependencies, maintaining persistent states, and adapting to dynamic goal shifts. Therefore, building a scalable swarm UAV autonomous flight dataset for ELH tasks is an urgent and challenging work for real-world deployment.

*Corresponding authors.
Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

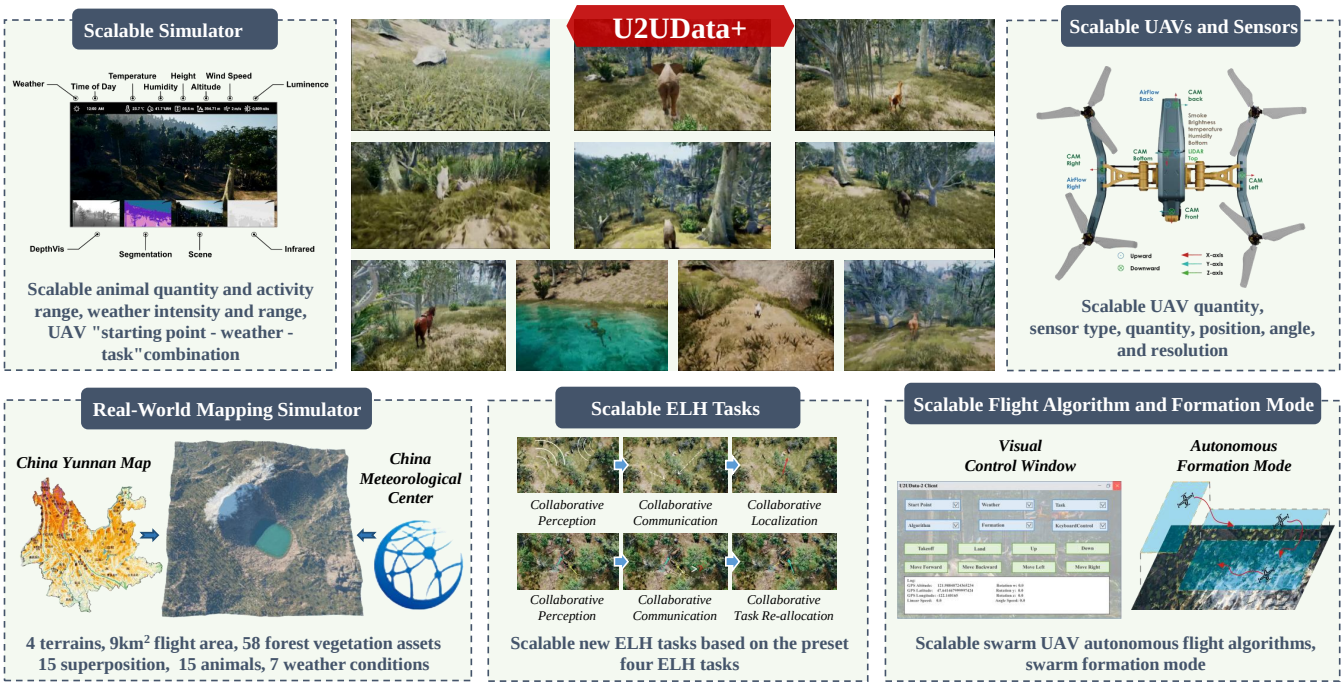


Figure 1: U2UData+ collects a large-scale swarm UAV autonomous flight dataset for ELH tasks. U2UData+ also bulid a scalable swarm UAV data online collection and algorithm closed-loop verification platform, supporting the customization of simulators, UAVs, sensors, flight algorithms, formation modes,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

In this paper, we present *U2UData+*, as shown in Figure 1, the first large-scale swarm UAV autonomous flight dataset for ELH tasks and the first scalable swarm UAV data online collection and algorithm closed-loop verification platform. 1) The dataset is captured by 15 UAVs in autonomous collaborative flight for ELH tasks (dataset size: 3.62T), comprising 12 scenes (weather and terrain combination), 720 traces, 120 hours (each trace 600 seconds), 4.32M LiDAR frames, 12.96M RGB, and 12.96M depth frames. This dataset also includes brightness, temperature, humidity, smoke, and airflow values covering all flight routes. 2) The platform supports the customization of simulators, UAVs, sensors, flight algorithms, formation modes, and ELH tasks. Through a visual control window, this platform allows users to collect customized datasets through one-click deployment online and to verify algorithms by closed-loop simulation, which can greatly alleviate the limitations of existing datasets on algorithm development. 3) U2UData+ also introduces an ELH task for wildlife conservation and provides 9 state-of-the-art swarm algorithms for benchmarking. All datasets, platforms, benchmarks, and video tutorials have been open-sourced and are available for public use. Our contributions can be summarized as follows:

- **Dataset.** We collect the first swarm UAV autonomous flight dataset for ELH tasks, with a size of over 3.62T.
- **Platform.** We build the first scalable swarm UAV data online collection and algorithm closed-loop verification platform, which allows users to collect customized datasets and verify algorithms.
- **Benchmark.** We introduce an ELH task for wildlife conservation and provide comprehensive benchmarks with 9 SOTA models.

Related Work

This section introduces the related work of Swarm UAV autonomous flight methods, simulators and datasets in detail.

Swarm UAV autonomous flight. Current low-altitude economy research mainly focuses on single-UAV autonomous flight and has matured core capabilities (Zhang et al. 2022; Xu et al. 2022a; Feng et al. 2024a), 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 (Hu et al. 2023; Lu et al. 2024), severely reducing situational awareness in dynamic open environments. (2) Onboard computational resources restrict real-time decision-making for dynamic obstacle negotiation and executing ELH tasks (Feng et al. 2025a; Li et al. 2025; Shen et al. 2025). (3) Operational robustness is inherently fragile (Sun et al. 2023; Zhao, Zhang, and Zou 2023), as hardware failures or unexpected obstacles often lead to task failure with no redundancy. Swarm UAV autonomous flight (Wang et al. 2020) can solve the inherent limitations of single-UAV through collaborative perception, localization, communication, navigation, tracking, and task re-allocation. Due to the lack of datasets, research on au-

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	Real Data	Tasks	Simulation	UAVs	ET-Lengh	Scalable
CoPerception-UAVs	2022	1	1	1	DF, FF	-	Basic	AirSim + Carla	5	-	N
CoPerception-UAVs+	2023	1	1	2	DF, FF	-	Basic	AirSim + Carla	10	-	N
U2UData	2024	4	7	8	DF, FF, AF	China	Basic	U2USim	3	15s	N
U2UData+	2025	4	7	8	DF, FF, AF	China	**ELH**	**U2USim⋆**	**15**	**600s**	**Y**

tonomous flight algorithms for swarm UAV has just begun.

Swarm UAV simulators. Existing swarm UAV simulators include FightGear (DigitalOcean 2024), XPlan (Research 2025), Jmavsim (PX4 2025), Gazebo (Robotics 2025), AirSim (Shah et al. 2018), Rfly-Sim (FEISILAB 2025), Isaac Sim (NVIDIA 2025), and U2USim (Han et al. 2024). Swarm UAV simulators need to more realistically simulate dynamic physical characteristics (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 is not open source. XPlan and Jmavsim can only interact with ROS. Gazebo, AirSim, and RflySim 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 and U2USim 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. Public swarm UAV datasets have significantly accelerated progress in UAV flight technologies in recent years. As shown in Table 1, existing swarm UAV datasets include CoPerception-UAVs (Hu et al. 2022), CoPerception-UAVs+ (Hu et al. 2023), and U2UData (Feng et al. 2024b). CoPerception-UAVs (Hu et al. 2022) and CoPerception-UAVs+ (Hu et al. 2023) rely on open-source simulators like AirSim (Shah et al. 2018) and CARLA (Dosovitskiy et al. 2017), 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 (Liu et al. 2024), 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 (Feng et al. 2024b) is the first large-scale cooperative perception dataset for swarm UAVs autonomous flight, which is collected by three UAVs flying autonomously in the U2USim (Han et al. 2024), 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. 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 and setting is fixed and cannot be expanded; only basic collaborative perception and tracking tasks can be designed. Complex ELH tasks (Liu et al. 2025; Wang et al. 2025) for swarm UAVs in dynamic open environments cannot be explored.

U2UData+ Platform

U2UData+ bulids the first scalable swarm UAV data online collection and algorithm closed-loop verification platform, as shown in Figure 1, supporting the customization of simulators, UAVs, sensors, flight algorithms, formation modes, and ELH tasks. Through a visual control window, U2UData+ allows users to collect customized datasets through one-click deployment online and to verify algorithms by closed-loop simulation. We have built a *video tutorial* for this platform. For each scalable operation, users can complete it one by one according to the video tutorial.

Real-world mapping simulator. The platform is based on U2USim (Han et al. 2024), 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, snow, and other weather conditions. This platform can represent complex ecological relationships between animals. It accurately models interactions such as predator-prey, show-

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

(a) Simulator map (b) Default starting point-weather-task (c) Custom starting point-weather-task

Figure 2: The scalable UAVs data online collection and algorithm closed-loop verification platform. Wind throughout the map.

ing how these relationships influence collective movement patterns. 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 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. As shown in Figure 2, 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, such as the Range and Number-Of-Channels of LiDAR sensors. The JSON file contains a total of 132 customizable parameters. Users can edit the JSON file to customize their own multirotor by selecting practical sensors and designing the sensor parameters. The specific meaning of each parameter and its modification range have been annotated in the open-source

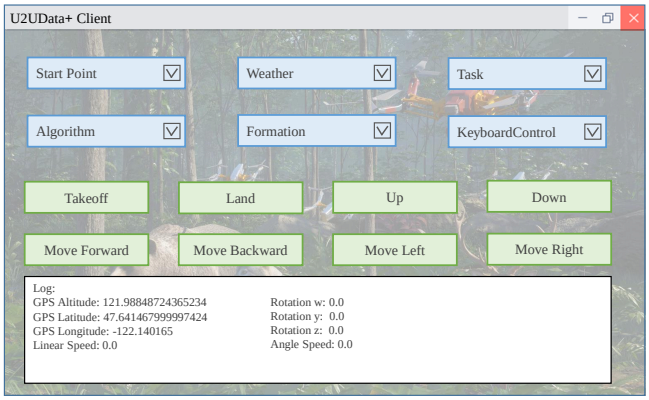


Figure 3: Visual control window. Users can collect customized datasets through one-click deployment online and verify algorithms by closed-loop simulation.

platform code.

Scalable ELH tasks. The platform provides four preset ELH 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 multisensor 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 ELH tasks (e.g., precision agriculture) can be added by modifying the UE5.2 simulator source code.

Scalable flight algorithms and formation modes. The platform supports four swarm UAV autonomous flight algorithms for four ELH 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

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

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

code blocks allow drag-and-drop replacement or augmentation of existing logic. New autonomous flight algorithms for custom ELH 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, ELH tasks, swarm UAV autonomous flight algorithms, and swarm formation mode. Users can also 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". We've also implemented XBOX controller control. Connect your XBOX and open the simulator to directly control the UAV with the controller. It's important to note that XBOX controller control and keyboard control are mutually exclusive.

U2UData+ Dataset

U2UData+ collects the first swarm UAV autonomous flight dataset for ELH tasks, with a size of over 3.62T. The dataset is collected by 15 UAVs in autonomous formation mode for the ELH task (wildlife conservation).

ELH tasks. U2UData+ only collects one ELH task: wildlife conservation, with a size of over 3.62T. Due to the huge amount of data, users can collect datasets for other ELH tasks on the U2UData+ platform on their own.

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 sensor, 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 multisensor 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.

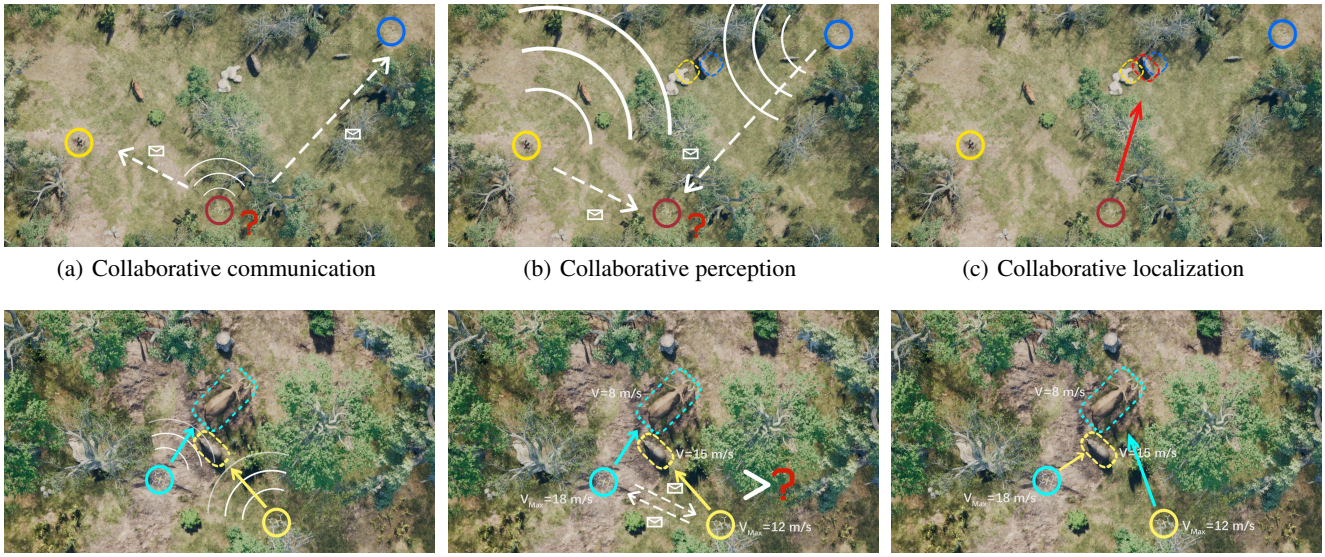
Table 3: Swarm UAV flight scene settings. ESTN: The trajectory number of each scene.

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

Table 4: Data collection settings between U2UData+ with existing swarm UAV datasets. ET-Length: The length of each trajectory. TNT: The total length of trajectories.

Datasets	UAVs	Scenes	ET-Length	TLT
CoPerception-UAVs	5	1	-	-
CoPerception-UAVs+	10	1	-	-
U2UData	3	7	15s	8.75h
U2UData+	15	12	600s	120h

Scene setting. The simulator map is first divided into 6 areas. Except for wind, which is located throughout the map, other weather is deployed in specific areas and has 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. For single-weather scenes, the trajectory of each scene collected by U2UData+ is 5. For cross-weather scenes, the trajectory of each scene collected



(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+ dataset. we select two swarm UAV collaboration clips and annotate them.

by U2UData+ is 3.

Dataset collection. As shown in Table 4, the dataset is collected by 15 UAVs in autonomous formation mode for the ELH task (wildlife conservation), comprising 12 scenes, 720 trajectories, and 600 seconds in length for each trajectory. The sampling interval of each sensor is 0.03 seconds and is synchronized in real time. 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. The total size of U2UData+ is 3.62T. The dataset has been open-sourced and are available for public use.

3D bounding boxes annotation. For annotating 3D bounding boxes on the gathered LiDAR data, we utilize SusTechPoint (Li et al. 2020), a robust open-source labeling tool. There are a total of 15 object classes, and we annotate their 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 IMU 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 the algorithm's performance compared to different papers.

U2UData+ vs. U2UData. U2UData+ significantly expands upon its predecessor (U2UData) by transitioning from basic collaborative perception and tracking tasks to scalable ELH tasks (wildlife conservation) based on multi-UAV collaborative perception, localization, communication, navigation, tracking, and task re-allocation. 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+ 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.

Table 5: A detailed comparison between U2UData and U2UData+. Basic tasks: collaborative perception and tracking. ELH tasks: wildlife conservation based on collaborative perception, localization, communication, navigation, tracking, and task re-allocation. ⋆ represents the scalable. ✔ represents the newly added function of U2UData+.

Comparison	U2UData	U2UData+
Tasks	Basic tasks	ELH 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	✔

Table 6: Swarm UAV collaborative tracking benchmark for ELH tasks in the U2UData+ dataset.

Methods	AMOTA(↑)	AMOTP(↑)	sAMOTA(↑)	MOTA(↑)	MT(↑)	ML(↓)
No Fusion	9.36	25.48	32.19	23.47	18.67	65.52
Late Fusion	14.62	31.68	47.96	37.41	27.93	37.28
Early Fusion	18.61	32.45	43.80	41.64	25.51	34.48
When2Com	20.16	34.32	49.47	45.74	30.69	32.51
DiscoNet	20.94	37.56	52.63	46.79	32.50	29.47
V2VNet	23.47	43.23	**57.82**	49.93	**35.68**	26.79
V2X-ViT	22.86	40.76	55.74	48.70	33.26	27.94
CoBEVT	**24.63**	**45.76**	54.73	**51.18**	34.79	27.25
Where2com	24.16	42.63	55.69	50.82	33.76	**26.42**

Most importantly, U2UData+ 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.

Dataset Visualization. U2UData+ dataset is the first large-scale swarm UAV autonomous flight dataset for the ELH 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 ELH tasks through collaborative perception, communication, and task re-allocation. These visualizations highlight the dataset's ability to design algorithms for ELH tasks.

U2UData+ Benchmark

SOTA Algorithms. Since the algorithms of swarm UAV autonomous flight for ELH tasks are still lacking, we provide a swarm UAV collaborative tracking benchmark for ELH tasks in the U2UData+ dataset. This benchmark uses 9 SOAT collaborative tracking algorithms, including No Fusion, Late Fusion, Early Fusion, When2Com (Liu et al. 2020), DiscoNet (Li et al. 2021), V2VNet (Wang et al. 2020), V2X-ViT (Xu et al. 2022b), CoBEVT (Xu et al. 2022a), and Where2com (Hu et al. 2022). This benchmark will be updated dynamically afterwards.

Evaluation metrics. We utilize the same evaluation metrics as outlined in (Weng et al. 2020) for object tracking. These metrics include: AMOTA, average multiobject tracking accuracy; AMOTP, average multiobject tracking precision; sAMOTA, scaled average multiobject tracking accuracy, which ensures a more linear representation across the entire [0, 1] range of significantly challenging tracking tasks; MOTA, multi object tracking accuracy; MT, mostly tracked trajectories; ML, mostly lost trajectories.

Tracker. We've chosen the AB3Dmot tracker (Weng et al. 2020) as our basic module of all SOAT algorithms. This tracker initially retrieves 3D object detections from a LiDAR point cloud. It subsequently integrates the 3D Kalman filter with the birth and death memory technique to guarantee efficient and resilient tracking performance. It attains state-of-the-art performance while maintaining the fastest speed.

Implementation details. We designate No Fusion as our baseline. To ensure a fair comparison, all models utilize PointPillar as the backbone for LiDAR feature extraction and use 32x feature compression (decompression) to save bandwidth. Among them, for CoBEVT, we only use the FuseBEVT module for feature aggregation without the SimBEVT module. During the training phase, we randomly designate one UAV as the ego UAV and train each model until achieving optimal task performance. During testing, we evaluate all compared models using a fixed ego UAV. For the tracking task, we utilize the previous three frames along with the current frame as inputs.

Results. As shown in Table 6, compared to the No Fusion method, AB3Dmot combined with cooperative algorithms significantly improves the tracking performance by at least 35.97% AMOTA and 32.88% sAMOTA. Compared with the Late Fusion method, the Intermediate Fusion method can improve the tracking performance by up to 27.48% AMOTA. Compared with the Early Fusion method, the Intermediate Fusion method can improve the tracking performance up to 8.33% AMOTA.

Conclusion

Swarm UAV autonomous flight for ELH tasks is crucial. U2UData+ is the first large-scale swarm UAV autonomous flight dataset for ELH tasks and the first scalable swarm UAV data online collection and algorithm closed-loop verification platform. The dataset is captured by 15 UAVs in autonomous collaborative flight for ELH tasks, comprising 12 scenes, 720 traces, 120 hours, and 600 seconds per trajectory. The platform supports the customization of simulators, UAVs, sensors, flight algorithms, formation modes, and ELH tasks. Through a visual control window, this platform allows users to collect customized datasets through one-click deployment online and to verify algorithms by closed-loop simulation. U2UData+ also provides a benchmark with 9 SOTA models. In the future, we hope U2UData+ can assist swarm UAV algorithms in being deployed in the real world.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 62222209), Beijing National Research Center for Information Science and Technology under Grant No. BNR2023TD03006, China Postdoctoral Science Foundation under Grant No. 2024M751688, Postdoctoral Fellowship Program of CPSF under Grant No. GZC20240827, Open Research Fund Program of Beijing National Research Center for Information Science and Technology, and Beijing Key Lab of Networked Multimedia.

References

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

DigitalOcean. 2024. FlightGear. https://www.flightgear.org/.

Dosovitskiy, A.; Ros, G.; Codevilla, F.; Lopez, A.; and Koltun, V. 2017. CARLA: An open urban driving simulator. In *Conference on robot learning*, 1–16. PMLR.

FEISILAB. 2025. RflySim. https://rflysim.com/doc/zh/.

Feng, T.; Li, Q.; Wang, X.; Wang, M.; Li, G.; and Zhu, W. 2024a. Multi-weather cross-view geo-localization using denoising diffusion models. In *Proceedings of the 2nd Workshop on UAVs in Multimedia: Capturing the World from a New Perspective*, 35–39.

Feng, T.; Wang, X.; Han, F.; Zhang, L.; and Zhu, W. 2024b. U2UData: A Large-scale Cooperative Perception Dataset for Swarm UAVs Autonomous Flight. In *ACM Multimedia 2024*.

Feng, T.; Wang, X.; Jiang, Y.-G.; and Zhu, W. 2025a. Embodied AI: From LLMs to World Models. In *IEEE Circuits and Systems Magazine*.

Feng, T.; Wang, X.; Zhou, Z.; Wang, R.; Zhan, Y.; Li, G.; Li, Q.; and Zhu, W. 2025b. EvoAgent: Agent Autonomous Evolution with Continual World Model for Long-Horizon Tasks. *arXiv preprint arXiv:2502.05907*.

Han, F.; Zhang, L.; Wang, X.; Zhao, K.-A.; Zhong, Y.; Su, Z.; Feng, T.; and Zhu, W. 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.

Hu, Y.; Fang, S.; Lei, Z.; Zhong, Y.; and Chen, S. 2022. Where2comm: Communication-efficient collaborative perception via spatial confidence maps. *Advances in neural information processing systems*, 35: 4874–4886.

Hu, Y.; Lu, Y.; Xu, R.; Xie, W.; Chen, S.; and Wang, Y. 2023. Collaboration helps camera overtake lidar in 3d detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9243–9252.

Li, E.; Wang, S.; Li, C.; Li, D.; Wu, X.; and Hao, Q. 2020. SUSTech POINTS: A Portable 3D Point Cloud Interactive Annotation Platform System. In *2020 IEEE Intelligent Vehicles Symposium (IV)*, 1108–1115.

Li, Y.; Cao, Y.; He, H.; Cheng, Q.; Fu, X.; Xiao, X.; Wang, T.; and Tang, R. 2025. M²IV: Towards Efficient and Fine-grained Multimodal In-Context Learning via Representation Engineering. In *Second Conference on Language Modeling*.

Li, Y.; Ren, S.; Wu, P.; Chen, S.; Feng, C.; and Zhang, W. 2021. Learning distilled collaboration graph for multi-agent perception. *Advances in Neural Information Processing Systems*, 34: 29541–29552.

Liu, H.; Huang, X.; Gu, J.; Shi, J.; He, N.; and Feng, T. 2025. TCDformer-based Momentum Transfer Model for Long-term Sports Prediction. *Expert Systems with Applications*, 128310.

Liu, M.; Yurtsever, E.; Fossaert, J.; Zhou, X.; Zimmer, W.; Cui, Y.; Zagar, B. L.; and Knoll, A. C. 2024. A Survey on Autonomous Driving Datasets: Statistics, Annotation Quality, and a Future Outlook. In *IEEE Transactions on Intelligent Vehicles*.

Liu, Y.-C.; Tian, J.; Glaser, N.; and Kira, Z. 2020. When2com: Multi-agent perception via communication graph grouping. In *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition*, 4106–4115.

Lu, Y.; Hu, Y.; Zhong, Y.; Wang, D.; Chen, S.; and Wang, Y. 2024. An Extensible Framework for Open Heterogeneous Collaborative Perception. *The Twelfth International Conference on Learning Representations*.

NVIDIA. 2025. Isaac Sim. https://developer.nvidia.com/isaac-sim.

Pan, Y.; Li, L.; Qin, J.; Chen, J.-J.; and Gardoni, P. 2024. Unmanned aerial vehicle–human collaboration route planning for intelligent infrastructure inspection. *Computer-Aided Civil and Infrastructure Engineering*, 39(14): 2074–2104.

PX4. 2025. Jmavsim. https://docs.px4.io/main/en/sim_jmavsim/.

Research, L. 2025. XPlan. https://www.x-plane.com/.

Robotics, O. 2025. Gazebo. https://gazebosim.org/home.

Shah, S.; Dey, D.; Lovett, C.; and Kapoor, A. 2018. Airsim: High-fidelity visual and physical simulation for autonomous vehicles. In *Field and Service Robotics: Results of the 11th International Conference*, 621–635.

Shen, Y.; Liu, H.; Liu, P.; Xia, R.; Yao, T.; Sun, Y.; and Feng, T. 2025. DETACH: Cross-domain Learning for Long-Horizon Tasks via Mixture of Disentangled Experts. *arXiv preprint arXiv:2508.07842*.

Sun, D.; Cheng, L.; Chen, S.; Li, C.; Xiao, Y.; and Luo, B. 2023. UAV-Ground Visual Tracking: A Unified Dataset and Collaborative Learning Approach. *IEEE Transactions on Circuits and Systems for Video Technology*.

Sun, G.; He, L.; Sun, Z.; Wu, Q.; Liang, S.; Li, J.; Niyato, D.; and Leung, V. C. 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): 8582–8600.

Wang, R.; Wang, X.; Feng, T.; Gong, X.; Li, G.; Zhan, Y.-W.; Li, Q.; and Zhu, W. 2025. Improving Compositional Generalization in Cross-Embodiment Learning via Mixture

of Disentangled Prototypes. In *Proceedings of the 33rd ACM International Conference on Multimedia*, 7162–7171.

Wang, T.-H.; Manivasagam, S.; Liang, M.; Yang, B.; Zeng, W.; and Urtasun, R. 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.

Weng, X.; Wang, J.; Held, D.; and Kitani, K. 2020. 3d multi-object tracking: A baseline and new evaluation metrics. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 10359–10366.

Xu, R.; Tu, Z.; Xiang, H.; Shao, W.; Zhou, B.; and Ma, J. 2022a. Cobevt: Cooperative bird's eye view semantic segmentation with sparse transformers. *arXiv preprint arXiv:2207.02202*.

Xu, R.; Xiang, H.; Tu, Z.; Xia, X.; Yang, M.-H.; and Ma, J. 2022b. V2x-vit: Vehicle-to-everything cooperative perception with vision transformer. In *European conference on computer vision*, 107–124.

Zhang, K.; Yang, T.; Ding, Z.; Yang, S.; Ma, T.; Li, M.; Xu, C.; and Gao, F. 2022. The visual-inertial-dynamical multirotor dataset. In *2022 International Conference on Robotics and Automation (ICRA)*, 7635–7641.

Zhao, B.; Zhang, W.; and Zou, Z. 2023. BM2CP: Efficient Collaborative Perception with LiDAR-Camera Modalities. *arXiv preprint arXiv:2310.14702*.