







S3E: A Multi-Robot Multimodal Dataset for Collaborative SLAM

Dapeng Feng , Yuhua Qi , Shipeng Zhong , Zhiqiang Chen, Qiming Chen, Hongbo Chen ,
Jin Wu , *Member, IEEE*, and Jun Ma , *Member, IEEE*

Abstract—The burgeoning demand for collaborative robotic systems to execute complex tasks collectively has intensified the research community’s focus on advancing simultaneous localization and mapping (SLAM) in a cooperative context. Despite this interest, the scalability and diversity of existing datasets for collaborative trajectories remain limited, especially in scenarios with constrained perspectives where the generalization capabilities of Collaborative SLAM (C-SLAM) are critical for the feasibility of multi-agent missions. Addressing this gap, we introduce S3E, an expansive multimodal dataset. Captured by a fleet of unmanned ground vehicles traversing four distinct collaborative trajectory paradigms, S3E encompasses 13 outdoor and 5 indoor sequences. These sequences feature meticulously synchronized and spatially calibrated data streams, including 360-degree LiDAR point cloud, high-resolution stereo imagery, high-frequency inertial measurement units (IMU), and Ultra-wideband (UWB) relative observations. Our dataset not only surpasses previous efforts in scale, scene diversity, and data intricacy but also provides a thorough analysis and benchmarks for both collaborative and individual SLAM methodologies.

Index Terms—Multi-robot SLAM, data sets for SLAM, SLAM.

I. INTRODUCTION

COLLABORATIVE Simultaneous Localization and Mapping (C-SLAM) is crucial for multi-agent cooperation, improving the robustness and efficiency of localization and mapping in shared spaces [1], [2], [3], [4]. Despite progress, repeatability and benchmarking in C-SLAM research face challenges due to: 1) **System Complexity**: The complexity of C-SLAM systems, which integrate intricate software architectures and a variety of hardware, complicates the process of achieving perfect replication, thereby emphasizing the need for publicly available datasets to facilitate methodological evaluation.

Received 3 September 2024; accepted 24 October 2024. Date of publication 1 November 2024; date of current version 11 November 2024. This article was recommended for publication by Associate Editor Brendan Englot and Editor Javier Civera upon evaluation of the reviewers’ comments. (*Corresponding author: Yuhua Qi.*)

Dapeng Feng, Yuhua Qi, Shipeng Zhong, Zhiqiang Chen, and Hongbo Chen are with Sun Yat-sen University, Guangzhou 510006, China (e-mail: fen gdp5@mail2.sysu.edu.cn; qiyh8@mail.sysu.edu.cn; zhongshp5@mail2.sysu.edu.cn; chenzhq56@mail2.sysu.edu.cn; chen hongbo@mail.sysu.edu.cn).

Qiming Chen is with YUNDRONE Technology Company, Ltd., Guangzhou 510642, China (e-mail: chenqiming629@gmail.com).

Jin Wu is with The Hong Kong University of Science and Technology, Hong Kong SAR, China (e-mail: jin_wu_uestc@hotmail.com).

Jun Ma is with the The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China (e-mail: jun.ma@ust.hk).

For access to the dataset and the latest information, please visit our repository at <https://pengyu-team.github.io/S3E>.

This letter has supplementary downloadable material available at <https://doi.org/10.1109/LRA.2024.3490402>, provided by the authors.

Digital Object Identifier 10.1109/LRA.2024.3490402

2) **Evaluation Methodology**: Traditional approaches, which divide single-agent SLAM datasets into segments for virtual agents, are impractical for real-world scenarios due to uniform perspectives and lighting at intersection points [5]. These limitations underscore the importance of developing more effective evaluation strategies and datasets to advance C-SLAM research.

Furthermore, effective trajectory design for multi-robot scenarios in C-SLAM must follow two key principles: 1) **Temporal and Spatial Diversity**: Trajectories should be crafted to provide multiple intra- and inter-loop closures at different times and places along the trajectories, ensuring robots have complementary perspectives. 2) **Communication Constraints**: Robots are typically limited to sharing information within close proximity, necessitating trajectory designs that maintain a reasonable interaction distance for effective communication. In this letter, we introduce four trajectory prototypes designed to meet these principles and evaluate the adaptability of C-SLAM methodologies across diverse closure strategies in multi-robot operations.

Publicly available datasets are crucial for the SLAM research community for two main reasons: 1) **Development Acceleration**: Conducting specialized SLAM experiments involves significant investment in advanced hardware and complex software, including calibration and ground truth generation. Datasets aid in streamlining this development process. 2) **Benchmarking and Analysis**: Datasets offer a standardized framework for comparing and assessing SLAM algorithms’ performance. While various SLAM datasets [6], [7] have been introduced over the years, contributing to sensor diversity and scenario breadth, most focus on single-agent systems. This focus may limit their utility for multi-agent scenarios, highlighting the need for more datasets that address multi-agent challenges.

The scarcity of datasets designed for C-SLAM is primarily due to the increasing challenges in data acquisition with more collaborating agents. Current C-SLAM datasets [8], [9], [10], [11], [12], [13] often focus on a single cooperative approach with significant spatial overlap, neglecting areas with minimal overlap. To fill this gap and enhance C-SLAM research, we introduce S3E dataset, offering a multimodal perspective with a variety of cooperative trajectory patterns in both outdoor and indoor environments.

Based on the S3E dataset, we conducted extensive experiments with cutting-edge SLAM methods across various applications, from single-robot to collaborative scenarios, using diverse sensor data. The findings reveal that current C-SLAM techniques struggle with inter-loop closures in some sequences, indicating a

need for further research to enhance the robustness and reliability of these systems.

In conclusion, our work makes several key contributions to the field:

- We have created a cutting-edge C-SLAM dataset using three ground robots, each equipped with a 16-beam 3D laser scanner, two high-resolution color cameras, a 9-axis IMU, a UWB receiver, and a dual-antenna RTK receiver. This dataset is the first to incorporate UWB relative distance measurements, providing a new research dimension.
- To assess C-SLAM's performance in environments with limited overlap, we have captured extensive long-term sequences using four meticulously designed trajectory paradigms that address various intra- and inter-robot loop closure scenarios.
- We have conducted a thorough and multifaceted evaluation of current C-SLAM techniques, employing both quantitative metrics and qualitative analyses to assess their performance and utility.

II. RELATED WORK

Single-Agent SLAM Datasets: High-quality, extensive datasets are crucial for advancing SLAM research. The KITTI dataset [6], known for its comprehensive urban data, has been a significant benchmark for vision and LiDAR-based SLAM techniques. However, it has limitations, such as the absence of perfectly synchronized data, offering only estimated time delays between sensors. To address this, the EuRoC dataset [7] was introduced, offering synchronized visual and inertial data from a micro aerial vehicle in controlled settings, using hardware clocks for synchronization.

Multi-Agent SLAM Datasets: The UTIAS dataset [8] was a pioneering work with five robots exploring a small area. The AirMuseum dataset [10] introduced a heterogeneous multi-robot setup in a warehouse, and the FordAV dataset [9] is collected by a fleet of autonomous vehicles, providing a multi-agent perspective with seasonal variations in weather, lighting, construction, and traffic conditions. However, these datasets often have significant co-view overlap, limiting comprehensive C-SLAM evaluation.

The GRACO dataset [12] attempted to address this by introducing a ground-aerial approach, but it had synchronization issues and did not fully handle dynamic environments. The SubT-MRS dataset [13] added more variety with multi-robot SLAM data across different environments. Kimera-Multi [11] advanced the state with a comprehensive collection featuring diverse trajectory designs and multiple robots.

It's important to note that existing datasets, including GRACO, SubT-MRS, and Kimera-Multi, often overlook the complexities of collaborative trajectory patterns. The ability to handle intra- and inter-loop closures is crucial for C-SLAM robustness and is not consistently addressed. Considering these collaborative patterns is essential for developing C-SLAM algorithms that can operate effectively in environments with varying agent interactions and dynamic scenes.

In this letter, we introduce a large-scale C-SLAM dataset that includes visual ambiguities and dynamic objects, captured under four distinct trajectory paradigms to assess C-SLAM generalizability under restricted overlap conditions. Our dataset also incorporates UWB relative measurement positioning, adding a significant dimension for C-SLAM research. For a comprehensive comparison of different datasets, including our contribution, refer to Table I.

III. S3E DATASET

The S3E dataset is meticulously assembled with a focus on high temporal precision in sensor data synchronization. Each sensor is calibrated to a shared timescale for harmonized multi-sensory information, facilitated by an advanced time synchronization mechanism. This is crucial for accurate data integration across different modalities. The dataset features two mobile robot platform versions:

- S3Ev1.0: Designed for indoor use with a compact design for exceptional maneuverability in tight spaces.
- S3Ev2.0: Enhanced with a wider frame to accommodate a high-precision ground truth system and an UWB module for improved localization accuracy in challenging environments.

For a visual representation of sensor placement, refer to Fig. 1

A. Sensor Configuration

Our S3E dataset encompasses a multimodal array of sensors, each selected for its operational range and noise characteristics, and meticulously synchronized to capture data with high temporal precision. The sensors are integrated onto the *Agilex Scout Mini*, a versatile all-terrain mobile platform capable of high-speed remote-controlled navigation, featuring four-wheel drive and a maximum speed of 10 km/h.

The technical specifications of the sensors and ground truth devices integrated into our payload are outlined on our project website.¹ This includes the sensor types, their resolution, measurement range, accuracy, and any other pertinent technical details that define their contribution to the SLAM system's performance. In particular, we utilize the centimeter-accuracy GNSS and RTK for outdoor environments, achieving a precision of ± 1 cm, to record the ground truth trajectories. For indoor environments, where GNSS signals are typically unavailable, we utilize a millimeter-precision motion capture system to obtain our ground truth data.

B. Sensor Synchronization

This section delves into the critical processes of time synchronization and sensor calibration, which are essential for achieving optimal sensor fusion and maximizing system performance within a dynamic, multimodal environment.

1) *Time Synchronization:* Time synchronization across all sensors is essential for the accurate co-registration of multi-sensor data, particularly in dynamic environments where precise

¹<https://pengyu-team.github.io/S3E>

TABLE I

COMPARATIVE ANALYSIS OF PROMINENT SLAM DATASETS. THE ABBREVIATIONS USED WITHIN THE TABLE ARE AS FOLLOWS: “SW” DENOTES SOFTWARE SYNCHRONIZATION, INDICATING THAT SYNCHRONIZATION ACROSS SENSORS OR SYSTEMS IS ACHIEVED THROUGH SOFTWARE MECHANISMS; “HW” REPRESENTS HARDWARE SYNCHRONIZATION, WHERE SYNCHRONIZATION IS MANAGED BY HARDWARE MEANS

Dataset	Platform	Sensors				Time Sync.		Trajectory		Environment		Ground Truth
		Camera	IMU	LiDAR	UWB	Intra-	Inter-	Overlap	Paradigm	Indoor	Outdoor	
KITTI [6]	Car	✓	✓	✓	✗	Sw	✗	-	-	✗	✓	GNSS/INS
EuRoC [7]	UAV	✓	✓	✗	✗	Hw	✗	-	-	✓	✗	Motion Capture
UTIAS [8]	5 UGVs	✓	✗	✗	✗	Sw	NTP	Large	1	✓	✗	Motion Capture
AirMuseum [10]	3 UGVs, 1 UAV	✓	✓	✗	✗	Sw	NTP	Large	1	✓	✗	SfM
FordAV [9]	3 Cars	✓	✓	✓	✗	-	GNSS	Large	1	✗	✓	GNSS/INS
GRACO [12]	1UGV, 1UAV	✓	✓	✓	✗	Hw	-	Restricted	1	✗	✓	GNSS/INS
Kimera-Multi [11]	8 UGVs	✓	✓	✓	✗	-	-	Restricted	1	✗	✓	GPS and total-station assisted LiDAR SLAM
SubT-MRS [13]	UGVs, UAVs, Handheld	✓	✓	✓	✗	Hw	-	Restricted	1	✓	✓	3D Scanner
Ours	3 UGVs	✓	✓	✓	✓	Hw	GNSS, PTPv2	Restricted	4	✓	✓	RTK, GNSS/INS Motion Capture

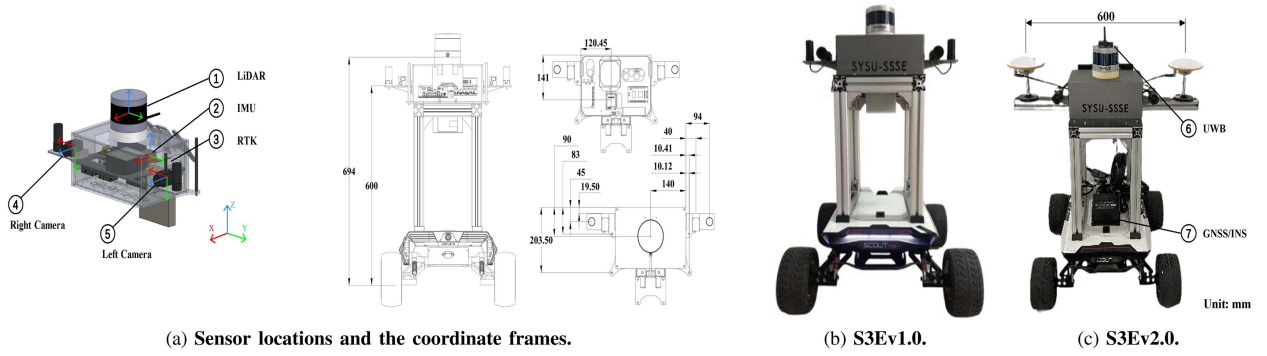


Fig. 1. **Mobile Platform Sensor Layout and Coordinate Systems.** The left part details the sensor locations and the coordinate frames that define their spatial orientation relative to the platform. In the right part, our mobile platforms are available in two versions, each designed for different operational requirements.

timing is critical to the quality of the fused dataset. This synchronization ensures that data from various sensors is temporally aligned, facilitating coherent integration and analysis.

Our synchronization system is built around an *Altera EP4CE10* FPGA board acting as the primary trigger device, with an *Intel NUC11TNKv7* serving as the host computer. The system is designed with a comprehensive set of I/O interfaces to accommodate the diverse requirements of different sensors. For synchronization across agents, we address two distinct scenarios:

- In outdoor settings with access to Global Navigation Satellite System (GNSS) signals, we use GNSS time as the global reference to synchronize the timing across agents.
- In GNSS-denied environments such as indoors and tunnels, agents synchronize their timers by obtaining external global time data from Alpha, which serves as a Precision Time Protocol version 2 (PTPv2) server, via a wireless network connection.

Considering transmission delays, all sensor readings are forwarded to the host computer, where they are timestamped upon arrival, organized, and packaged to ensure accurate temporal referencing for subsequent analysis.

2) *Sensor Calibration:* Our sensor suite is governed by a unified coordinate system framework, adhering to the right-hand rule, which ensures uniformity in data orientation and facilitates standardized analysis.

The calibration process was meticulously executed through five distinct stages: camera-intrinsic, IMU-intrinsic, camera-camera extrinsic, camera-IMU extrinsic, and camera-lidar extrinsic. Each stage was crucial for ensuring the accurate alignment and integration of sensor data.

Upon completion of these calibration stages, the resulting parameters were meticulously stored in YAML files for each agent, ensuring that the calibrated parameters are readily accessible and reproducible.

C. Ground Truth

To create accurate ground truth tracks for our dataset, we use a three-pronged approach:

For outdoor environments with good GNSS signal reception, a dual-antenna RTK device is used to achieve highly accurate localization data with centimeter-level precision. The S3Ev2.0 platform features a GNSS/INS unit that significantly enhances ground truth instrumentation. This integration allows for high-frequency positioning data output even during GNSS signal outages, such as in tunnels. The system ensures continuous and reliable ground truth data by fusing RTK measurements with IMU data using post-processing fusion techniques.

In indoor environments without GNSS signals, like laboratories, a motion capture system with 17 high-frequency cameras is used to record track start and endpoints due to the high costs

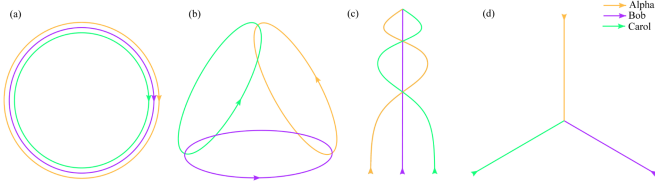


Fig. 2. **Illustration of Trajectory Paradigms for C-SLAM.** The distinct trajectory paradigms adopted by three agents, designated as Alpha, Bob, and Carol, to demonstrate different interaction and information exchange patterns in a multi-agent SLAM context. (a) **Concentric Circles.** (b) **Intersecting Circles.** (c) **Intersection Curve.** (d) **Rays.**

of a large-scale system. In other indoor areas lacking a motion capture system, such as teaching buildings, the RTK device is used to record the initial and terminal points of the tracks outside the buildings.

By combining these methods, we are able to generate comprehensive ground truth positions that cater to both outdoor and indoor environments, thereby facilitating a robust evaluation of C-SLAM algorithms across diverse real-world scenarios.

D. Trajectory Paradigms

The S3E dataset incorporates a diverse set of trajectory paradigms to simulate various collaborative scenarios that agents may encounter during SLAM operations. As depicted in Fig. 2, we have designed these paradigms to address different loop closure conditions, both within individual agents (intra-loop) and across multiple agents (inter-loop). The following are the four trajectory paradigms we have included in our dataset:

Concentric Circles: This paradigm involves coordinated unmanned platforms that stay connected and collect similar environmental data. This allows for efficient area coverage and tasks like environmental monitoring and mapping. Real-time data sharing improves accuracy, but the similarity of observations can make it challenging to detect and adapt to rapid environmental changes.

Intersecting Circles: In this paradigm, unmanned platforms work separately in different areas and only share data when they intersect. This approach reduces the need for constant communication and coordination but may limit the consistency of overall mapping and positioning due to less frequent data exchange.

Intersection Curve: This paradigm involves unmanned platforms starting from different locations, meeting at specific points, and eventually converging at a final destination. They communicate and share data only during these meetings, which helps in large-scale exploration and mapping by reducing cumulative errors. Each platform's independent operation boosts efficiency and coverage. However, this approach requires advanced navigation for planning effective meeting points and routes.

Rays: In this paradigm, autonomous platforms begin at various starting points, work independently, and come together at a common endpoint for communication and data exchange. This method is effective for large-scale, independent operations like disaster recovery and mining exploration, where each platform operates on its own. The final convergence helps in sharing

TABLE II
COMPARATIVE ANALYSIS OF TRAJECTORY PARADIGMS FOR LOOP CLOSURE DETECTION IN C-SLAM. THE SYMBOLS WITHIN THE TABLE INDICATE THE PRESENCE (●) OR ABSENCE (○) OF LOOP CLOSURES FOR EACH PARADIGM, PROVIDING A QUICK REFERENCE FOR THE SUITABILITY OF EACH STRATEGY IN DIFFERENT COLLABORATIVE ROBOTIC MISSIONS

Trajectory Paradigm	Closures		Applications
	Intra-loop	Inter-loop	
Concentric Circles	●	●	Extensive coverage and detailed exploration
Intersecting Circles	●	●	Distributed search and rescue operations
Intersection Curve	○	●	Large-scale distributed exploration, patrol, and mapping
Rays	○	○	Independent exploration and mapping

historical data and optimizing local maps, but the limited interaction can affect the overall consistency of mapping and positioning, thus requiring advanced autonomous navigation and decision-making skills.

Each paradigm is meticulously crafted to offer a robust framework for assessing C-SLAM algorithms across a multitude of real-world collaborative robotic applications. Table II provides a comparative analysis of the loop closure detection situations corresponding to these paradigms, highlighting their respective advantages and potential application areas.

E. Dataset Analysis

To test collaborative mission scenarios and trajectory paradigms, we conducted data collection at Sun Yat-sen University's Guangzhou East Campus using three tele-operated robots under human control for safety. The dataset, detailed in Table III, includes diverse environments like squares and libraries, with 13 outdoor and 5 indoor sequences. It totals over 263.2GB, 475.2 minutes of footage, and covers more than 28.1 km. With an average sequence length of 527.9 seconds, it's suitable for long-term C-SLAM technology evaluations.

To thoroughly assess the accuracy and robustness of C-SLAM algorithms in complex, real-world scenarios, our dataset encompasses a diverse range of environments, each presenting unique challenges:

Dormitory: High pedestrian and bicycle traffic test perception and tracking. Regular layouts are ideal for assessing C-SLAM precision and consistency across platforms.

Campus Road: Long distances and expansive views evaluate endurance and large-scale exploration. Long-distance and multi-cycle data assess stability, accuracy, and efficiency over extended operations.

Playground: Open spaces with fewer obstructions challenge feature extraction and optimization. Data collected at different times of day test adaptability to lighting conditions and performance under rapid motion.

Laboratory: Confined spaces with complex layouts and rich semantic content test advanced scene understanding and mapping capabilities.

Teaching Building and Tunnel: Poor lighting and similar geometric structures challenge robustness in maintaining accurate positioning and mapping, especially in tunnels and corridors.

Our dataset features unique areas like libraries and squares, adding environmental variety and perceptual challenges. Fig. 3

TABLE III
ANALYSIS OF S3E DATASET

Scenario		Time[s]	Trajectory				Ground Truth	Length[m]			Size[GB]	Sensors			
Env.	Region		a	b	c	d		Alpha (α)	Bob (β)	Carol (γ)		Camera	IMU	LiDAR	UWB
Outdoor	v1.0	Square_1	460		✓		RTK	546.0	496.5	529.2	17.8	✓	✓	✓	✗
		Square_2	255			✓	RTK	-	250.6	246.4	9.4	✓	✓	✓	✗
		Library_1	454	✓			RTK	507.6	517.2	498.9	16.3	✓	✓	✓	✗
		Campus_Road_1	878		✓		RTK	920.5	995.9	1072.3	29.4	✓	✓	✓	✗
		Playground_1	298	✓			RTK	407.7	425.6	445.5	8.7	✓	✓	✓	✗
		Playground_2	222			✓	RTK	265.6	315.7	456.4	6.3	✓	✓	✓	✗
		Dormitory_1	671			✓	RTK	727.0	719.3	721.9	23.5	✓	✓	✓	✗
	v2.0	Square_3	466			✓	GNSS/INS	487.4	569.6	563.8	8.5	✗	✓	✓	✓
		Library_2	491	✓			GNSS/INS	521.6	523.9	520.3	9.3	✗	✓	✓	✓
		Campus_Road_2	1594	✓			GNSS/INS	1938.6	1934.2	1950.1	30.9	✗	✓	✓	✓
		Campus_Road_3	907	✓			GNSS/INS	983.3	967.6	986.9	17.5	✗	✓	✓	✓
		Playground_3	111		✓		GNSS/INS	84.7	91.3	110.7	1.8	✗	✓	✓	✓
		Tunnel_1	425	✓			GNSS/INS	521.9	502.4	501.1	8.2	✗	✓	✓	✓
	Indoor	Teaching_Building_1	798		✓		RTK	617.2	734.4	643.4	27.3	✓	✓	✓	✗
		Laboratory_1	292		✓		Motion Capture	147.7	161.5	141.0	9.6	✓	✓	✓	✗
		Laboratory_2	391		✓		Motion Capture	215.3	199.1	160.3	12.7	✓	✓	✓	✗
		Laboratory_3	410		✓		Motion Capture	219.1	202.2	204.2	13.3	✓	✓	✓	✗
		Laboratory_4	380	✓			Motion Capture	173.7	177.2	180.0	12.7	✓	✓	✓	✗

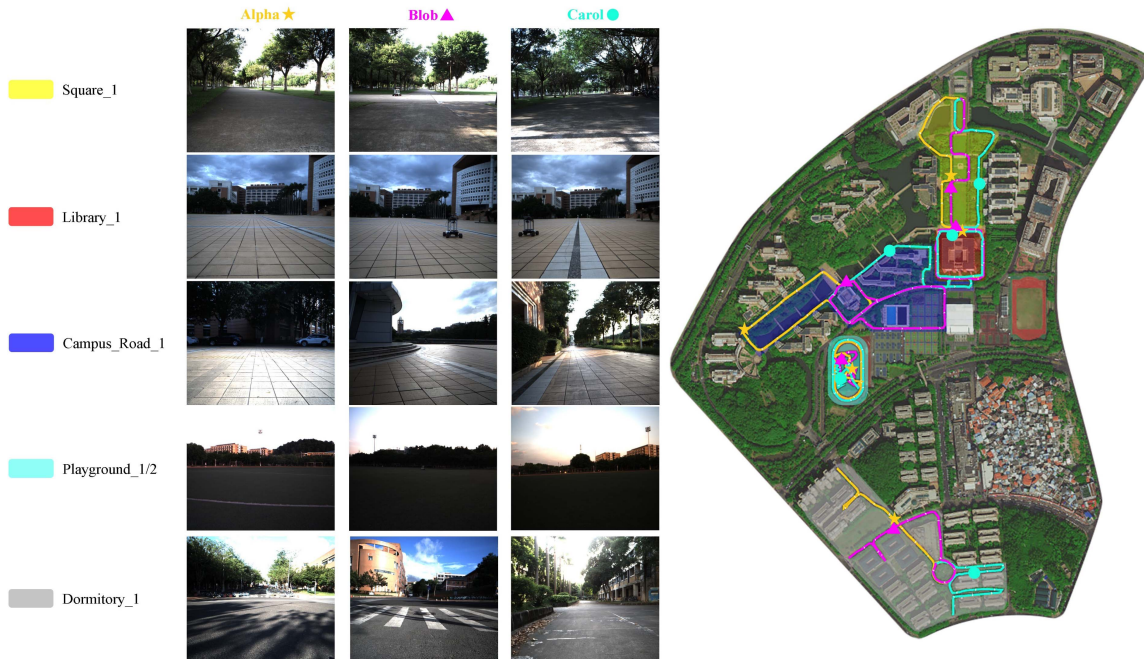


Fig. 3. **Visualization of Outdoor Trajectory in the S3E Dataset.** The outdoor trajectories captured in the S3E dataset by three tele-operated mobile platforms, designated as Alpha (α), Bob (β), and Carol (γ). The trajectories are distinctly annotated with Orange, Purple, and Cyan. The annotations \star , \triangle , and \bullet indicate the specific positions where data was recorded in each sequence. The left portion of the figure presents the synchronized data capture at the annotated points, demonstrating the collaborative data collection process.

shows the trajectory paradigms in outdoor settings within the S3E dataset. The dataset covers a range of challenging environments that C-SLAM algorithms may encounter, including dynamic objects, long operation times, perceptual aliasing, indoor settings, and significant motion. This diversity is crucial for evaluating C-SLAM performance, adaptability, and robustness, which are key for advancing collaborative robotic navigation and mapping technologies.

F. Dataset Format

Our research utilizes the ROS2 [14] bag format for sensor data storage, a standard in robotics known for efficient data management and playback. To simplify data replay and ensure

easy access to synchronized sensor streams from our Alpha (α), Blob (β), and Carol (γ) robots, we've implemented a strategy to merge data from the same operational sequence into one ROS2 bag file.

Data Organization: Our dataset, as shown in Fig. 4, is meticulously organized with a clear directory structure. Calibration parameters are detailed in YAML files within a specific Calibration folder. Each data sequence is complete with a `.db3` file for primary sensor measurements and a `meta-data.yaml` file outlining sequence specifics like sensor configurations and capture conditions. Additionally, to aid in thorough performance assessments, we've included auxiliary files like `<agent_id>_gt.txt`, which contain ground truth data.

TABLE IV
DESCRIPTORS OF ROS2 TOPICS INCLUDED IN OUR DATASET. NOTE THAT **AGENT_ID** STANDS FOR ALPHA, BLOB, AND CAROL

Sensor	Topic	Type	Description	Frequency
LiDAR	/agent_id/velodyne_points	sensor_msgs/msg/PointCloud2	Raw velodyne pointcloud with $[x, y, z, intensity, ring, time]$	10 Hz
Camera ¹	/agent_id/left_camera/compressed	sensor_msgs/msg/CompressedImage	Compressed RGB8 image with 1224×1024	10 Hz
	/agent_id/right_camera/compressed	sensor_msgs/msg/CompressedImage	Compressed RGB8 image with 1224×1024	
IMU	/agent_id/imu/data	sensor_msgs/msg/Imu	Raw imu data	100 Hz
RTK ²	/agent_id/fix	sensor_msgs/msg/NavSatFix	RTK-GNSS data with $[latitude, longitude, altitude]$	1 Hz
UWB ³	/agent_id/nlink_linktrack_nodeframe2	std_msgs/msg/Float64MultiArray	UWB data comprises 13 data fields ⁴	50 Hz
	/agent_id/time_reference	sensor_msgs/msg/TimeReference	GPS time reference anchored at the epoch of 1980-01-06 00:00:00 UTC	
	/agent_id/fix	sensor_msgs/msg/NavSatFix	RTK-GNSS data with $[latitude, longitude, altitude]$	
GNSS/INS ³	/agent_id/correct_imu	sensor_msgs/msg/Imu	Raw build-in imu data	100 Hz
	/agent_id/vel	geometry_msgs/msg/TwistStamped	Timestamped linear velocities (m/s) of mobile platform along $[east, north, up]$	
	/agent_id/heating	geometry_msgs/msg/QuaternionStamped	Timestamped quaternions map mobile platform attitudes	

¹ Available in all sequences of S3Ev1.0. ² Available in outdoor sequences of S3Ev1.0. ³ Available in all sequences of S3Ev2.0.

⁴ Detailed definitions of UWB data fields are documented and accessible on our project website.

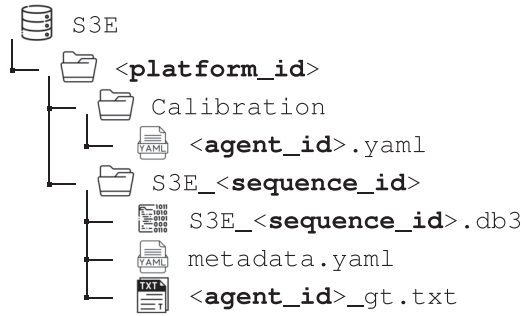


Fig. 4. S3E dataset organizational structure.

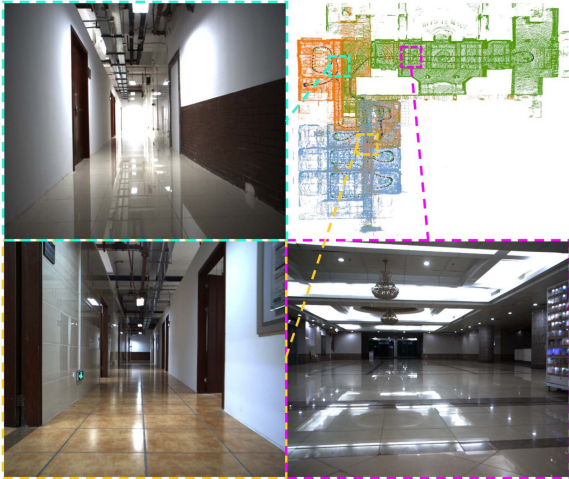


Fig. 5. Map in laboratory_1 with CoRLIO.

Ground Truth Format: The ground truth data, which is essential for evaluating the accuracy of C-SLAM algorithms, is provided as TXT files. These files contain timestamped poses in UTM coordinates and orientation quaternions, formatted as follows: $[timestamp, t_x, t_y, t_z, q_x, q_y, q_z, q_w]$. The position

$t_a \in \{x, y, z\}$ denotes the robot's location in UTM coordinates, while $q_a \in \{x, y, z, w\}$ represents the orientation in quaternion form.

Sensor Data Topics: Table IV outlines the sensor topics included in the dataset, detailing the type of data each topic contains, its description, and the frequency of data capture.

IV. EXPERIMENTS

A. Baselines

We have implemented four single-agent SLAM systems, namely ORB-SLAM3 [15], VINS-Fusion [16], LIO-SAM [17], and LVI-SAM [18]. Additionally, we have incorporated five collaborative SLAM (C-SLAM) systems into our study, which include COVINS [1], DiSCo-SLAM [2], Swarm-SLAM [19], DCL-SLAM [3], and CoLRIO [4]. These systems have been evaluated using the S3E dataset.

For most of the baselines, we only modify the intrinsic and extrinsic of the sensors and use the left camera for evaluation.

B. Results

Our experimental evaluation of the S3E dataset provides valuable insights into the performance of various state-of-the-art SLAM methodologies under diverse real-world conditions. The results, summarized in Tables V and VI, reveal the absolute trajectory error (ATE) for both single-agent and collaborative SLAM (C-SLAM) systems in outdoor and indoor environments without UWB measurement.

• Key Findings:

- **Inter-Loop Closures Detection Impact:** The results highlight the increased performance of C-SLAM systems when successful in detecting inter-loop closures, showcasing the benefits of collaborative information sharing among agents.
- **Front-end State Estimation Impact:** The performance of C-SLAM is heavily influenced by the accuracy of

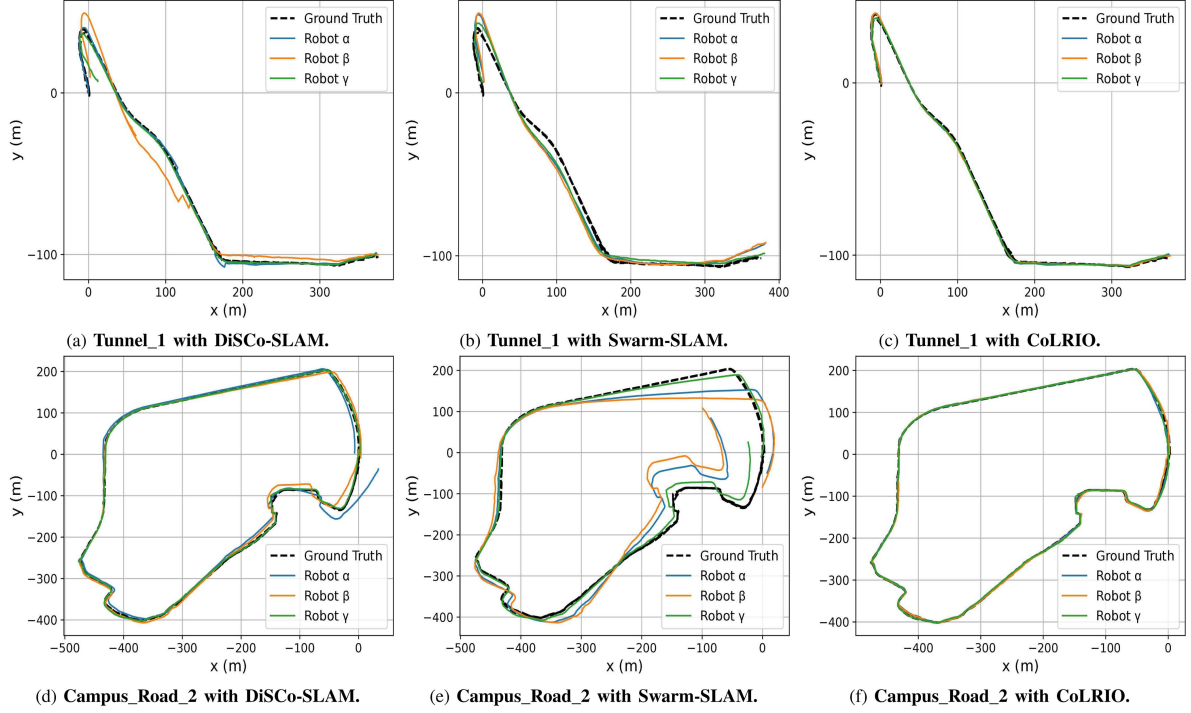


Fig. 6. The qualitative results of outdoor environments.

TABLE V

ATE [m] FOR SINGLE SLAM AND C-SLAM IN THE S3EV1.0 OUTDOOR ENVIRONMENT WITHOUT UWB MEASUREMENT. \times FAILS TO INITIALIZE OR TRACK FRAMES. IF INTER-LOOP CLOSURES DETECTION FAILS, WE MARK IT “FAILED”

Methods	Square_1				Square_2				Library_1				Campus_Road_1				Playground_1				Playground_2				Dormitory_1		
	α	β	γ		α	β	γ		α	β	γ		α	β	γ		α	β	γ		α	β	γ				
ORB-SLAM3	1.16	15.5	\times	-	2.81	\times	\times	-	11.4	\times	\times	\times	-	\times	55.5	\times	0.87	\times	\times	\times	3.29	\times	\times	\times	-	\times	\times
VINS-Fusion	1.81	\times	4.83	-	1.51	0.62	7.95	7.86	5.56	\times	16.5	\times	-	2.21	1.35	Failed	3.24	7.36	4.31	-	31.3	\times	\times	\times	6.67	3.97	7.70
LIO-SAM	1.19	1.75	\times	-	0.73	0.36	1.12	1.52	1.14	2.06	3.25	2.43	-	2.06	3.25	2.43	\times	\times	0.86	-	\times	\times	0.68	0.63	1.44	0.91	
LVI-SAM	1.21	0.88	\times	-	0.79	0.40	1.89	1.67	1.31	2.44	3.14	1.30	-	2.44	3.14	1.30	\times	1.59	0.76	-	6.78	6.10	0.72	0.86	1.48	0.94	
CoLRIO (front-end)	0.90	0.57	0.65	-	0.49	0.19	1.37	1.54	0.88	1.06	0.91	1.10	-	1.06	0.91	1.10	0.39	0.45	0.20	-	0.69	0.35	0.29	0.47	0.83	0.81	
COVINS	1.63	0.83	1.75	-	Failed	Failed	-	-	\times	18.3	9.61	54.7	-	2.21	1.35	Failed	0.27	0.37	0.38	-	\times	\times	-	0.54	1.47	\times	
DiSCo-SLAM	8.13	7.11	2.13	-	1.39	0.52	3.16	2.71	3.22	7.62	11.3	6.36	-	7.62	11.3	6.36	3.36	3.38	1.70	-	1.36	1.32	2.05	10.0	5.61	4.50	
Swarm-SLAM	1.23	0.86	0.66	-	0.50	0.26	0.58	1.26	1.17	1.51	1.48	1.87	-	1.51	1.48	1.87	0.33	0.40	0.32	-	\times	\times	-	0.52	1.37	0.51	
DCL-SLAM	0.91	0.53	0.80	-	0.56	0.20	1.06	1.05	1.02	1.08	0.79	1.08	-	1.08	0.79	1.08	0.39	0.60	0.16	-	0.44	0.32	0.29	0.54	0.87	0.84	

TABLE VI

ATE [m] FOR SINGLE SLAM AND C-SLAM IN THE S3EV1.0 INDOOR ENVIRONMENT WITHOUT UWB MEASUREMENT

Methods	Laboratory_1			Laboratory_2			Laboratory_3			Laboratory_4		
	α	β	γ	α	β	γ	α	β	γ	α	β	γ
LIO-SAM	0.47	1.82	2.51	12.2	6.12	8.88	6.24	8.57	7.31	1.96	0.88	0.65
FAST-LIO2	7.27	11.2	5.13	10.2	2.14	15.0	12.8	6.80	7.33	0.79	0.88	0.37
CoLRIO (front-end)	0.40	0.51	0.44	0.83	0.79	0.96	1.79	0.91	2.43	0.67	0.65	0.87
DiSCo-SLAM	0.56	0.43	0.33	2.02	0.36	4.60	2.01	0.31	1.28	1.40	2.02	3.45
Swarm-SLAM	9.86	1.14	3.29	4.56	7.03	15.3	7.06	12.2	8.10	1.79	3.40	3.47
CoLRIO	0.43	0.50	0.44	0.40	0.07	0.08	0.49	0.91	0.13	0.95	0.65	0.18

its front-end state estimation. For example, CoLRIO’s success in certain scenarios is attributed to its effective front-end tracker.

- *Specific Observations:*

- In scenarios with large spatial overlap, C-SLAM systems leveraged inter-robot measurements to enhance state estimation accuracy. However, in areas with limited overlap, reducing drift remained a challenge.

- The incorporation of UWB measurements in CoLRIO significantly improved localization robustness and accuracy, as demonstrated in Table VII, showcasing the benefits of additional ranging data for C-SLAM systems.

The results from our experiments on the S3E dataset underscore the importance of continuous innovation in C-SLAM algorithms. The varying performance across different scenarios and the impact of collaborative data sharing highlight the need for further research. The qualitative results presented in Figs. 5 and 6, visually complement the performance metrics, providing a deeper understanding of the C-SLAM systems’ behavior in diverse environments.

V. KNOWN ISSUES

The S3E dataset, while a substantial contribution to the field of C-SLAM, presents certain limitations. This section delves into

TABLE VII
ATE [m] FOR C-SLAM IN THE S3Ev2.0 OUTDOOR ENVIRONMENT WITH UWB MEASUREMENT

Methods	Square_3			Library_2			Campus_Road_2			Campus_Road_3			Playground_3			Tunnel_1		
	α	β	γ	α	β	γ	α	β	γ	α	β	γ	α	β	γ	α	β	γ
DiSCo-SLAM	1.15	1.26	12.11	1.07	1.02	0.98	15.41	14.60	3.37	1.27	1.32	39.76	9.01	0.72	2.32	5.06	11.90	4.40
Swarm-SLAM	8.52	5.64	2.80	3.15	0.91	0.94	48.96	64.23	22.37	8.79	9.60	4.90	1.11	3.24	5.47	10.49	11.06	10.06
CoLRIO (w/o UWB)	1.79	1.62	1.30	0.91	0.89	0.88	1.29	2.97	1.21	1.04	1.13	1.07	0.24	0.19	0.30	5.49	4.69	5.15
CoLRIO (w UWB)	1.71	1.52	1.54	0.89	0.87	0.88	1.37	1.59	1.34	0.77	0.61	0.90	0.23	0.20	0.30	4.01	3.21	3.60

the known issues associated with the dataset and the broader challenges of C-SLAM systems:

- 1) Scalability: C-SLAM systems must efficiently handle operations with a large number of robots without significant performance degradation.
- 2) Resource Management: There are strict limitations on communication and computation capabilities of mobile robots, which C-SLAM must overcome to achieve real-time performance.

VI. CONCLUSION

The S3E dataset provides a comprehensive and multifaceted platform for evaluating C-SLAM systems under a variety of real-world conditions. It features four distinct trajectory paradigms for broad scenario evaluation. Captured across diverse indoor and outdoor environments, the dataset includes meticulously synchronized sensor data, offering high detail and complexity. Our experiments using this dataset have highlighted the improved robustness of C-SLAM systems, especially in handling inter-loop closures. The addition of UWB measurements has significantly enhanced localization accuracy and reliability. The S3E dataset establishes a new benchmark for C-SLAM, fostering further research and innovation in multi-agent robotic navigation and mapping, and advancing the field.

ACKNOWLEDGMENT

This research greatly benefited from the guidance and expertise of many contributors. The authors extend their profound gratitude to our colleagues, Yizhen Yin and Haoxin Zhang from Sun Yat-sen University, for significantly improving our dataset's quality and applicability, especially in data collection. Special acknowledgment goes to Prof. Tao Jiang and his student Yudu Jiao from Chongqing University for evaluating the single-agent SLAM algorithms on our dataset.

REFERENCES

- [1] P. Schmuck, T. Ziegler, M. Karrer, J. Perraudin, and M. Chli, "COVINS: Visual-inertial SLAM for centralized collaboration," in *Proc. 2021 IEEE Int. Symp. Mixed Augmented Reality Adjunct*, Oct. 2021, pp. 171–176.
- [2] Y. Huang, T. Shan, F. Chen, and B. Englot, "DiSCo-SLAM: Distributed scan context-enabled multi-robot LiDAR SLAM with two-stage global-local graph optimization," *IEEE Robot. Automat. Lett.*, vol. 7, no. 2, pp. 1150–1157, Apr. 2022.
- [3] S. Zhong, Y. Qi, Z. Chen, J. Wu, H. Chen, and M. Liu, "DCL-SLAM: A distributed collaborative LiDAR SLAM framework for a robotic swarm," *IEEE Sensors J.*, vol. 24, no. 4, pp. 4786–4797, Feb. 2024.
- [4] S. Zhong et al., "CoLRIO: LiDAR-ranging-inertial centralized state estimation for robotic swarms," in *Proc. 2024 IEEE Int. Conf. Robot. Automat.*, 2024, pp. 3920–3926.
- [5] P.-Y. Lajoie, B. Ramtola, F. Wu, and G. Beltrame, "Towards collaborative simultaneous localization and mapping: A survey of the current research landscape," *Field Robot.*, vol. 2, no. 1, pp. 971–1000, Mar. 2022.
- [6] A. Geiger, P. Lenz, C. Stillner, and R. Urtasun, "Vision meets robotics: The KITTI dataset," *Int. J. Robot. Res.*, vol. 32, no. 11, pp. 1231–1237, 2013.
- [7] M. Burri et al., "The EuRoC micro aerial vehicle datasets," *Int. J. Robot. Res.*, vol. 35, no. 10, pp. 1157–1163, 2016.
- [8] K. Y. Leung, Y. Halpern, T. D. Barfoot, and H. H. Liu, "The UTIAS multi-robot cooperative localization and mapping dataset," *Int. J. Robot. Res.*, vol. 30, no. 8, pp. 969–974, 2011.
- [9] S. Agarwal, A. Vora, G. Pandey, W. Williams, H. Kourous, and J. McBride, "Ford multi-AV seasonal dataset," *Int. J. Robot. Res.*, vol. 39, no. 12, pp. 1367–1376, 2020.
- [10] R. Dubois, A. Eudes, and V. Frémont, "AirMuseum: A heterogeneous multi-robot dataset for stereo-visual and inertial simultaneous localization and mapping," in *Proc. 2020 IEEE Int. Conf. Multisensor Fusion Integration Intell. Syst.*, Sep. 2020, pp. 166–172.
- [11] Y. Tian et al., "Resilient and distributed multi-robot visual SLAM: Datasets, experiments, and lessons learned," in *Proc. 2023 IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2023, pp. 11027–11034.
- [12] Y. Zhu, Y. Kong, Y. Jie, S. Xu, and H. Cheng, "GRACO: A multimodal dataset for ground and aerial cooperative localization and mapping," *IEEE Robot. Automat. Lett.*, vol. 8, no. 2, pp. 966–973, Feb. 2023.
- [13] S. Zhao et al., "SubT-MRS dataset: Pushing SLAM towards all-weather environments," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2024, pp. 22647–22657.
- [14] S. Macenski, T. Foote, B. Gerkey, C. Lalancette, and W. Woodall, "Robot operating system 2: Design, architecture, and uses in the wild," *Sci. Robot.*, vol. 7, no. 66, 2022, Art. no. eabm6074.
- [15] C. Campos, R. Elvira, J. J. G. Rodríguez, J. M. M. Montiel, and J. D. Tardós, "ORB-SLAM3: An accurate open-source library for visual, visual-inertial, and multimap SLAM," *IEEE Trans. Robot.*, vol. 37, no. 6, pp. 1874–1890, Dec. 2021.
- [16] T. Qin, J. Pan, S. Cao, and S. Shen, "A general optimization-based framework for local odometry estimation with multiple sensors," 2019, *arXiv:1901.03638*.
- [17] T. Shan, B. Englot, D. Meyers, W. Wang, C. Ratti, and D. Rus, "LIO-SAM: Tightly-coupled LiDAR inertial odometry via smoothing and mapping," in *Proc. 2020 IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Oct. 2020, pp. 5135–5142.
- [18] T. Shan, B. Englot, C. Ratti, and D. Rus, "LVI-SAM: Tightly-coupled LiDAR-visual-inertial odometry via smoothing and mapping," in *Proc. 2021 IEEE Int. Conf. Robot. Automat.*, May 2021, pp. 5692–5698.
- [19] P.-Y. Lajoie and G. Beltrame, "Swarm-SLAM: Sparse decentralized collaborative simultaneous localization and mapping framework for multi-robot systems," *IEEE Robot. Automat. Lett.*, vol. 9, no. 1, pp. 475–482, Jan. 2024.