

Article

PIRATE—Precision Imaging Real-Time Autonomous Tracker & Explorer

Dan Zlotnikov  and Ohad Ben-Shahar * 

Faculty of Computer and Information Science, School of Brain Sciences and Cognition, Ben Gurion University of the Negev, Be'er Sheva 8410501, Israel; danzlot@post.bgu.ac.il

* Correspondence: ben-shahar@cs.bgu.ac.il

Abstract

We present PIRATE (Precision Imaging Real-time Autonomous Tracker and Explorer), a fully autonomous unmanned surface vehicle designed to enable self-operating data collection and persistent tracking of mobile underwater targets through the tight integration of acoustic localization, onboard visual perception, and closed-loop navigation. PIRATE employs a single mobile acoustic receiver to estimate target position using time-difference-of-arrival (TDoA) measurements acquired at different times and locations through planned autonomous motion and uses these estimates to drive adaptive vehicle behavior and activate fine-grained visual sensing in real time. This architecture enables sustained target-driven operation, in which navigation, acoustic monitoring, and visual processing are dynamically coordinated based on mission context and localization uncertainty. The system integrates real-time AI-based visual detection and tracking with automatic mission control, allowing visual perception to operate opportunistically within an acoustically guided tracking loop rather than as a standalone sensing modality. Field experiments in a shallow-water environment demonstrate reliable autonomous navigation, single-receiver acoustic localization with meter-scale accuracy, and stable onboard visual inference under sustained operation. By enabling coupled acoustic tracking and onboard visual perception in a fully autonomous surface platform free of external infrastructure, PIRATE provides a practical foundation for fine-scale behavioral observation, adaptive marine monitoring, and long-duration studies of mobile underwater organisms. We demonstrate this advantage with two possible applications.

Keywords: marine ROV; USV; unmanned surface vehicle; marine robotics; autonomous robot; underwater data collection; autonomous tracking; underwater computer vision; acoustic tracking



Academic Editor: Zhiguang Feng

Received: 20 February 2026

Revised: 7 March 2026

Accepted: 12 March 2026

Published: 17 March 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

The marine environment plays a central role in a wide range of scientific disciplines, including physical oceanography, marine ecology, climate science, and environmental monitoring. Many key oceanic processes that influence global climate, geochemical cycles, and ecosystem dynamics remain only partially characterized, constrained not by theory, but by the scarcity of direct observations across large spatial and temporal scales [1]. Despite decades of technological development, the ocean remains one of the least directly observed environments on Earth.

This observational sparsity is particularly pronounced for underwater visual data. It has been estimated that less than 0.001% of the deep seafloor has been visually observed [2],

underscoring the extreme limitations in direct imagery in such conditions. These limitations are especially restrictive for the study of biological processes and animal behavior, which often require continuous, spatially resolved observations. For example, understanding how marine animals such as penguins, turtles, and other marine animals move through heterogeneous habitats and respond to environmental variability remains challenging due to the limited availability of fine-scale, continuous tracking data [3,4].

While physical and chemical oceanographic variables are increasingly sampled at scale using autonomous observing platforms—such as underwater gliders that routinely collect temperature, salinity, dissolved oxygen, and other measurements over extended missions [5,6]—equivalent advances in sustained visual and spatially resolved biological observation have been more limited. As a result, many ecological datasets emphasize species presence and broad distribution patterns rather than fine-scale movement, interaction, and behavioral dynamics of individual organisms. Furthermore, underwater data collection faces inherent challenges arising from harsh and highly variable environmental conditions. Optical sensing is affected by signal attenuation, scattering, and distortion, as well as changes in color balance, illumination, and noise characteristics [7]. These effects complicate both human interpretation and automated analysis of underwater imagery and have motivated extensive research into robust detection and enhancement techniques [8].

A key technical limitation in fine-grained marine data collection is reliable underwater target localization. Unlike surface or aerial systems, underwater platforms cannot use Global Navigation Satellite System (GNSS) signals, as radio-frequency transmissions do not propagate through water. Instead, localization relies primarily on acoustics, in which a transmitting tag can be localized using acoustic receivers. One established approach employs static receiver arrays [9,10], which, under favorable geometry and environmental conditions, can achieve sub-meter to meter-scale localization accuracy, but require extensive seabed-mounted infrastructure and are restricted to fixed spatial footprints.

A second class of methods uses dynamic multi-platform configurations, in which multiple mobile platforms cooperatively estimate target position [11,12]. Such approaches typically achieve meter- to decameter-scale accuracy depending on platform coordination, communication latency, and synchronization quality, while offering greater spatial flexibility at the cost of increased system complexity and sensitivity to communication errors.

A third widely used approach is Ultra-Short Baseline (USBL) acoustic positioning, which estimates target bearing and range from a compact hydrophone array mounted on a surface or vehicle platform. USBL systems commonly provide meter- to multi-meter-scale accuracy in shallow and coastal environments, with performance strongly influenced by sound-speed uncertainty, array geometry, platform motion, and acoustic multipath [13]. In addition to being very costly, many commercial USBL configurations are optimized for tracking vehicles or divers using active transponders, which are typically larger and higher-power than wildlife telemetry tags, potentially limiting direct applicability to small marine organisms without hardware adaptation.

Table 1 highlights the trade-offs between infrastructure dependence, autonomy level, and achievable localization accuracy across representative underwater localization paradigms.

Beyond sensing modality alone, trajectory design plays a critical role in determining the effectiveness of underwater acoustic localization and data collection. In recent years, communication-aware and constraint-driven marine trajectory planning has received increasing attention, particularly in the context of multi-vehicle systems. Yang et al. [14] formulated underwater information collection as a distributed traveling salesman problem (DTSP) for coordinated unmanned marine vehicles, integrating clustering, Dubins-constrained motion, and a pseudo-trajectory homotopy method to ensure kinematic feasibility under underwater communication constraints. Their framework explicitly

couples vehicle path geometry with acoustic transmission considerations, highlighting the importance of motion-generated spatial diversity in infrastructure-free underwater sensing. Similarly, Zhu et al. [15] proposed an event-driven cooperative localization–formation control strategy for multiple USVs, demonstrating how coordinated trajectory control and adaptive formation management can improve localization robustness and responsiveness in distributed marine systems. Collectively, these works emphasize that localization quality and data collection efficiency are tightly linked to trajectory geometry, vehicle coordination, and communication structure. However, most existing approaches focus on multi-platform coordination and algorithmic optimization, with validation typically limited to simulation or controlled experiments. The integration of geometry-aware trajectory planning with real-time onboard perception and fully autonomous, field-validated operation remains comparatively underexplored.

Table 1. Comparison of representative underwater localization approaches.

Metric	Fixed Arrays	Multi-Veh.	ROV + USBL	PIRATE
Localization Error	<1–5 m	1–10 m	1–5 m	~1–5 m
Infrastructure	Seabed array	Multi-vehicle	Vessel	Vessel
Autonomy Level	Low	High	Low	Full
Operational Range	Fixed	>1 km	Tether-limited	>1 km
Multimodal integration	No	No	No	Yes
Relative Cost	High	High	Very High	Low–Med

Consequently, the combined challenges of limited visual coverage, environmental sensing constraints, acoustic localization limitations, and the dependence of localization performance on vehicle trajectory geometry have strongly shaped the design of marine observation systems, highlighting the need for integrated platforms that can autonomously combine adaptive motion, acoustic sensing, and real-time perception to enable scalable, infrastructure-independent underwater data collection.

Early approaches to underwater data collection relied primarily on human-operated and infrastructure-based observation methods. Diver-based visual surveys have long been used to document benthic habitats and reef-associated fauna, providing high-quality localized observations but remaining inherently limited by depth, dive duration, safety constraints, and logistical cost [16,17]. In parallel, stationary and cabled underwater observatories enabled continuous monitoring at fixed locations, supporting long-term ecological and environmental measurements but offering no spatial mobility and limited responsiveness to dynamic events [18,19]. Tethered, remotely operated vehicles (ROVs) further extended observational reach by enabling controlled visual inspection beyond diver limits; however, their operation remains dependent on surface vessels, tether management and range, and continuous human supervision [20].

With time, advances in marine engineering, and in robotics in general, have opened avenues for development of untethered ROV systems capable of executing predefined routes or inspection patterns under remote human control. These platforms improved spatial coverage and operational flexibility and have been widely adopted for tasks such as seabed inspection and data collection [21,22]. Nevertheless, such systems remain fundamentally reactive, with perception, interpretation, and decision making performed by a remote operator or deferred to post-mission analysis, limiting their ability to adapt behavior in response to observed underwater dynamics.

More recently, autonomous and semi-autonomous marine platforms—including unmanned underwater vehicles (UUVs) and unmanned surface vehicles (USVs)—have emerged as key components of modern ocean observing systems, enabling persistent

data collection over extended spatial and temporal scales while reducing reliance on crewed operations [1,6,23]. These systems typically support basic autonomy, such as way-point navigation, predefined survey execution, or infrastructure inspection missions [24]. While such capabilities represent a substantial advance, autonomous behavior is often limited to mission execution rather than interpretation-driven adaptation, with sensing and higher-level decision making performed offline or supported by intermittent shore-based communication [3]. As a result, these platforms are highly effective for static surveying and episodic measurements but remain less suited for sustained interaction with dynamic underwater targets.

To enable more fine-grained and higher-quality underwater observations, there is a growing need for autonomous systems that tightly couple data collection with real-time responsiveness to the acquired data. Achieving this capability requires robust sensor fusion across heterogeneous sensing modalities—such as acoustic localization and visual perception—combined with onboard intelligence capable of autonomously managing perception, decision making, and control. A unified, onboard computational core that performs real-time sensor fusion and adaptive behavior selection allows a platform to respond directly to observed phenomena, improving spatial resolution, temporal continuity, and overall data relevance.

Large-scale marine animal movements are often punctuated by extended periods of localized residency, providing natural temporal windows for targeted observation. Green sea turtles (*Chelonia mydas*), for example, undertake long-distance migrations between developmental, foraging, and breeding areas, yet high-resolution Fastloc-GPS telemetry (that requires periodic animal resurfacing) demonstrates that individuals frequently establish distinct residence areas where they remain spatially constrained for extended periods, conducting repeated short excursions within localized habitat patches [25]. These residence phases may persist for weeks to months and encompass ecologically significant behaviors related to foraging, refuge use, and energy balance that are poorly resolved by coarse or episodic observations. In such regimes, real-time, in situ processing enables autonomous systems to exploit predictable residency windows through adaptive motion and sensing, rather than relying solely on static or preprogrammed survey strategies.

Similar periods of localized, meter-scale stationarity are common across many benthic and reef-associated species, including coral trouts that exhibit strong site fidelity near shelters [26], rays that remain largely stationary while resting on sandy substrates [27], and territorial reef fishes such as damselfish that exhibit strong site fidelity, occupying and defending localized habitat patches while restricting their movements largely to small, well-defined territories [28].

Despite substantial progress in marine robotics, several critical gaps remain. First, infrastructure-free acoustic localization using a single mobile receiver has rarely been validated under sustained real-world field conditions; most prior work relies either on fixed receiver arrays or multi-platform coordination, with limited embodied experimental verification. Second, while real-time underwater visual detection and tracking have advanced considerably, these capabilities are typically deployed as standalone perception modules rather than tightly coupled with acoustic localization in a unified closed-loop autonomy framework. Third, although geometry-aware trajectory design has been studied extensively in theory and simulation, its integration with onboard perception and its validation in fully autonomous marine field deployments remain comparatively underexplored. Addressing these gaps requires not only algorithmic development, but the experimental demonstration of a tightly integrated, field-operational system capable of sustained target-driven behavior without external infrastructure. PIRATE attempts to make an important step in this direction.

Contributions

This work advances the state of autonomous marine observation through the following contributions:

1. Field-validated single-receiver acoustic localization. We experimentally demonstrate that motion-generated spatial diversity driven by a single mobile acoustic receiver enables meter-scale underwater target localization under real marine conditions. Unlike infrastructure-extensive array systems, this approach requires no fixed receivers or synchronized multi-platform deployments, and is validated through repeated field experiments with quantified error analysis.
2. Closed-loop integration of acoustic localization and onboard visual perception. We present and validate a tightly coupled autonomy framework in which acoustic target estimates directly regulate vehicle motion and opportunistically trigger embedded visual inference. This establishes a perception-driven control loop operating fully onboard, rather than treating acoustic localization and visual detection as independent sensing modalities.
3. Infrastructure-free persistent tracking of mobile underwater targets. We demonstrate sustained, target-driven operation in which localization, pursuit, and re-localization occur autonomously within a single continuous mission, without reliance on fixed acoustic arrays, shore-based processing, or continuous human supervision.
4. Embodied validation of geometry-aware localization principles. Through systematic variation of tracking polygon scale and field-based error quantification, we experimentally verify how trajectory geometry influences localization robustness. This bridges theoretical trajectory-localization coupling concepts with real-world marine deployment.
5. A practical framework for fine-scale behavioral observation of marine life. By combining real-time onboard inference with infrastructure-free acoustic tracking, the proposed system establishes a deployable methodology for adaptive, high-resolution monitoring of mobile marine organisms, supporting scientific investigations that require spatial and temporal continuity beyond static survey methods.

2. Hardware Overview

2.1. USV Platform

PIRATE is implemented on a custom unmanned surface vehicle (USV) employing a catamaran hull configuration, selected to provide passive roll stability, redundancy, and low-speed maneuverability during autonomous operation (Figure 1). Differential thrust control is achieved through two electrically driven stern-mounted thrusters, enabling precise heading regulation and controlled motion during acoustic localization and tracking maneuvers.

The physical platform follows a modular catamaran design inspired by a previously reported surface vehicle architecture [29], but in our case, it hosts a very different computational backbone and hardware, integrated acoustic telemetry, onboard GPU-enabled visual perception, and closed-loop autonomous control. Specifically, all system integration, onboard software architecture, sensing pipelines, and autonomous behaviors described in this work are independent, novel, and unique to the PIRATE system.

The vehicle structure consists of a polyethylene deck supporting two watertight enclosures that house onboard electronics and power systems. Both the acoustic receiver and the visual sensor are mounted below the waterline with the camera installed in a sealed housing on a custom fin to minimize hydrodynamic disturbance while maintaining a stable downward-facing viewing geometry.

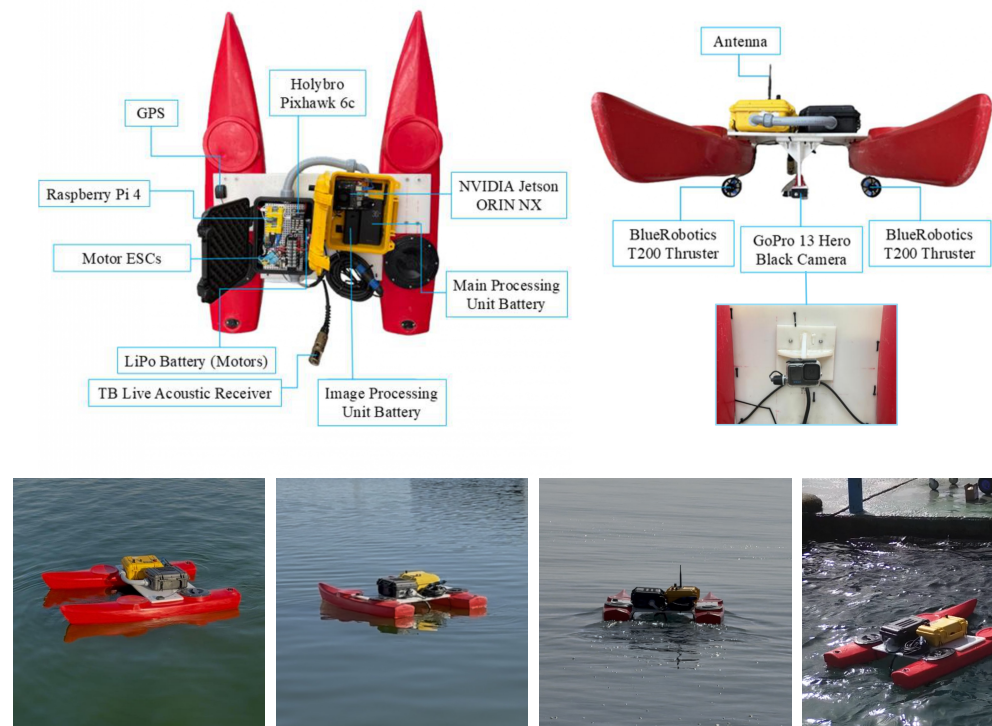


Figure 1. PIRATE platform overview and field deployment. (**Top**): system component layout showing onboard sensing, computation, propulsion configuration, acoustic receiver placement, and camera orientation. (**Bottom**): representative field deployments during autonomous operation under real marine conditions.

The platform is designed to support multi-hour autonomous missions under calm to moderate sea states, providing sufficient payload capacity and electrical power for continuous sensing, onboard computation, and closed-loop navigation. Key physical, propulsion, and power-related parameters of the platform are summarized in Table A1.

2.2. Vehicle Kinematics and Motion Characteristics

For navigation and maneuver execution, PIRATE employs a deliberately simple planar kinematic approximation, as maneuverability and complex movement is not a focus of this work.

PIRATE is actuated by two independently controlled stern-mounted thrusters separated by a lateral distance $b = 0.6$ m, forming a differential-thrust configuration. At the operating speeds used for acoustic tracking and navigation ($v_{\max} \approx 2$ m/s), the vehicle motion can be approximated using a planar kinematic model:

$$\dot{x} = v \cos \psi, \quad \dot{y} = v \sin \psi, \quad \dot{\psi} = \omega \quad (1)$$

where (x, y) denote global position, ψ is the heading angle, v is the forward velocity, and ω is the yaw rate. Forward and angular velocities are generated through differential thrust according to

$$v = k_v(T_L + T_R), \quad \omega = k_\omega(T_R - T_L), \quad (2)$$

with T_L and T_R denoting normalized left and right thruster commands, and k_v, k_ω representing empirical propulsion constants.

Two main behaviors are implemented. If $T_L = T_R$, then $\omega = 0$ and the vehicle moves in a straight line along its current heading with velocity $v = 2k_v T$. If, however, one thruster

is commanded forward ($T_R = T$) and the opposite thruster remains idle ($T_L = 0$), the resulting motion becomes

$$v = k_v T, \quad \omega = k_\omega T, \quad (3)$$

yielding a turning radius

$$R = \frac{v}{\omega} = \frac{k_v}{k_\omega}, \quad (4)$$

which produces tight rotational motion about a point located near the inactive side of the vessel. The angular velocity can also be expressed from a simplified moment balance as

$$\omega = \frac{Fb}{2I_z}, \quad (5)$$

where F is the thrust force, b is the thruster spacing, and I_z is the yaw moment of inertia.

This kinematic characterization links differential thrust actuation to forward motion and rapid heading reorientation, supporting the closed-loop tracking and acoustic listening maneuvers implemented in PIRATE.

2.3. Acoustic Sensing Hardware

Acoustic target detection is performed using a commercially available single-receiver underwater acoustic telemetry unit. Tagged targets emit periodic acoustic signals that are detected and timestamped onboard during vehicle operation. Each detection is synchronized with the vehicle state estimate and logged locally for real-time processing and post-mission analysis.

Importantly, the system operates without reliance on fixed acoustic infrastructure or synchronized receiver arrays, and instead uses a dynamic, single-receiver approach (see Section 3.3). Fundamental to this method are localization-relevant measurements obtained through deliberate vehicle motion, facilitating infrastructure-free acoustic tracking in unrestricted environments, all under autonomous control. Acoustic hardware specifications are described in Table A2.

2.4. Visual Sensing Hardware

Visual observations are acquired using an onboard action camera mounted in a downward-facing configuration relative to the vehicle frame. Though not restricted to specific visual sensors, the one used in our PIRATE prototype was optimized for shallow-water operation under ambient illumination and provides wide-angle coverage suitable for opportunistic visual detection and tracking during acoustic-guided missions.

If needed, the visual sensing subsystem can be activated selectively based on mission context, allowing computationally intensive perception to operate only when target proximity, as inferred acoustically, makes visual observation viable. Table A2 summarizes the camera specifications.

2.5. GPS Localization

Global positioning is provided by a compact GNSS receiver connected to the main mission controller. GNSS measurements support waypoint navigation, trajectory logging, and synchronization of acoustic detections with vehicle position.

2.6. Communication

Short-range communication between the USV and a shore-based station is achieved via a wireless peer-to-peer link. This channel is used exclusively for mission initialization, monitoring, and configuration. All sensing, decision-making, and control logic execute fully onboard during autonomous operation.

2.7. Onboard Computation Hardware

PIRATE employs a dual-layer onboard computation architecture that separates mission control from high-throughput perception. A low-power embedded controller—Raspberry Pi 4—manages vehicle navigation, mission sequencing, sensor coordination, and safety logic, while a dedicated GPU—NVIDIA Jetson Orin NX—performs real-time visual inference. The GPU’s specifications are summarized in Table A3.

This separation allows computationally demanding perception tasks to be executed without interfering with safety-critical control loops, while enabling sustained real-time operation under embedded power and thermal constraints. These capacities are achieved while using cheap hardware and software components, specifically avoiding the need for a real-time operating system (OS) that would otherwise be required for mission control. The perception unit runs a Linux-based operating system and executes deep-learning-based detection and tracking models during autonomous deployment.

3. Software Architecture

3.1. System Software Stack

PIRATE adopts a modular software architecture in which low-level vehicle actuation—including motor arming and thrust control—is handled through standard MAVLink communication with an onboard PX4 autopilot. All higher-level system functionality, including navigation logic, mission sequencing, and target-driven decision making, is implemented on a main Raspberry Pi controller using custom Python3.10.13 modules, while computationally intensive tasks, most notably AI-based visual detection and tracking, are delegated to a dedicated onboard GPU (Figure 2).

Most peripheral sensors interface directly with the Pi via USB, while the primary visual sensor is connected directly to the GPU unit over USB-C to minimize input latency and avoid unnecessary data routing through the control layer.

To ensure deterministic behavior and fine-grained control over autonomous operation, the software stack was developed without reliance on external behavior-control frameworks. This design enables full modularity and configurability across all operational stages while reducing dependency on third-party software updates that may introduce unintended behavioral changes.

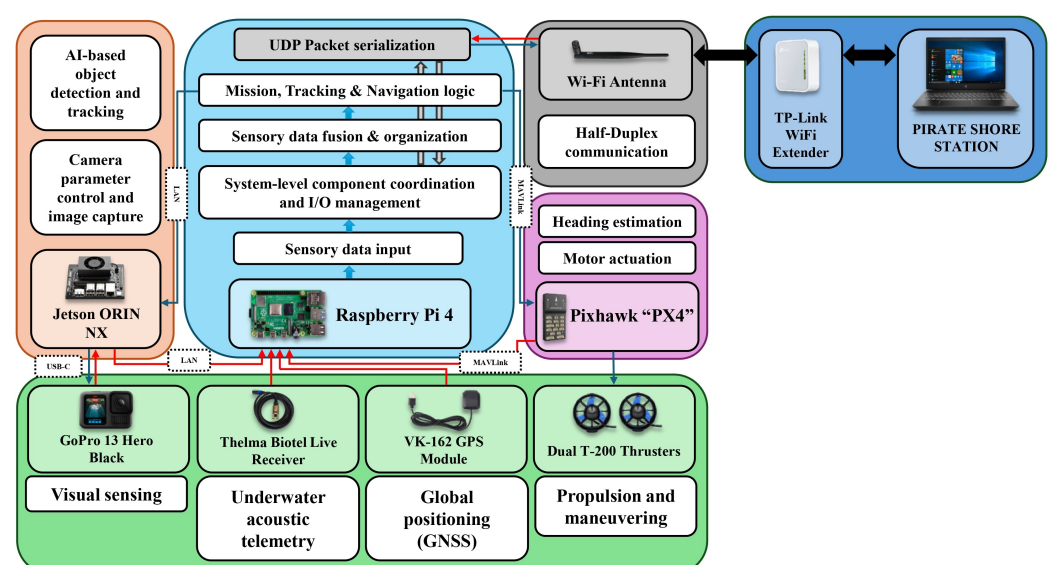


Figure 2. PIRATE system architecture illustrating the separation between low-level vehicle control, high-level mission and navigation logic, and onboard sensing and perception, with computationally intensive visual processing offloaded to a dedicated GPU.

3.2. Mission Control and System Coordination

A Raspberry Pi 4 serves as the primary mission controller and system coordinator within the PIRATE architecture. It manages autonomous mission execution, orchestrates communication between subsystems, and aggregates data from onboard sensors and processing units. Visual and acoustic data streams are synchronized to support real-time decision-making and logged for post-mission analysis. The Pi acts as a supervisory control layer that governs mission execution, coordinates subsystem interactions, and enforces safety-critical constraints across heterogeneous hardware components. During operation, the Pi issues navigation and actuation commands to the PX4 flight controller, which performs real-time motor control and stabilization.

Computationally extensive tasks such as object detection and tracking are performed upon request by the NVIDIA Jetson GPU, triggered by the Pi to begin or terminate their execution. The GPU sends computation results back to the Pi, e.g., target location or detected object count in the current camera frame.

PIRATE is built to operate completely autonomously. However, mission initialization and pre-configuration can be performed using the shore station GUI. Communication with the shore station is established via a 2.4 GHz Wi-Fi link, enabling transmission of mission status, telemetry, and selected sensor data, as well as reception of high-level commands when connectivity is available. To minimize latency under bandwidth and connectivity constraints, a lightweight custom UDP-based communication protocol is employed.

Mission execution is structured as a finite state machine (FSM) governing distinct operational modes, including idle, navigation, tracking, visual processing, return-to-home (RTH), and fault handling. This FSM-based design enforces deterministic state transitions, prevents conflicting behaviors between subsystems, and guarantees bounded responses to safety-critical events. As illustrated in Figure 3, high-level operational states are organized hierarchically, with composite modes encapsulating navigation and perception substates. Transition triggers are event-driven, and a global interrupt mechanism allows immediate transition to the RTH state from any active mode, ensuring fail-safe recovery.

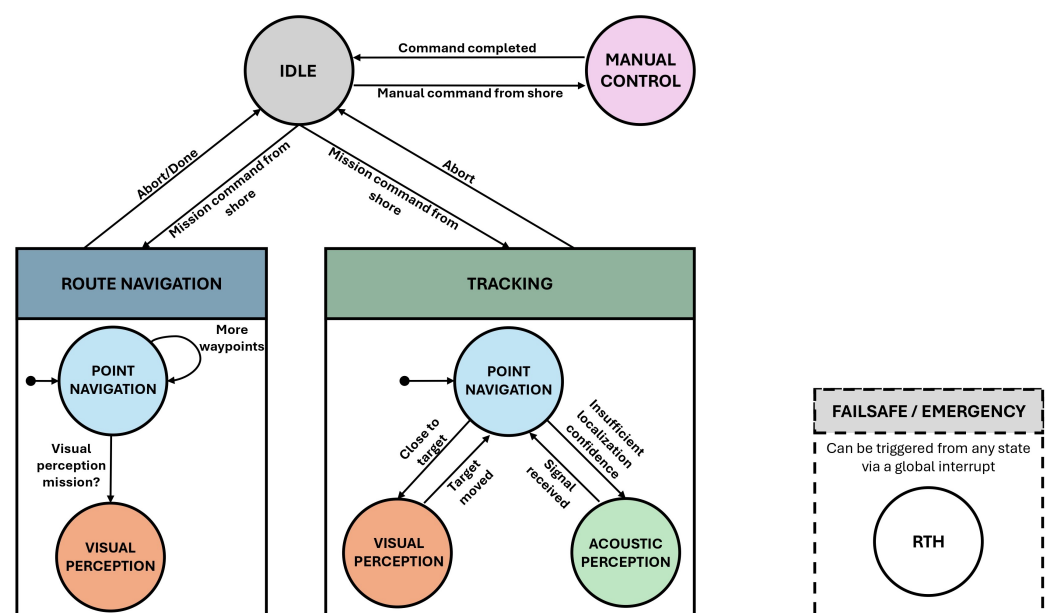


Figure 3. Finite state machine governing mission execution in PIRATE. High-level operational modes are represented as composite states, with internal substates coordinating navigation, visual perception, and acoustic sensing. A global failsafe mechanism enables RTH from any state via an interrupt-driven transition.

By centralizing mission logic and inter-device coordination on the Raspberry Pi, the PIRATE system achieves a modular architecture in which sensing, actuation, and perception components are logically integrated while physically decoupled. This design supports reliable autonomous operation under intermittent or non-existing communication conditions and facilitates integration of additional sensors and payloads for mission-specific extensions.

PIRATE's RTH capabilities are fully customizable and can be triggered upon communication timeout, mission end, or various failure modes.

3.3. Single-Receiver Acoustic Localization

Acoustic telemetry is widely used for tracking tagged marine animals, traditionally relying on time-difference-of-arrival (TDoA) measurements obtained from spatially distributed, synchronized receiver arrays. While such configurations provide high localization accuracy, their deployment requires fixed and costly infrastructure that limits operational flexibility and spatial coverage.

PIRATE adopts an alternative single-receiver localization strategy, in which spatial diversity is achieved through deliberate motion of a mobile acoustic receiver. Rather than relying on simultaneous receptions across multiple receivers, the system exploits successive acoustic detections acquired at different vehicle positions over time. This approach enables infrastructure-free localization while retaining the geometric constraints characteristic of TDoA-based methods.

Within PIRATE, each acoustic detection is timestamped and associated with the vehicle's estimated position at the time of reception. During tracking missions, the platform autonomously executes predefined polygonal listening trajectories to generate sufficient spatial diversity for localization. Target position estimates are updated incrementally on-board using motion-compensated TDoA calculations that account for receiver displacement and potential target motion between transmissions.

The feasibility of single-receiver acoustic tracking of marine animals using autonomous surface platforms has been demonstrated in prior work [29], which reported meter to decameter-scale localization accuracy under field conditions using the simplest acoustic tags. PIRATE builds on this foundation by integrating single-receiver localization directly into a closed-loop autonomy framework, where acoustic position estimates actively guide vehicle motion and trigger complementary sensing modalities. No explicit constant-velocity model is imposed; instead, localization is recomputed sequentially from consecutive receiver positions, and motion-induced inconsistencies are resolved through the underlying TDoA optimization framework [29].

In the PIRATE system, acoustic localization is performed entirely onboard and operates continuously during tracking missions. The resulting target estimates are used to adapt navigation behavior in real time, enabling persistent, target-driven operation without reliance on fixed receiver arrays or continuous human supervision. Additional background on TDoA geometry and single-receiver localization principles [29] is provided in Appendix B.

3.4. Acoustic Localization Enhancement via ToF Range Estimation

In addition to its role in time-difference-of-arrival (TDoA)-based localization, the acoustic receiver is also used for continuous target-range estimation. After the initial target position is obtained using the single-receiver triangulation procedure, the system estimates the acoustic tag's transmission time by back-propagating previously recorded receptions, employing their timestamps, the speed of sound, and the estimated target location. This estimated transmission time enables subsequent time-of-flight (ToF)-based range

measurements using the tag's regular, periodic transmissions. By continuously listening to the acoustic signal stream, PIRATE maintains an online estimate of the target–vehicle range, which is integrated into the mission logic. This range estimate supports adaptive decision-making, such as enabling or disabling visual sensing, triggering re-localization when range uncertainty increases, and adjusting mission states during target approach and tracking.

3.5. Visual Detection and Tracking Pipeline

Once PIRATE determines that it is within an operationally viable range of a tracked target, the onboard visual perception pipeline is activated.

To balance robustness and computational efficiency under embedded constraints, PIRATE employs a detector–tracker orchestration strategy that combines a deep-learning-based object detector with a single-object visual tracker. Object detection provides global scene awareness and is used for initial target acquisition and periodic verification, while single-object tracking exploits temporal continuity to achieve efficient frame-to-frame localization between detection steps.

Closed-loop vision-based target tracking architectures that couple visual detection, recursive state estimation, and feedback control have been widely investigated in robotics [30,31]. These systems typically rely on Kalman-filter-based motion modeling to propagate target state between observations and to stabilize frame-to-frame localization under measurement noise and short-term occlusion, while lightweight controllers regulate vehicle motion to maintain visual lock in real time. PIRATE adopts a related tracking-centric perception–prediction paradigm, but extends it to transformer-based single-object tracking and burst-orchestrated detection under embedded GPU constraints in dynamic marine environments.

In the current implementation, a YOLO-based [32] detector (specifically, YOLOv7 [33]) is paired with a transformer-based STARK [34] tracker. YOLO-style detectors are well-suited for target initialization and re-identification in cluttered marine scenes due to their strong appearance modeling, while STARK explicitly models motion and temporal consistency, enabling robust short-term tracking under appearance changes and partial occlusions. The detector was pretrained on a composite underwater dataset comprising samples from Fish4Knowledge [35], DeepFish [36], and Brackish [37]. These datasets span a range of turbidity levels, illumination conditions, background complexity, and target appearances.

Within PIRATE, visual tracking is initialized using a detector-derived bounding box, after which the tracker operates for a fixed burst of N consecutive frames, where N is a configurable system parameter. After each tracking interval, the detector is re-invoked to verify target identity and spatial consistency, enabling correction of accumulated tracker drift or recovery from target loss. Detector outputs are incorporated into the tracking state only when the associated detection confidence exceeds a predefined threshold; otherwise, the existing tracker state is maintained to prevent disruption due to spurious detections. A visual representation of this pipeline is provided in Figure 4.

This alternating detection–tracking cycle enables sustained real-time operation on the embedded hardware while preserving robustness in visually challenging marine environments characterized by haze, clutter, variable illumination, and dynamic backgrounds. Benchmark performance of the selected detection and tracking models on standard datasets, as reported in the original publications, is provided in Appendix C.

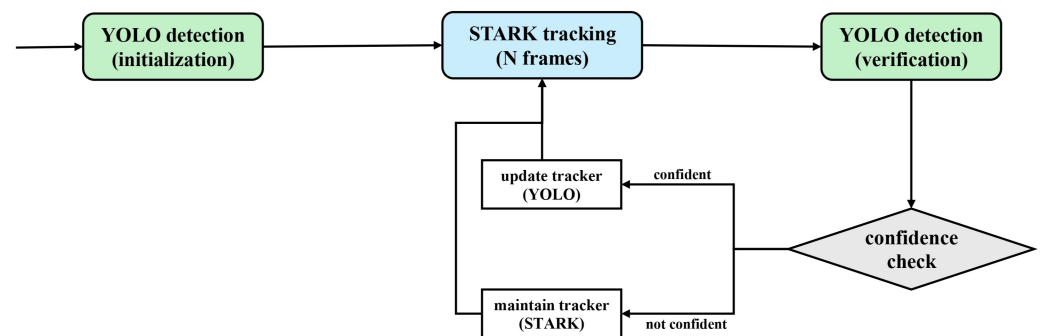


Figure 4. Schematic overview of the detector–tracker orchestration strategy employed in PIRATE. Object detection is used for target initialization and periodic verification, while single-object tracking operates for fixed intervals between detection steps. Tracker updates are applied only when detection confidence exceeds a predefined threshold.

4. Experimental Validation

Field experiments were conducted at three distinct locations selected to provide complementary environmental conditions and to progressively evaluate the capabilities of the PIRATE platform:

Be’er Sheva Ecological Lake: (31.237250° N, 34.825240° E). This location offers a calm, shallow-water environment with minimal surface disturbance, making it well-suited for controlled testing of surface vehicle operation, autonomous navigation, wireless communication, and acoustic sensing. Due to limited water clarity at this site, underwater visual observation was not feasible, and visual perception experiments were therefore not conducted. The mean wind speed during the deployment was 2.6 m/s, with predominantly easterly winds and maximum gusts of 6.7 m/s.

Sedot Yam Bay: (32.493822° N, 34.889102° E), a shallow coastal area in the Mediterranean Sea characterized by mild wave activity and sufficient underwater visibility. Experiments at this site were conducted in coordination with marine scientists from the Sedot Yam Research Station. The mean wind speed was approximately 2.8 m/s, with maximum gusts reaching 10 m/s. Wave height was small (<0.5 m).

Ashkelon Marina: (31.683569° N, 34.555747° E), an active Mediterranean marina offering a sheltered, semi-enclosed environment with reduced wave exposure due to breakwaters. Shallow waters and adequate underwater visibility enabled visual perception, navigation, and acoustic experiments under realistic marina conditions, including reflections from harbor structures and intermittent vessel activity. Mean wind speed was approximately 1.2 m/s, with maximum gusts up to 5 m/s, and wave height remained small (<0.5 m).

Figure 5 depicts the sites and the locations where experiments were conducted. All experiments were performed using the fully integrated PIRATE platform, including onboard propulsion, sensing, computation, and communication subsystems. The experimental campaign was designed to incrementally validate system capabilities, progressing from basic mechanical operation and navigation in a controlled environment to autonomous visual perception and sensing in a realistic coastal setting. During all field trials, PIRATE was configured to use 50% thrust for forward motion and 25% thrust for rotational maneuvers. Acoustic receiver functionality was achieved by deploying a commercially available Thelma Biotel acoustic tag (69 kHz carrier frequency, 10 s transmission interval) placed in the water less than 2 m from the deployment site.



Figure 5. Experimental deployment sites used for field validation of the PIRATE platform. (a) Be'er Sheva Ecological Lake, a controlled shallow-water environment used for system commissioning and autonomous navigation validation. The green region indicates the area used for navigation experiments, while the indigo region denotes the area used for acoustic tracking. (b) Sedot Yam Bay, a Mediterranean coastal site used for autonomous visual perception and acoustic tracking experiments; the red region indicates the area used for visual data collection. (c) Ashkelon Marina, an operational Mediterranean marina site used for experimental validation of visual perception, navigation, and acoustic localization under sheltered harbor conditions. Plus symbols indicate initial PIRATE deployment points; the yellow region indicates the area used for pursuit and visual perception experiments.

4.1. Experimental Validation 1—Autonomous Navigation

This experiment evaluated the autonomous navigation capabilities of the PIRATE system under real-world field conditions. The objective was to assess waypoint-based navigation accuracy, route-following behavior, and system robustness prior to initiating target-driven tracking missions. The PIRATE platform and shore station were initialized at predefined locations, and all navigation commands were issued via the shore station graphical user interface or loaded from mission configuration files stored onboard. The Return-To-Home (RTH) safety mechanism was enabled throughout the experiment and configured to automatically activate following 60 s of sustained communication loss with the shore station.

Two navigation scenarios were evaluated: (i) single-point navigation and (ii) multi-point navigation.

Single-Point Navigation Setup

In the single-point navigation scenario, PIRATE was commanded to autonomously navigate from its current position to a single target waypoint. Each navigation mission consisted of a straight-line planned route between the start position and the specified waypoint. Waypoint acceptance was defined by a configurable radial tolerance, allowing mission termination once the platform entered the acceptance region around the target.

An example of a planned single-point navigation route used during the experiment is shown in Figure 6.

Eight independent single-point navigation trials were executed using different target waypoints. This was done to evaluate PIRATE's performance across different wind, drift and route length parameters. During each trial, the platform autonomously executed the navigation mission while logging platform state, position estimates, and control outputs for subsequent analysis.

Multi-Point Path Navigation Setup

In the multi-point navigation scenario, PIRATE was tasked with autonomously executing a predefined route composed of multiple sequential waypoints. The planned route consisted of a semi-closed polygonal path designed to evaluate path-following accuracy across multiple short navigation legs and frequent heading changes.

An example of a planned multi-point navigation route used during the experiment is shown in Figure 7.



Figure 6. Example planned route for a single-point navigation mission. The route consists of a straight-line path between the start position and a single target waypoint.

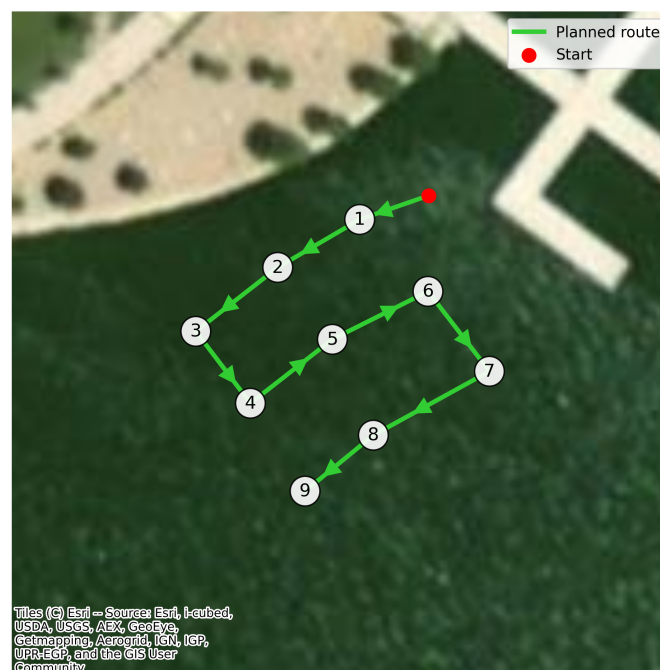


Figure 7. Example planned route for a multi-point navigation mission. The route is defined by multiple sequential waypoints forming a compact polygonal trajectory.

The same multi-point route was executed three times. In the first execution, the route was selected and transmitted from the shore station graphical interface. In the subsequent executions, the identical route was loaded from a predefined mission file stored onboard the PIRATE platform and executed autonomously without operator intervention. All navigation missions were logged in full for post-mission performance evaluation.

4.2. Experimental Validation 2—Acoustic Target Localization

This experiment evaluated the acoustic localization capability of the PIRATE system using a single mobile receiver under real-world field conditions. The objective was to assess the system's ability to estimate the position of an acoustic target by actively maneuvering

the platform to create spatial diversity through controlled motion, without relying on fixed receiver infrastructure.

As mentioned above, the acoustic target was a Thelma Biotel 9 mm transmitter tag, transmitting in 10 s intervals, placed 1 m deep underwater at (31.237493° N, 34.824349° E). All localization trials were conducted with the same fixed target position and comparable operating ranges to isolate the influence of listening geometry.

To provide a reference for localization accuracy, the tag was statically fixed at the deployment point and PIRATE was positioned directly above it. The onboard GNSS position was averaged over a 60 s station-keeping interval to obtain a reference horizontal location. Localization error was computed as the Euclidean distance between the estimated acoustic position and this reference. As the reference is GNSS-derived, the reported errors represent a conservative upper bound on positioning accuracy under field conditions.

Each localization mission followed a structured four-stage procedure:

1. Definition of a search polygon according to predefined edge length and polygon direction parameters.
2. Autonomous navigation to each polygon vertex
3. Acoustic signal reception and logging.
4. Upon completion of the full polygon traversal—estimation of the target position using the collected acoustic measurements.

An example of the planned search polygon is shown in Figure 8, where predefined listening vertices are visited sequentially to generate motion-induced spatial diversity.



Figure 8. Example planned search polygon used in the acoustic localization experiments. The polygon vertices denote predefined listening locations visited sequentially by the PIRATE platform (solid arrows indicate traversal order). The dashed edge illustrates the closing segment of the polygon. The pink cross marks the ground-truth position of the acoustic tag, which was unknown to the system during mission execution.

The search polygon geometry was specified prior to each mission execution and remained fixed throughout each individual run. To evaluate the effect of spatial scale on localization performance, three different polygon sizes were defined. Each polygon configuration was executed in three independent trials, resulting in a total of nine acoustic localization experiments. The platform completed the full navigation and listening sequence for each run before computing a target position estimate.

All raw acoustic detections, platform positions, and timestamps were logged during each mission to enable post-mission analysis and comparison across polygon sizes and repeated trials. All post-processing was performed automatically using a custom library to produce the reported figures directly from the mission logs.

Scope of Trajectory Evaluation

The objective of these experiments was to validate motion-generated spatial diversity within the integrated closed-loop autonomy framework, rather than to exhaustively optimize trajectory geometry. To isolate geometric conditioning effects under field conditions, polygon side length was systematically varied while maintaining a consistent traversal structure. Other trajectory parameters—such as shape (e.g., triangular, circular), vehicle speed, and listening dwell time—were not independently optimized in this study and left for future work. Systematic optimization of trajectory shape and kinematic parameters was not pursued, as the emphasis was on validating the integrated autonomy framework under controlled geometric scaling. The proposed architecture remains compatible with uncertainty-aware or optimization-based trajectory planning, in which localization residuals or geometric conditioning metrics guide autonomous adjustment of listening geometry.

4.3. Experimental Validation 3—Visual Perception Under Controlled Conditions

Experiment 3 evaluated the visual perception subsystem of PIRATE under controlled laboratory conditions; Instead of acquiring them live in the field, the Jetson was pre-loaded with video frame sequences and configured to stream and process them. The experiments focused on real-time inference performance, detection–tracking scheduling, thermal behavior during sustained execution, and qualitative tracking behavior across visually diverse underwater environments. All experiments were conducted on the Jetson platform using an identical software pipeline. As mentioned before, the perception system was initialized with pretrained detector weights configured for a single *fish* class and executed continuously without manual intervention or runtime resets.

Real-Time Inference Performance. The first laboratory experiment evaluated real-time inference behavior under fixed, repeatable visual input. A prerecorded underwater video sequence consisting of 1835 frames was streamed to the platform in a live-camera simulation mode. Inference was performed at four input resolutions (360p, 480p, 640p, and 720p, at 16:9 ratio), and each configuration was executed ten independent times.

The number of consecutive STARK tracking frames between YOLO detector invocations (STARK BURST) was varied to study the effect of detector–tracker scheduling on computational load and temporal propagation. A burst length of zero corresponds to detector-only operation, while increasing burst lengths shift execution toward tracker-based propagation. Numerous values were included to sample a range of scheduling regimes. Inference throughput and latency measurements were recorded over the full sequence and aggregated across repeated runs.

YOLO-STARK Detector Evaluation. The second laboratory experiment examined detection–tracking behavior across heterogeneous underwater visual conditions. The YOLO–STARK orchestration pipeline was executed on representative sequences drawn from Fish4Knowledge, DeepFish, and Brackish datasets.

All sequences were processed using identical model weights and orchestration parameters. Tracking outputs were recorded for subsequent qualitative analysis of temporal continuity, detector re-initialization behavior, and robustness to domain variation.

Thermal Behavior Test. The third laboratory experiment examined thermal behavior during prolonged inference execution. Images from multiple underwater datasets were streamed continuously to the Jetson platform, enabling uninterrupted inference over more

than 25,000 consecutive frames. GPU temperature was logged throughout the experiment to characterize thermal response under sustained computational load.

5. Demo Applications

In the final part of this study, the platform was evaluated through two field-based demonstration experiments designed to approximate realistic scientific missions rather than controlled technical benchmarks. The objective is not only to validate individual subsystems, but to assess whether the integrated architecture can sustain meaningful, end-to-end operation under field conditions. These demonstrations are structured to simulate practical deployment scenarios in which target localization, motion planning, and onboard perception must operate continuously and coherently, thereby illustrating PIRATE's capacity to execute missions with genuine scientific value.

5.1. Demo Application 1—Visual Perception in the Field

In this experiment, PIRATE employed detector-only visual inference using a pretrained YOLO-based model. This configuration was selected because the experiment did not involve a single persistent visual target. Instead, the objective was to simulate an exploratory or seafloor-mapping mission scenario, in which the platform encounters multiple transient objects of interest within the field of view rather than tracking a specific individual over time. In this context, continuous object detection provides a more appropriate perception modality than single-object tracking.

The demonstration application was conducted during two separate field deployments: one at Sedot Yam Bay and one at Ashkelon Marina. Prior to each deployment, the onboard Jetson computing unit was initialized manually. Upon initiation of the visual perception script, the onboard GoPro camera was programmatically activated, and continuous video acquisition and real-time inference were executed entirely onboard. No post-processing or offline analysis was performed during the experiments.

During both deployments, the platform traversed shallow-water regions indicated in Figure 5b,c. Navigation multi-path waypoints were selected dynamically from the shore station based on real-time visual observations of fish presence, allowing the perception pipeline to be exposed to a wide range of natural viewing conditions. These included variations in water depth, background structure, illumination, and target motion. A fixed and relatively low detection confidence threshold (0.1 in our case) was used to favor recall and capture intermittent target visibility during exploratory operation; detection confidence was not treated as a measure of classification accuracy in this study.

Video data were recorded continuously throughout both deployments at a resolution of 1280×720 pixels and a frame rate capped at 10 frames per second. The recorded video streams, together with the corresponding real-time detection outputs produced onboard, were saved for subsequent analysis. This experiment was designed to validate the feasibility of sustained real-time visual perception with live underwater targets during autonomous operation and to exercise the perception subsystem in end-to-end field scenarios representative of exploratory marine observation and seafloor-mapping missions.

5.2. Demo Application 2—Dynamic Target Re-Localization and Pursuit

This experiment was conducted to validate the ability of the PIRATE system to autonomously alternate between acoustic localization and active pursuit when of a target that undergoes multiple spatial relocations during a single mission. This experiment explicitly tested the system's responsiveness to target movement and its capacity to re-initiate localization and pursuit without external intervention, with the aim of simulating a class

of marine organisms that exhibit the move-loiter pattern. The experiment took place in Ashkelon Marina (Figure 5c).

The experiment consisted of three consecutive tracking–pursuit cycles executed within a single continuous deployment. In each cycle, PIRATE first performed an acoustic triangulation phase by navigating through a predefined tracking geometry while receiving acoustic range measurements from the tag. These measurements were processed onboard to estimate the target location using the single-receiver localization framework. Following target estimation, PIRATE transitioned into a pursuit phase and navigated directly toward the estimated target position.

Upon completion of the first pursuit phase, PIRATE entered a loiter mode in the vicinity of the estimated target position. During this phase, the vehicle continued receiving acoustic transmissions and used standard time-of-flight (ToF) range estimates, together with the previously estimated target location, to continuously assess target–vehicle distance while collecting underwater visual data.

After approximately 60 s (corresponding to roughly six acoustic transmissions), the acoustic tag was manually relocated to a new position within the marina. This relocation was not communicated to the vehicle. The purpose of the displacement was to inspect PIRATE’s performance when the target changes its location after a period of data collection. In contrast to the previous experiment, where the tag was left completely stationary while PIRATE performed its localization maneuver, for this demo application the tag was continuously moved in a pseudo-random manual manner, within a 1 m radius, to emulate slight displacements of a real marine animal. Although far from a complete simulation of a real underwater organism, it provided a better reference for validating PIRATE’s acoustic localization performance on a more realistic target.

The localization–pursuit loop was conducted a total of 3 times. Figure 9 illustrates the three true tag locations and the corresponding triangulation geometries executed in response to each unannounced relocation.

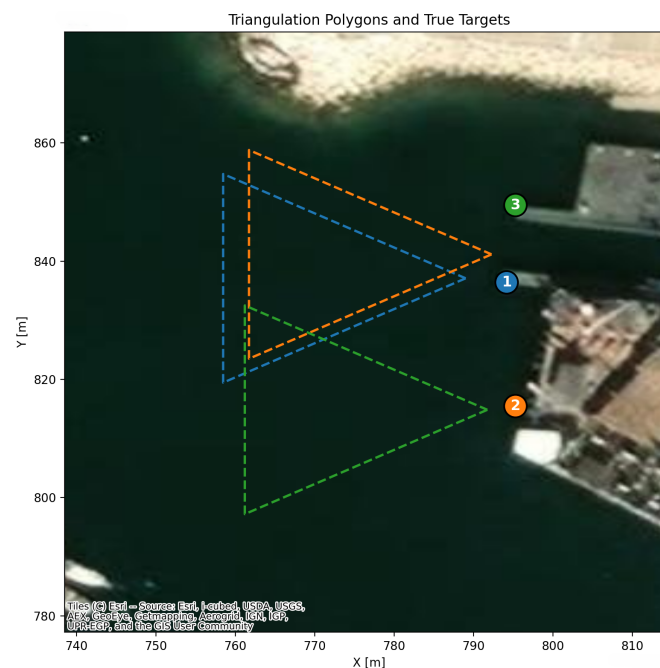


Figure 9. Sequential re-localization experiment in Ashkelon Marina. Dashed polygons indicate the triangulation geometries executed during each localization phase. Colored markers (1–3) denote the true static tag locations for the three consecutive placements. The figure illustrates how distinct tracking polygons were generated in response to unannounced target relocations within a single continuous mission.

6. Results

6.1. Results (Experimental Validation 1)—Autonomous Navigation

Autonomous navigation experiments confirm that PIRATE successfully executed both single-point and multi-point navigation missions under real-world operating conditions. All commanded missions were completed autonomously, including correct activation of the Return-To-Home (RTH) mechanism following deliberately induced communication loss.

Single-Point Navigation

Representative single-point navigation results are shown in Figure 10. The composite visualization presents the planned straight-line route, executed trajectory, platform speed over time (inferred from timed GPS measurements), and the corresponding cross-track error (CTE) distribution for a representative run.

Across eight single-point trials, PIRATE maintained stable forward motion with bounded deviations from the planned path. Aggregate metrics indicate mean cross-track errors on the order of several meters, with larger deviations typically occurring during the mid-portion of longer traverses.

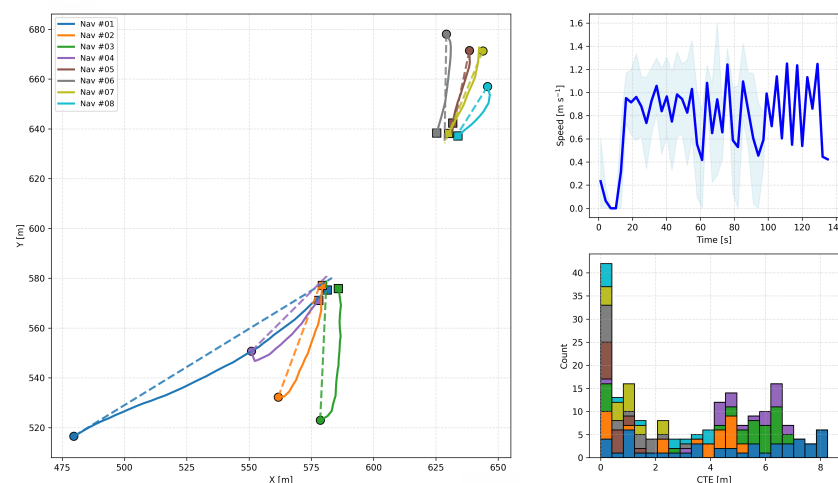


Figure 10. Composite visualization of multiple single-point navigation trials. The (left) panel shows the planned routes (dashed) and the corresponding executed trajectories (solid) for each mission. The (top-right) panel shows the mean platform speed over time, with ± 1 standard deviation across trials (shaded band). The (bottom-right) panel displays the cross-track error (CTE) distributions for all missions (color-coded as the trajectories).

Multi-Point Path Navigation

Figure 11 shows a composite visualization of three repeated multi-point navigation executions following the same predefined waypoint route. The planned trajectory is overlaid with the three executed paths (labeled Nav #01–#03), demonstrating consistent route completion across runs.

Nav #01 and Nav #02 maintained movement patterns throughout the mission, whereas Nav #03 exhibits increased trajectory deviation near mission completion. This behavior coincides with a strong wind event observed during the experiment, indicating PIRATE's ability to complete waypoint sequences under adverse environmental conditions. The corresponding CTE distributions (bottom-right) show that, despite the disturbance observed in Nav #03, overall navigation accuracy remains within 10m bounds, and more importantly, PIRATE shows robustness to such path disturbances.

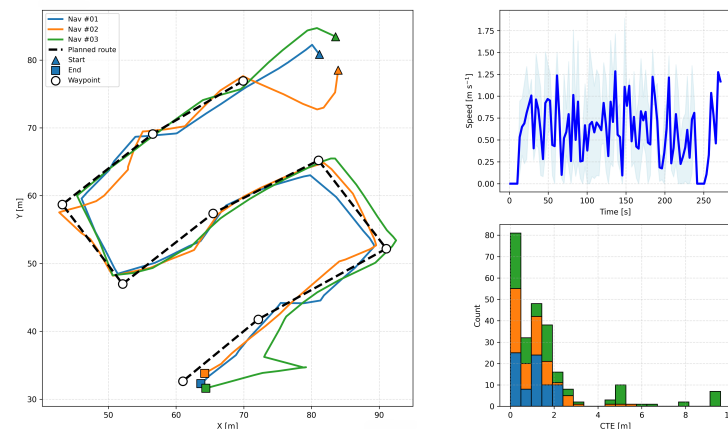


Figure 11. Combined composite of repeated multi-point navigation trials. The left panel shows the planned route and executed trajectories, the top-right panel shows the average speed, and the bottom-right panel presents combined cross-track error (CTE) distributions. Nav #03 illustrates the effect of environmental disturbance near the end of the mission.

A consolidated comparison of navigation performance across both mission types is provided in Table 2. Note that the slightly negative mean excess path observed for single-point missions reflects early mission termination upon entering the predefined waypoint acceptance radius, set to 5 m in our experiments.

Table 2. Aggregated autonomous navigation performance metrics for single-point and multi-point navigation missions.

Mission Type	Trials	Mean Excess Path [%]	Mean Speed [m s^{-1}]	Max Speed [m s^{-1}]	Mean CTE [m]	90th CTE [m]	Max CTE [m]
Single-point	8	−4.24	0.70	1.63	3.25	6.58	8.27
Multi-point	3	1.30	0.61	1.91	1.54	3.33	9.60

Overall, navigation experiments show that PIRATE is capable of performing tight maneuvers and complex multi-point navigation paths, demonstrating the required capabilities for subsequent tracking and target pursuit experiments.

6.2. Results (Experimental Validation 2)—Acoustic Target Localization

The acoustic localization experiments evaluated single-receiver target estimation accuracy as a function of tracking polygon scale. Figure 12 summarizes the tracking missions grouped by polygon edge length, with planned tracking geometries and executed trajectories shown separately.

The top row of Figure 12 presents all planned tracking polygons for each rib-length configuration, together with the true target location (black cross). Each dashed polygon corresponds to an individual tracking mission, with colors used consistently across planned and executed views. For the smallest configuration (edge ≈ 18 m), only a single tracking mission produced a target estimate. The other two experiments failed to produce one because the polygon geometry was too small to provide adequate spatial information, resulting in a single planned polygon for the ≈ 18 m edge length. For larger edge lengths (≈ 24 m, ≈ 30 m, and ≈ 36 m), multiple planned polygons are shown, reflecting repeated tracking missions with comparable geometric scale but varying offsets.

The bottom row of Figure 12 shows the corresponding executed tracking trajectories and final target estimates. Each trajectory and its associated estimate share the same color as the planned polygon above, enabling direct visual association between planning and execution. The true target position is again indicated by a black cross. At intermediate

edge lengths (≈ 24 m), multiple estimates are obtained with moderate dispersion around the true target. As edge length increases (≈ 30 m and ≈ 36 m), the estimates become increasingly tightly clustered, indicating improved robustness and repeatability of the localization process.

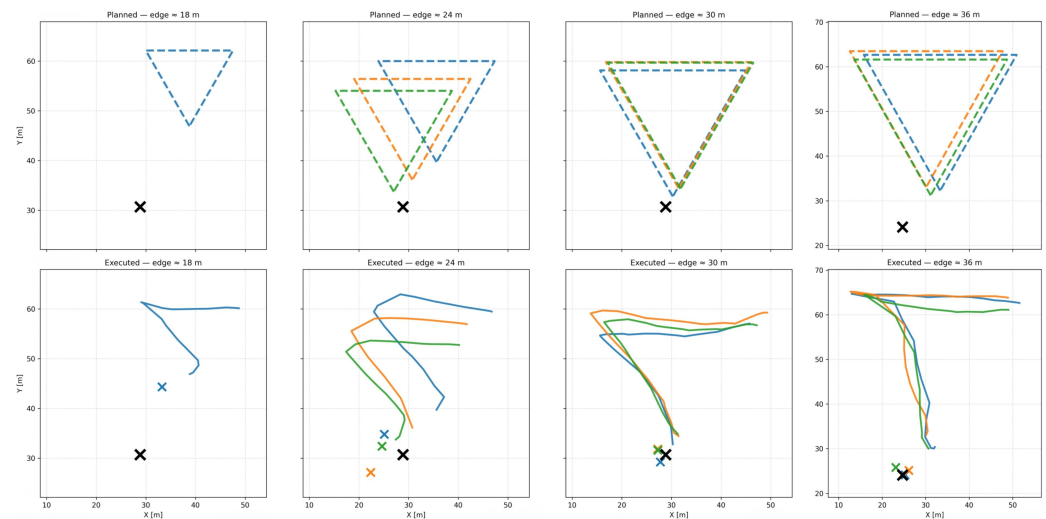


Figure 12. Planned and executed tracking missions grouped by polygon edge length. (**Top**) row: planned tracking polygons (dashed). (**Bottom**) row: executed tracking trajectories and final acoustic target estimates, color-coded to match their corresponding planned polygons. The true target location is shown as a black cross in each subplot.

Table 3 reports the aggregated localization error statistics for each polygon scale. Successful trials indicate those in which a target estimate was produced.

Table 3. Localization error as a function of edge length.

Edge Length [m]	Mean Error [m]	Successful Trials
18	14.33	1
24	5.84	3
30	1.89	3
36	1.55	3

6.3. Results (Experimental Validation 3)—Visual Perception Under Controlled Conditions

Figure 13 summarizes inference throughput as a function of STARK burst length for multiple input resolutions. Solid lines indicate mean frames per second (FPS) aggregated across ten independent runs of the same 1835-frame video sequence. Increasing burst length reduces detector invocation frequency and shifts computation toward tracker-based propagation, producing predictable changes in throughput across resolutions. As shown in the figure, throughput increases monotonically with burst length across all resolutions, with diminishing gains beyond moderate burst values, indicating a practical trade-off between detection refresh rate and computational efficiency.

Quantitative performance statistics for a representative STARK Burst ($N = 8$) are summarized in Table 4. Mean inference rates remain tightly bounded across repeated runs, with minimum and maximum FPS differing by less than 2% in all cases. Mean 95th-percentile latency remains under 60 ms across all resolutions, representing an upper bound for worst-case per-frame processing delay, allowing for continuous real-time processing of at least 16 fps. Complete per-burst performance results for all resolutions are provided in Appendix D.

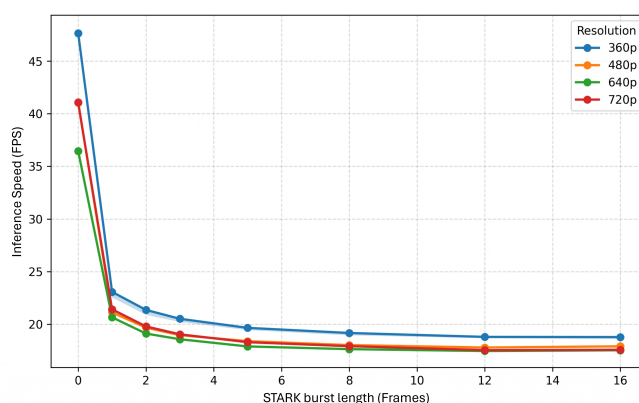


Figure 13. Inference throughput as a function of STARK burst length for multiple input resolutions. Solid lines indicate mean FPS aggregated across ten independent runs of the same 1835-frame video sequence. Variance in all cases was too small to discern visually in the plot.

Table 4. Summary of visual perception performance at a representative STARK burst length ($N = 8$), aggregated across ten runs of the same 1835-frame video sequence. $N = 8$ corresponds to approximately 0.5 s of temporal context at the observed steady-state inference rate (~ 17 – 19 FPS).

Resolution	Burst (Frames)	Mean FPS	Min FPS	Max FPS	Mean p95 Latency (ms)
360p	8	19.16	18.94	19.24	55.78
480p	8	18.01	17.93	18.07	58.38
640p	8	17.63	17.58	17.68	58.63
720p	8	17.91	17.85	17.97	57.48

Thermal measurements collected during prolonged inference execution are shown in Figure 14. GPU temperature exhibits an initial warm-up phase followed by convergence to a steady operating regime during more than 25,000 consecutive inference frames (≈ 40 min of continuous operation). Periodic thermal regulation events (e.g., fan activation and GPU/CPU throttling) were observed but did not coincide with measurable changes in inference throughput or execution stability. This thermal behavior represents an upper bound on real-world behavior, considering that the underwater environment increases heat dissipation.

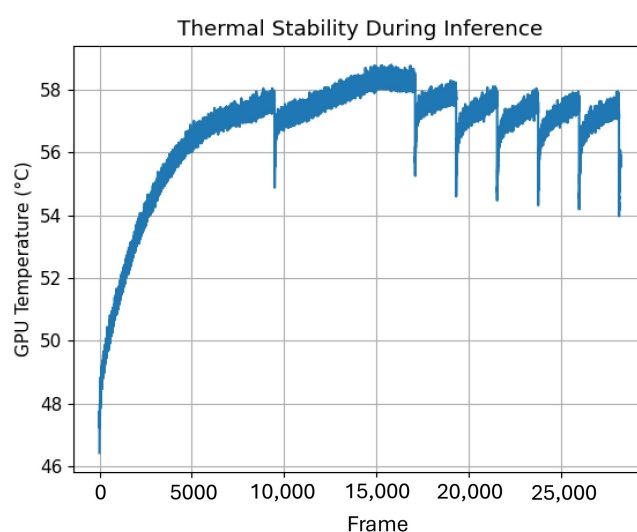


Figure 14. GPU temperature evolution during sustained visual inference on the onboard computing unit during more than 25,000 consecutive frames.

The temporal interaction between detection and tracking within the YOLO–STARK orchestration pipeline is illustrated in Figure 15. The sequence demonstrates how detector invocations periodically reinitialize or correct the target estimate while the tracker propagates the bounding box between detection frames.

Overall, these results demonstrate that the visual perception pipeline operates in a stable, repeatable, and thermally robust manner under sustained embedded execution. The combination of bounded throughput, low tail latency, and consistent detection–tracking behavior across heterogeneous underwater environments provides a reliable foundation for integration into closed-loop autonomous field experiments.

YOLO–STARK Detection–Tracking Sequence

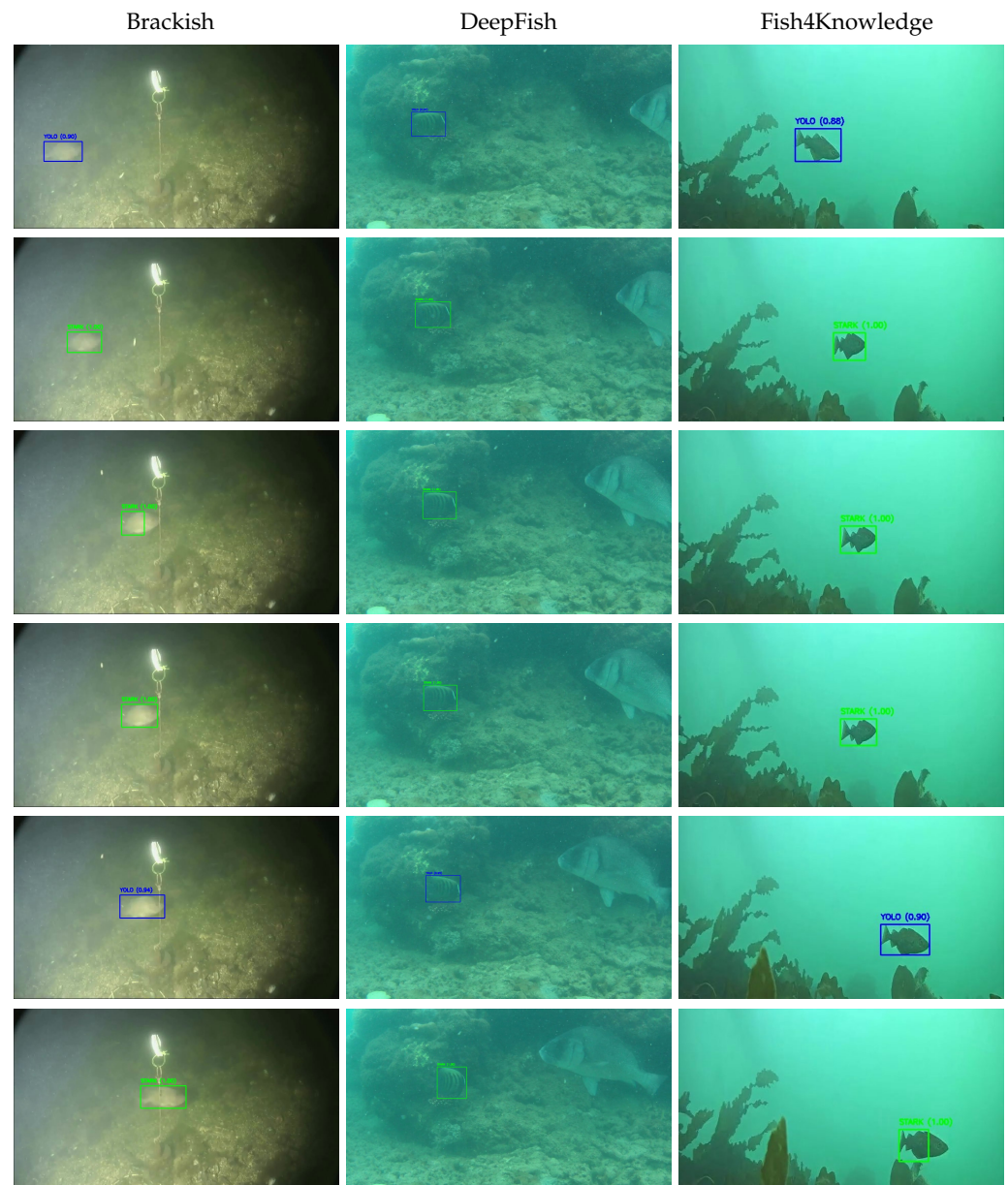


Figure 15. Exemplified temporal behavior of the YOLO–STARK orchestration pipeline with a STARK burst length of $N = 3$. Blue bounding boxes indicate detector-based target initialization or correction, while green bounding boxes denote tracker-based propagation between detector invocations. Frames are ordered chronologically from top to bottom.

6.4. Results (Demo Application 1)—Visual Perception in the Field

As mentioned before, this demo application was meant to demonstrate successful end-to-end operation of the PIRATE platform under real-world marine conditions. Performed at two distinct deployment locations (Sedot Yam and Ashkelon Marina) the platform operated fully autonomously once assigned a multipath mission, while the onboard visual perception pipeline executed continuously on the Jetson computing unit, producing real-time detection outputs from the live GoPro video stream without interruption or manual intervention.

At each site, continuous inference was maintained for the full duration of the autonomous mission, with no observed pipeline stalls, dropped frames, or manual restarts. Across both deployments, several hundred consecutive video frames were processed onboard during autonomous navigation, confirming the feasibility of sustained real-time visual perception in natural shallow-water environments under fully autonomous conditions.

To characterize the behavior of onboard visual perception during field operation, we report the instantaneous number of detections per frame over time for each deployment location. Bounding boxes were extracted directly from onboard inference output using a fixed confidence threshold; frames with no bounding box output were explicitly recorded as zero detections. These metrics are not intended as measures of detection accuracy or precision, but rather to illustrate sustained real-time operation and continuous generation of mission-relevant visual data during exploratory autonomous deployment.

The temporal evolution of visual detections during autonomous field deployments is shown in Figure 16. The observed detection patterns correspond to a perceptually feasible behavioral pattern of underwater animals swimming in packs, where groups intermittently enter and exit the camera field of view, producing short bursts of multiple simultaneous detections followed by quieter intervals.

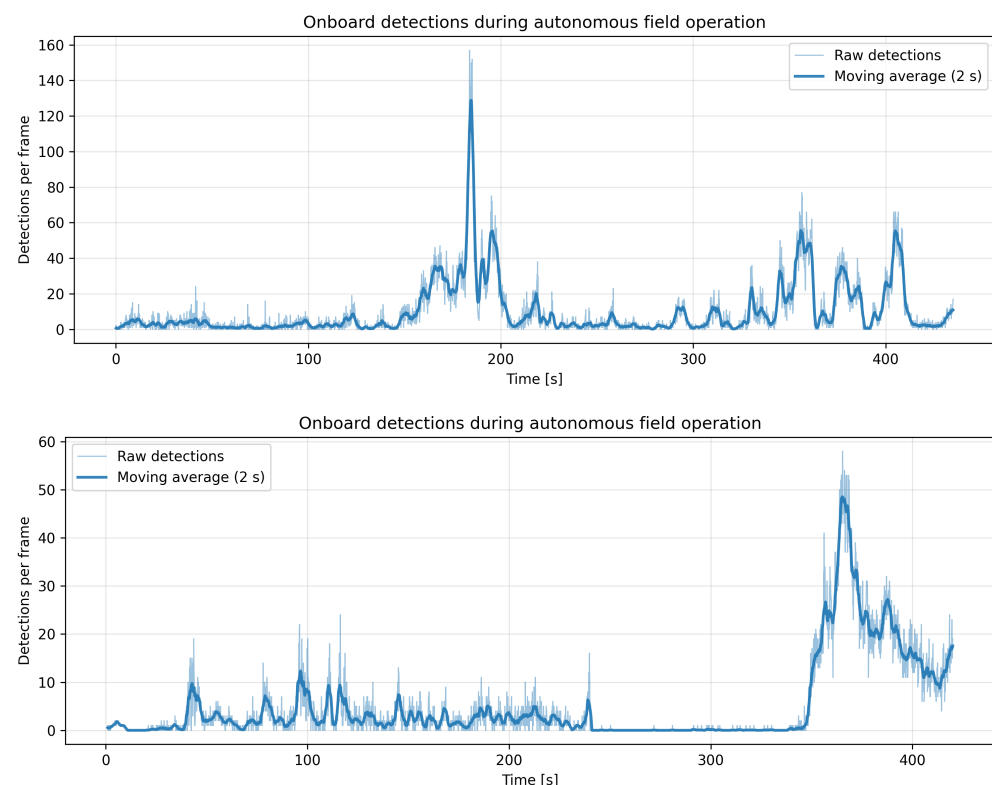


Figure 16. Temporal evolution of onboard visual detections during autonomous field deployments at two locations. **(Top):** Sedot Yam. **(Bottom):** Ashkelon Marina. In both plots, the thin curve shows raw frame-level detection counts, while the thick curve indicates a 2 s moving average to emphasize temporal trends during continuous autonomous operation.

Representative annotated frames from both deployments are shown in Figure 17. The examples illustrate fish detections under varying viewing conditions and site-specific environmental characteristics, including changes in target size, range, background structure, water turbidity, and illumination. Together, the results demonstrate consistent onboard detection performance during continuous autonomous operation across multiple real-world coastal environments.

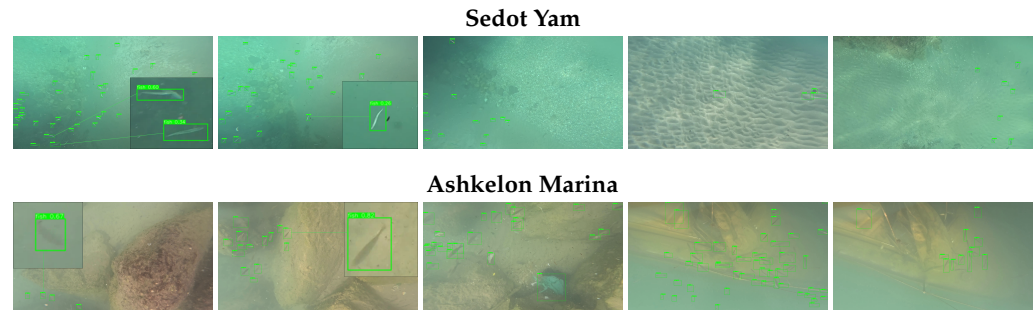


Figure 17. Representative annotated frames recorded during autonomous field visual observation experiments at two deployment sites. Bounding boxes indicate real-time fish detections produced onboard by the visual perception pipeline during continuous autonomous operation. Note the successful detection even under very challenging conditions. Post-processing was applied only to render the real-time detection bounding boxes on the recorded video.

6.5. Results (Demo Application 2)—Dynamic Target Re-Localization and Pursuit

The dynamic re-localization experiment demonstrates stable closed-loop behavior under repeated unannounced target displacement. In all three cycles, PIRATE successfully detected target relocation through deviations in acoustic range measurements, autonomously re-entered the localization phase, re-estimated the target position using the single-receiver triangulation framework, and resumed pursuit. Across the full mission, transitions between loiter, re-localization, and pursuit modes were executed without reset or operator intervention, maintaining meter-scale localization accuracy.

Figure 18a–c present detailed views of each individual localization–pursuit cycle. For every cycle, the planned triangulation polygon is shown together with the executed triangulation path, followed by the planned and executed pursuit trajectories. The bold paths represent executed pursuit segments. In each case, the vehicle closely followed the intended triangulation geometry before transitioning to pursuit.

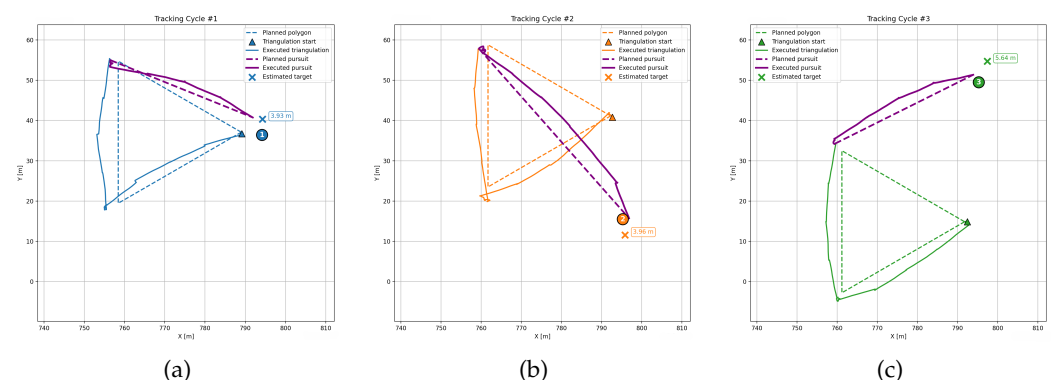


Figure 18. Detailed views of the three consecutive localization–pursuit cycles. Dashed polygons indicate planned triangulation geometry, solid lines show executed triangulation paths, dashed lines represent planned pursuit trajectories, and bold solid lines denote executed pursuit paths. Cross markers indicate estimated target positions and annotated values report the final localization error relative to ground truth.

The annotated values in each subfigure denote the final Euclidean localization error relative to the ground-truth tag location. Across all three cycles, the system maintained consistent meter-level estimation accuracy despite repeated target relocation and pseudo-random tag movement.

Figure 19 presents the complete executed trajectory of PIRATE during the entire dynamic re-localization experiment. The plot illustrates the three localization–pursuit cycles performed within a single uninterrupted deployment. Each color corresponds to a distinct target placement, showing the executed triangulation maneuver followed by the corresponding pursuit trajectory toward the estimated target location. The execution order of the cycles is blue → orange → green.. The continuous trajectory confirms seamless transitions between successive re-localization and pursuit phases throughout the mission.

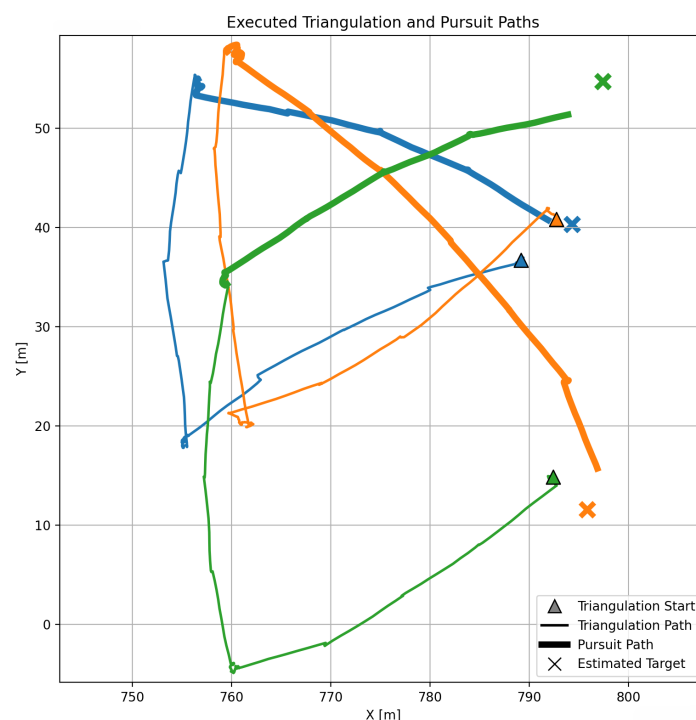


Figure 19. Full executed trajectory of the dynamic re-localization experiment. Colored paths represent the three consecutive localization–pursuit missions executed within a single continuous deployment. The plot does not include the loiter location sequences, where PIRATE was only affected by wind and current, while capturing underwater video data. The execution order is blue → orange → green.

The camera was successfully initiated, and visual data was captured in the loitering stage of each tracking cycle, after the target was deemed close enough for visual data perception. Representative images from each loiter location are presented in Figure 20.



Figure 20. Representative images captured from each loiter location.

7. Discussion and Conclusions

This work presented PIRATE, an autonomous surface platform in which navigation, acoustic sensing, and visual perception are performed and coordinated through onboard

control logic and artificial intelligence. Rather than operating as isolated subsystems, these components are integrated in a closed-loop architecture that allows sensing and motion to be conditioned on mission state and available information. This design enables the platform to execute deliberate acoustic localization maneuvers, maintain stable navigation, and perform real-time visual inference without reliance on fixed sensing infrastructure or continuous shore-side supervision.

Field experiments have shown that the integrated architecture operates reliably under real-world marine conditions. In particular, PIRATE maintained stable navigation during structured listening maneuvers, achieved meter-scale single-receiver localization under appropriate geometric configurations, and sustained continuous onboard visual inference during autonomous operation. These results validate the technical feasibility of coordinating motion, acoustic sensing, and perception through onboard control logic.

The demonstration applications further showed that this integration can support mission-level behavior with scientific relevance. In these scenarios, the platform detected target displacement, autonomously re-initiated localization, and resumed pursuit without operator intervention, while maintaining real-time perception. Together, these demonstrations illustrate that PIRATE is capable not only of executing individual subsystems in isolation, but of performing coherent, adaptive, target-driven missions representative of practical field deployments.

7.1. From Survey Autonomy to Target-Driven Autonomy

Most autonomous marine platforms excel at repeatable survey execution, but typically remain limited in their ability to redirect sensing effort when the environment (or the target) changes. PIRATE's core architectural shift is that acoustic localization is not merely logged for offline analysis; it actively drives vehicle behavior and gates additional sensing. In particular, PIRATE's single-receiver approach achieves spatial diversity through deliberate motion and successive receptions, enabling infrastructure-free tracking without the deployment footprint of fixed synchronized arrays. This matters for biological observation as many ecologically relevant targets exhibit "move-loiter" dynamics: extended localized residency punctuated by displacement. PIRATE is designed around this regime, where autonomy must exploit loiter windows to collect high-value visual data while remaining capable of re-localizing after movement.

7.2. System-Level Performance

The experimental suite performed on PIRATE supports three system-level claims, which together establish the capacities embodied in its name, namely a real-time autonomous tracker and explorer:

Operational autonomy under field conditions: The navigation experiments establish a stable motion and control baseline on which sensing-driven behaviors can be layered, including reliable mission completion under the tested disturbances and the ability to execute structured listening geometries.

Practical single-receiver localization: The results show that trajectory geometry is a first-order determinant of localization quality: sufficiently large listening polygons yield consistent meter-scale estimates for stationary targets, while small geometries can fail to provide enough information. This explicitly characterizes a key trade-off between footprint/duration and accuracy that is directly actionable for mission design.

Closed-loop re-localization, pursuit and visual perception without operator intervention: The dynamic re-localization experiment is particularly important because it demonstrates autonomous state transitions driven by onboard signals rather than scripted phases: PIRATE detects displacement through sustained range deviations, re-enters triangulation,

and resumes pursuit across multiple unannounced target moves in a single continuous mission. This is a concrete step beyond *autonomous data collection* toward *autonomous interaction* with a moving, partially observed target.

7.3. Implications for Underwater Data Collection in the Intended Target Class

PIRATE's primary value for underwater data collection is not only that it can localize, but that it can allocate sensing effort where it is most likely to produce interpretable observations. Acoustic localization provides a coarse-to-fine mechanism: it supports search, reduces the spatial uncertainty needed to make visual sensing viable, and provides a persistent cue for detecting when the target has departed the vicinity.

This architecture is well aligned with shallow-water targets that are (a) taggable with periodic acoustic transmissions, and (b) likely to produce visually meaningful data when the platform approaches within a practical range. In this operating regime, PIRATE can convert intermittent acoustic receptions into directed visual opportunities, enabling datasets that capture behavior and context rather than only sparse positions.

7.4. Advantages Relative to Common Alternatives

PIRATE's benefits can be summarized by contrasting the operational constraints of three common approaches:

1. Advantages over fixed receiver arrays: Synchronized multi-receiver arrays can deliver high accuracy but impose deployment cost, a fixed spatial footprint, and limited portability. PIRATE trades instantaneous spatial diversity for motion-generated diversity, preserving TDoA-style constraints while removing the need for distributed infrastructure.
2. Advantages over human-in-the-loop visual observation: Diver/ROV-style workflows can produce high-quality imagery but require continuous supervision and typically lack repeatable closed-loop behavior. PIRATE demonstrates continuous onboard perception during autonomous operation, and onboard decision making that relies on visual perception, positioning vision as an operational modality rather than a post-mission add-on.
3. Advantages over survey-oriented autonomy: Waypoint-following survey USVs or UUVs are effective for coverage but are not inherently target-contingent. PIRATE's distinguishing capability is stateful coordination: it can select listening trajectories, pursue, loiter, and re-localize based on onboard evidence, enabling persistent target-driven observation over a continuous mission.

7.5. Future Work

Like most methodological tools and technologies, our presented PIRATE is not necessarily a final stepping stone but better be considered an inaugural standard leading to future improvements and developments. The following are several such directions inspired by observed limitations in PIRATE's current operational envelope.

Hydrodynamic Disturbance and Navigation Robustness: All experiments were conducted under mild to moderate sea conditions (wave height < 0.5 m, wind speed ≤ 10 m/s). Within this regime, navigation errors remained bounded and acoustic localization achieved meter-scale accuracy (Tables 1 and 2). Wind, surface currents, and wave-induced yaw moments affect trajectory repeatability and listening-geometry consistency. The current platform relies on passive catamaran stability and differential thrust control without adaptive disturbance rejection. Operation in harsher sea states (e.g., wave height > 1 m, wind speed > 15 m/s) is expected to require robust or adaptive control strategies and improved disturbance-aware state estimation, as explored in recent adaptive and prescribed-

performance control frameworks for marine vehicles [38,39]. Accordingly, the validated operational envelope is presently limited to sheltered and nearshore environments. A more robust marine-grade platform incorporating enhanced structural stability and disturbance-aware control is currently under development, and future experiments will evaluate system performance under significantly harsher sea conditions now that the core autonomy principles and scientific applications have been validated.

Onboard obstacle detection and avoidance: PIRATE currently lacks dedicated onboard obstacle detection and reactive avoidance capabilities. Successful deployments thus depend on conservative site selection and pre-mission planning to mitigate collision risk. The platform cannot autonomously detect or avoid hazards such as moorings, debris, shorelines, or sea-level traffic (vessels or humans), limiting operation in cluttered or dynamic environments. A priority direction is the integration of dedicated perception hardware, including a mast- or top-mounted wide-FOV camera (optionally complemented by side cameras) and, where optical sensing is unreliable, lightweight ranging sensors. While slightly increasing cost, these sensors and inputs should support a layered safety stack with geofencing and stand-off constraints, together with a local reactive controller capable of detouring around obstacles and smoothly rejoining planned trajectories.

Mission and behavior control: Several decision points in the current implementation rely on fixed thresholds, including listening polygon scale, detection confidence levels, and state-switching distances. These values were selected to enable controlled validation of the closed-loop autonomy framework under stable environmental conditions. However, environmental variability—such as acoustic attenuation, reception intermittency, turbidity, and current-induced drift—may shift the optimal operating regime. The modular FSM-based architecture allows integration of adaptive mechanisms based on online confidence and uncertainty estimates. Metrics such as acoustic reception rate, localization residual statistics, geometric conditioning indicators, detection confidence trends, or drift magnitude could be used to dynamically adjust listening geometry, localization residuals, threshold values, or behavioral transitions (e.g., expanding the polygon, loitering, pursuing, or reinitializing). Learning-based approaches may further be explored to optimize action selection with respect to mission success and data-collection quality.

Energy budget and endurance: Given the limited energy budget onboard, repeated localization maneuvers and extended perception runs constrain mission duration. Extending endurance is therefore a key objective. One promising avenue is the integration of deck-mounted marine-grade solar panels combined with appropriate charge control and energy monitoring, enabling opportunistic energy harvesting during low-speed observation or loitering phases without increasing battery mass.

Scalability to multi-vehicle operation: A natural extension of PIRATE is a coordinated multi-vehicle fleet operating as a collaborative swarm. Vehicles could divide functional roles (search, acoustic listening, close-range visual observation, communications relay) and share hypotheses to adapt behavior online, improving coverage, reacquisition speed, persistence, and redundancy while preserving infrastructure-free deployment.

7.6. Conclusions

PIRATE should be viewed primarily as a system-level contribution: a deployable autonomy framework that unifies mobility, acoustic telemetry, and visual perception into a single operational capability for target-oriented underwater observation. By treating sensing and motion as coupled resources governed by onboard state and evidence, the platform establishes a pathway from sparse, intermittent underwater signals to purposeful, repeatable data acquisition behaviors that can be executed in the field without fixed infrastructure or continuous human control. This integration-centric design differentiates

PIRATE from survey-first systems and from observation tools that depend on external instrumentation, and it provides a modular basis for extending capability through safety layers (obstacle-aware autonomy), stronger decision policies, endurance enhancements (including onboard solar recharging), and coordinated multi-vehicle operation. As these directions are developed, PIRATE can evolve from a single-platform demonstrator into a scalable approach for persistent, high-value underwater data collection in regimes where targets are mobile, partially observed, and best characterized by combining coarse acoustic cues with close-range visual evidence.

Author Contributions: Conceptualization, D.Z. and O.B.-S.; Methodology, D.Z. and O.B.-S.; Software, D.Z.; Validation, D.Z. and O.B.-S.; Formal analysis, D.Z.; Investigation, D.Z.; Resources, O.B.-S.; Data curation, D.Z.; Writing—original draft, D.Z.; Writing—review & editing, D.Z. and O.B.-S.; Visualization, D.Z.; Supervision, O.B.-S.; Project administration, O.B.-S.; Funding acquisition, O.B.-S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded in part by the Israel Science Foundation under Grants No. 501/25 and 504/25.

Data Availability Statement: The original data presented in the study are openly available in the project web page at <https://icvl.cs.bgu.ac.il/pages/researches/PIRATE.html> (accessed on 25 February 2026).

Acknowledgments: We would like to acknowledge The River Park in Be'er Sheva, The Interuniversity Institute for Marine Sciences in Eilat, Kadas Nexus, Leon H. Charney School of Marine Sciences, University of Haifa, Israel, and Ashkelon Marina for providing us with suitable locations for field experiments and demo applications. We would also like to thank Moshe Kiflawi for constructive discussions about the desired specifications of PIRATE-like devices. During the preparation of this manuscript/study, the authors used ChatGPT 5.2 and Gemini 3 for the purposes of code writing, code evaluation, latex format generation, and native English-grade text generation. These materials were reviewed, edited, and revised by the authors, who take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Appendix A. Detailed Hardware Specifications

Appendix A.1. USV Platform Parameters

This table the primary physical, propulsion, and power-related characteristics of the PIRATE unmanned surface vehicle that are relevant to autonomous operation and payload integration.

Table A1. Detailed physical parameters of the PIRATE USV.

Parameter	Value
Length	1.20 m
Width	0.88 m
Height	0.40 m
Draft	0.10 m
Weight (without payload)	16 kg
Additional payload capacity	20 kg
Ingress protection rating	IP67
Main Motor Battery	
Operating voltage range	7–20 V
Battery capacity	12 Ah (177.6 Wh)
Nominal battery voltage	14.8 V

Appendix A.2. Sensing and Communication Hardware

Table A2 lists the detailed specifications of the acoustic receiver, visual sensing unit, and short-range communication hardware integrated into the PIRATE platform.

Table A2. Detailed sensing and communication hardware specifications.

Parameter	Value
Acoustic Receiver	
Manufacturer	Thelma Biotel
Model	TB Live Receiver
Mounting depth	0.12 m
Dry weight (with enclosure)	545 g
Power supply	3.6–9 V
Supported frequency range	63–77 kHz
Visual Sensor	
Manufacturer	GoPro
Model	13 Hero Black
Maximum resolution	5312 × 2988
Maximum frame rate	Up to 240 fps
Field of view	122°/177° w. extra lens
Communication	
Wireless interface	Dual-band Wi-Fi
Operating frequency	2.4 GHz/5 GHz
Nominal range	~50 m
Protocol	UDP

Appendix A.3. GPU-Enabled Computing Unit

This subsection reports the technical specifications of the embedded GPU-enabled computing unit used for onboard visual inference and real-time perception tasks.

Table A3. Technical specifications of the embedded GPU computing unit.

Parameter	Value
Manufacturer	NVIDIA
Model	Jetson Orin NX
Processor	6-core ARM Cortex-A78AE
GPU architecture	NVIDIA Ampere
CUDA cores	1024
Tensor cores	32
Memory	8 GB LPDDR5
Power modes	10–25 W
Peak AI performance	Up to 70 TOPS (INT8)

Appendix B. Background on TDoA-Based and Single-Receiver Acoustic Localization

This appendix summarizes the geometric principles underlying time-difference-of-arrival (TDoA) localization and its extension to single-receiver configurations. The material is provided for completeness and background reference and is not specific to the PIRATE system implementation.

Appendix B.1. Classical TDoA Localization

In classical TDoA-based localization, acoustic signals emitted by a transmitter are received at different times by multiple synchronized receivers with known positions. Differ-

ences in reception times define hyperbolic constraints on the possible transmitter location, and the intersection of multiple such constraints yields an estimate of the target position.

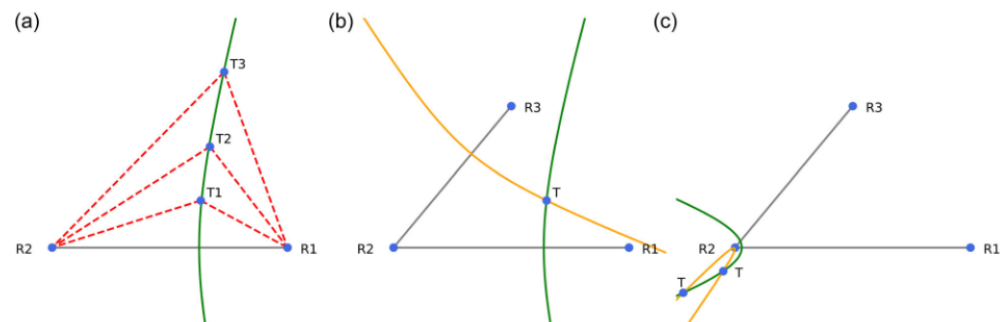


Figure A1. Time Difference of Arrival (TDoA) principles [29]. (a) Given the known locations of receivers R_1 and R_2 and a specified range difference, the possible locations of a transmitter T lie on a hyperbola. (b) The intersection of two hyperbolae defined by receiver pairs (R_1, R_2) and (R_2, R_3) yields candidate transmitter locations. (c) Large range differences may result in multiple feasible solutions.

The geometric constraint underlying TDoA localization can be expressed as

$$\|\mathbf{x} - \mathbf{r}_i\| - \|\mathbf{x} - \mathbf{r}_j\| = c(t_i - t_j), \quad (\text{A1})$$

where $\mathbf{x}$ denotes the unknown target position, $\mathbf{r}_i$ and $\mathbf{r}_j$ are receiver positions, t_i and t_j are the corresponding reception times, and c is the speed of sound in water. Each receiver pair contributes a hyperbolic constraint, and increased geometric diversity improves localization accuracy.

Appendix B.2. Single-Receiver Localization via Receiver Motion

Single-receiver localization replaces spatially distributed receivers with a single mobile receiver whose motion provides the required spatial diversity. TDoA measurements are computed across successive receptions obtained at different receiver positions over time. Each pair of receptions defines a hyperbolic constraint, and the intersection of multiple such constraints yields a target position estimate.

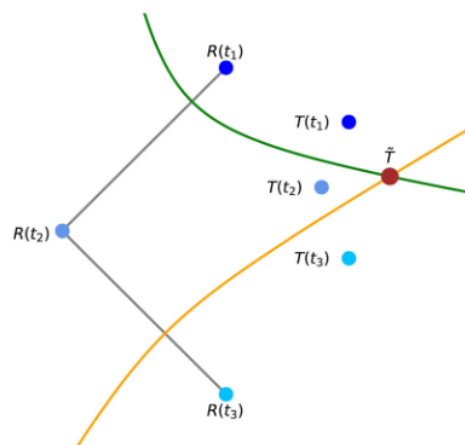


Figure A2. Illustration of single-receiver TDoA localization using a mobile receiver [29]. Receiver positions $R(t)$ correspond to successive receptions of transmissions from the target $T(t)$. Each reception pair $(R(t_i), R(t_{i+1}))$ defines a TDoA hyperbola, and target position estimates are obtained by intersecting constraints from multiple reception pairs.

Appendix C. Benchmark Performance of Detection and Tracking Models

This appendix summarizes representative benchmark performance of the object detection and visual tracking models used in PIRATE, as reported in their original publications. The results are included for reference to motivate model selection and are not intended for direct comparison with field experiments conducted in this study.

Table A4. Representative benchmark performance of detection and tracking models used in PIRATE, reported on standard datasets in the original publications.

Model	Task	Dataset	mAP@[0.5:0.5:0.95]	AP@0.5	AO	Success	Precision
YOLOv7	Detection	COCO 2017	51.4	69.7	–	–	–
YOLOv7-Tiny	Detection	COCO 2017	37.5	56.8	–	–	–
YOLOv7	Detection	Pascal VOC 2007	–	83.4	–	–	–
YOLOv7-Tiny	Detection	Pascal VOC 2007	–	72.1	–	–	–
STARK-S	Tracking	GOT-10k	–	–	67.1	–	–
STARK-ST	Tracking	GOT-10k	–	–	68.8	–	–
STARK-S	Tracking	LaSOT	–	–	–	66.4	77.8
STARK-ST	Tracking	LaSOT	–	–	–	67.1	78.1
STARK-S	Tracking	TrackingNet	–	–	–	81.3	87.2
STARK-ST	Tracking	TrackingNet	–	–	–	82.0	88.3

Appendix D. Per-Burst Visual Perception Performance

This appendix provides complete per-burst performance results for the visual perception pipeline evaluated in Experiment 3. Results are reported for all tested input resolutions and STARK burst lengths, aggregated across ten independent runs of the same 1835-frame video sequence. Metrics include mean inference throughput, observed minimum and maximum mean FPS across runs, and mean 95th-percentile per-frame latency, providing a comprehensive view of performance trends and absolute bounds under identical workload conditions.

Appendix D.1. 360p Resolution

Table A5. Per-burst visual perception performance at 360p resolution.

Burst	Mean FPS	Min FPS	Max FPS	Mean p95 Latency (ms)
0	47.65	46.90	47.84	20.21
1	23.07	22.54	23.16	53.76
2	21.36	20.89	21.44	55.14
3	20.51	20.21	20.58	55.64
5	19.65	19.44	19.72	56.01
8	19.16	18.94	19.24	55.78
12	18.79	18.66	18.88	55.81
16	18.78	18.60	18.86	55.22

Appendix D.2. 480p Resolution

Table A6. Per-burst visual perception performance at 480p resolution.

Burst	Mean FPS	Min FPS	Max FPS	Mean p95 Latency (ms)
0	41.05	40.78	41.28	23.10
1	21.12	21.09	21.17	56.40
2	19.67	19.61	19.73	57.97
3	18.96	18.93	18.99	58.58
5	18.39	18.35	18.43	58.62
8	18.01	17.93	18.07	58.38
12	17.78	17.75	17.85	58.18
16	17.90	17.82	17.94	57.25

Appendix D.3. 640p Resolution

Table A7. Per-burst visual perception performance at 640p resolution.

Burst	Mean FPS	Min FPS	Max FPS	Mean p95 Latency (ms)
0	36.45	36.22	36.55	25.43
1	20.65	20.56	20.69	56.49
2	19.11	19.05	19.15	58.32
3	18.57	18.50	18.61	58.73
5	17.89	17.84	17.94	59.19
8	17.63	17.58	17.68	58.63
12	17.45	17.41	17.50	58.26
16	17.52	17.46	17.55	57.44

Appendix D.4. 720p Resolution

Table A8. Per-burst visual perception performance at 720p resolution.

Burst	Mean FPS	Min FPS	Max FPS	Mean p95 Latency (ms)
0	41.06	40.77	41.29	21.79
1	21.41	21.36	21.46	55.10
2	19.79	19.71	19.85	56.82
3	19.04	19.00	19.06	57.54
5	18.29	18.26	18.32	57.90
8	17.91	17.85	17.97	57.48
12	17.53	17.50	17.57	57.71
16	17.57	17.48	17.64	56.96

Note on Resolution-Dependent Performance.

A slight non-monotonic performance relationship is observed in regard to the 720p configurations, with marginally higher throughput at 720p for several STARK burst lengths. This behavior is attributable to GPU kernel alignment effects and preprocessing overhead on embedded hardware, where performance does not scale strictly with input pixel count. Such effects are commonly observed in practice and are consistent with the tight run-to-run variability and stable execution behavior reported in this study.

References

- Hassoun, A.E.R.; Tanhua, T.; Lips, I.; Heslop, E.; Petihakis, G.; Karstensen, J. The European Ocean Observing Community: Urgent gaps and recommendations to implement during the UN Ocean Decade. *Front. Mar. Sci.* **2024**, *11*, 1394984. [\[CrossRef\]](#)
- Bell, K.L.C.; Johannes, K.N.; Kennedy, B.R.C.; Poulton, S.E. How little we've seen: A visual coverage estimate of the deep seafloor. *Sci. Adv.* **2025**, *11*, eadp8602. [\[CrossRef\]](#) [\[PubMed\]](#)
- Hussey, N.E.; Kessel, S.T.; Aarestrup, K.; Cooke, S.J.; Cowley, P.D.; Fisk, A.T.; Harcourt, R.G.; Holland, K.N.; Iverson, S.J.; Kocik, J.F.; et al. Aquatic animal telemetry: A panoramic window into the underwater world. *Science* **2015**, *348*, 1255642. [\[CrossRef\]](#) [\[PubMed\]](#)
- Hays, G.C.; Ferreira, L.C.; Sequeira, A.M.M.; Meekan, M.G.; Duarte, C.M.; Bailey, H.; Bailleul, F.; Bowen, W.D.; Caley, M.J.; Costa, D.P.; et al. Key questions in marine megafauna movement ecology. *Trends Ecol. Evol.* **2016**, *31*, 463–475. [\[CrossRef\]](#)
- Garcia, C.A.; Barone, B.; Ferrón, S.; Poulos, S.; Karl, D.M. A multi-year, seasonally resolved dataset of Seaglider observations at Station ALOHA in the North Pacific Subtropical Gyre. *Sci. Data* **2025**, *12*, 641. [\[CrossRef\]](#)
- Font, E. Autonomous underwater gliders to observe the ocean. *Nat. Rev. Earth Environ.* **2024**, *5*, 610. [\[CrossRef\]](#)
- Lai, Y.L.; Ang, T.F.; Bhatti, U.A.; Ku, C.S.; Han, Q.; Por, L.Y. Color correction methods for underwater image enhancement: A systematic literature review. *PLoS ONE* **2025**, *20*, e0317306. [\[CrossRef\]](#) [\[PubMed\]](#)
- Nabahirwa, E.; Song, W.; Zhang, M.; Fang, Y.; Ni, Z. A Structured Review of Underwater Object Detection Challenges and Solutions: From Traditional to Large Vision Language Models. *arXiv* **2025**, arXiv:2509.08490. [\[CrossRef\]](#)
- Bangley, C.W.; Cudney, J.L.; Dell'Apa, A.; Diaddorio, E.; Rulifson, R.A. Development of a continental shelf acoustic telemetry array to support behavioral research of fish in a high energy ocean environment. *Fish. Res.* **2022**, *247*, 106177. [\[CrossRef\]](#)

10. Reyes-Betancort, J.A.; Aspillaga, E.; Martorell-Barceló, M.; Follana-Berná, G.; Arlinghaus, R. Performance of a novel system for high-resolution tracking of marine fish societies. *Anim. Biotelem.* **2021**, *9*, 1. [\[CrossRef\]](#)
11. Gadre, A.S.; Maczka, D.K.; Spinello, D.; McCarter, B.R.; Stilwell, D.J.; Neu, W.; Roan, M.J.; Hennage, J.B. Cooperative localization of an acoustic source using towed hydrophone arrays. In Proceedings of the 2008 IEEE/OES Autonomous Underwater Vehicles, Woods Hole, MA, USA, 13–14 October 2008. [\[CrossRef\]](#)
12. Campos, R.; Gracias, N.; Ridao, P. Underwater multi-vehicle trajectory alignment and mapping using acoustic and optical constraints. *Sensors* **2016**, *16*, 387. [\[CrossRef\]](#)
13. Yu, Y.; He, C.; Sun, S.; Liu, Z. Error Analysis on Ultra-Short Baseline Positioning System. *J. Phys. Conf. Ser.* **2023**, *2486*, 012058. [\[CrossRef\]](#)
14. Yang, Z.; Fu, J.; Sun, Y.; Li, Y. A Pseudo-Trajectory Homotopy Method for UUVs Information Collection IoT System With an Underwater Communication Constraint. *IEEE Internet Things J.* **2025**, *12*, 43255–43267. [\[CrossRef\]](#)
15. Zhu, H.; Xia, X.; Zhang, Y.; Yu, H.; Fan, S.; Gao, W. Event-Driven Control Approach of Cooperative Localization-Formation Hybrid for Internet of Multi-USV. *IEEE Trans. Aerosp. Electron. Syst.* **2026**, *early access*. [\[CrossRef\]](#)
16. Larson, S.; Christiansen, J.; Olsen, A.Y.; Walsh, W.J.; Teague, C.H.; Tissot, B.; Randell, Z. A Unique 100 Meter Underwater Survey Method Documents Changes in Abundance, Richness, and Community Structure of Hawai'i Reef Fishes. *Front. Mar. Sci.* **2022**, *9*, 892261. [\[CrossRef\]](#)
17. Schram, M.J.; Emory, M.E.; Kilborn, J.P.; Peake, J.A.; Wall, K.R.; Williams, I.; Stallings, C.D. Reef fish assemblages differ both compositionally and functionally on artificial and natural reefs in the eastern Gulf of Mexico. *ICES J. Mar. Sci.* **2024**, *81*, 1150–1163. [\[CrossRef\]](#)
18. Gröger, J.P.; Cisewski, B.; Badri-Hoeher, S.; Boos, K.; Clemmesen, C.; Dauben, V.; Hoeher, P.A. Development and operation of a novel non-invasive opto-acoustic underwater fish observatory in Kiel Bight, Southwestern Baltic Sea. *Front. Mar. Sci.* **2024**, *11*, 1425259. [\[CrossRef\]](#)
19. Tibone, M.; Francescangeli, M.; Stefanni, S.; Aguzzi, J.; O'Neill, B.; Del Rio, J.; Toma, D.M.; Mirimin, L. Enhancing fish community monitoring at a cabled observatory by combining environmental DNA and imaging analysis. *J. Fish Biol.* **2025**, *Early View*. [\[CrossRef\]](#)
20. Kanashiro, R.A.; Paiva, J.C.P.; Nunes, W.R.B.M.; de Melo, L.F. Remotely Operated and Autonomous Underwater Vehicles in Offshore Wind Farms: A Review on Applications, Challenges, and Sustainability Perspectives. *Sustainability* **2026**, *18*, 2. [\[CrossRef\]](#)
21. Duraš, A.; Ilioudi, A.; Wolf, B.J.; Palunko, I.; De Schutter, B. A Dataset for Detection and Segmentation of Underwater Marine Debris in Shallow Waters. *Sci. Data* **2024**, *11*, 921. [\[CrossRef\]](#)
22. Ohrem, S.J.; Haugaløkken, B.O.A.; Kelasidi, E. SOLAQUA: SINTEF Ocean Large Aquaculture Robotics Dataset. *arXiv* **2025**, arXiv:2504.01790. [\[CrossRef\]](#)
23. Handegård, N.O. Uncrewed Surface Vehicles (USVs) as Platforms for Fisheries and Plankton Acoustics. *ICES J. Mar. Sci.* **2024**, *81*, 1712–1722. [\[CrossRef\]](#)
24. Lu, Z.; Li, W.; Zhang, X.; Wang, J.; Zhuang, Z.; Liu, C. Design and Testing of an Autonomous Navigation Unmanned Surface Vehicle for Buoy Inspection. *J. Mar. Sci. Eng.* **2024**, *12*, 819. [\[CrossRef\]](#)
25. Hardy, R.F.; Meylan, A.B.; Gray, J.A.; Meylan, P.A. Daily, seasonal, and long-distance movements inferred from Fastloc-GPS telemetry of immature green turtles (*Chelonia mydas*) at a high-latitude, mid-ocean developmental site. *PLoS ONE* **2023**, *18*, e0292235. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Zeller, D.C. Home range and activity patterns of the coral trout *Plectropomus leopardus* (Serranidae). *Mar. Ecol. Prog. Ser.* **1997**, *154*, 65–77. [\[CrossRef\]](#)
27. Flowers, K.I.; Heithaus, M.R.; Papastamatiou, Y.P. Buried in the sand: Uncovering the ecological roles and importance of rays. *Fish Fish.* **2021**, *22*, 105–127. [\[CrossRef\]](#)
28. Klumpp, D.W.; McKinnon, A.D.; Daniel, P. Damselfish territories: Zones of high productivity on coral reefs. *Mar. Ecol. Prog. Ser.* **1987**, *40*, 41–51. [\[CrossRef\]](#)
29. Git, I.; Samina, M.; Givon, S.; Segev, R.; Kiflawi, M.; Ben-Shahar, O. Acoustic Tracking of Moving Marine Targets Using a Single Autonomous Surface Receiver. *J. Field Robot.* **2024**, *41*, 1498–1520. [\[CrossRef\]](#)
30. Gharsa, O.; Toubia, M.M.; Boumehraz, M.; Agram, N. Autonomous Vision-Based Object Detection and Tracking System for Quadrotor Unmanned Aerial Vehicles. *Sensors* **2025**, *25*, 6403. [\[CrossRef\]](#)
31. Ji, D.; Ren, S.; Zheng, R.; Yi, R.; Zhao, H.; Lin, Y. A tracking control method of ASV following AUV. In Proceedings of the 2013 OCEANS, San Diego, CA, USA, 23–27 September 2013; pp. 1–4.
32. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788. [\[CrossRef\]](#)

33. Wang, C.Y.; Bochkovskiy, A.; Liao, H.Y.M. YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 17–24 June 2023; pp. 7464–7475.
34. Yan, B.; Peng, H.; Fu, J.; Wang, D.; Lu, H. Learning Spatio-Temporal Transformer for Visual Tracking. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Virtual, 11–17 October 2021; pp. 10448–10457.
35. Fisher, R.B.; Chen-Burger, Y.H.; Giordano, D.; Hardman, L.; Lin, F.P. *Fish4Knowledge: Collecting and Analyzing Massive Coral Reef Fish Video Data*; Intelligent Systems Reference Library; Springer: Berlin/Heidelberg, Germany, 2016; Volume 104. [[CrossRef](#)]
36. Alzayat, S.; Laradji, I.H.; Konovalov, D.A.; Bradley, M.; Vazquez, D.; Sheaves, M. A Realistic Fish-Habitat Dataset to Evaluate Algorithms for Underwater Visual Analysis. *arXiv* **2020**, arXiv:2008.12603. [[CrossRef](#)]
37. Pedersen, M.; Bruslund Haurum, J.; Gade, R.; Moeslund, T.B. Detection of Marine Animals in a New Underwater Dataset with Varying Visibility. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Long Beach, CA, USA, 16–17 June 2019; pp. 18–26.
38. Gao, X.; Xin, H.; Liu, J.; Li, T. Event-Driven Prescribed Optimal Disturbance Rejection for Dynamic Positioning of Ships via Reinforcement Learning. *IEEE Trans. Autom. Sci. Eng.* **2026**, *23*, 4615–4626. [[CrossRef](#)]
39. Cao, L.; Qin, Y.; Pan, Y.; Liang, H. Prescribed Performance-Based Optimal Formation Control for USVs With Position Constraints and Yaw Angle Time-Varying Partial Constraints. *IEEE Trans. Intell. Transp. Syst.* **2025**, *26*, 4109–4121. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.