drones

MDPI

*Article*

# AI-Driven Tethered Drone Surveillance for Maritime Security in Ports and Coastal Areas

**Alberto Belmonte-Hernández [1], Briac Grauby [2], Anaida Fernández García [1], Solange Tardi [2], Torbjørn Houge [3], Hidalgo García Bango [4] and Álvaro Gutiérrez [5,*]**

[1] Grupo de Aplicación de Telecomunicaciones Visuales (GATV), Universidad Politécnica de Madrid, 28040 Madrid, Spain; alberto.belmonte@upm.es (A.B.-H.); anaida.fernandez@upm.es (A.F.G.)
[2] Elistair, 3 Chemin du Jubin, 69570 Dardilly, France; b.grauby@elistair.com (B.G.); s.tardi@elistair.com (S.T.)
[3] Maritime Robotics, Brattørkaia 11, 7010 Trondheim, Norway; torbjorn.houge@maritimerobotics.com
[4] Base Marítima Ribadeo, Agencia Estatal de Administración Tributaria, 27700 Ribadeo, Spain; hidalgo.garcia@correo.aeat.es
[5] ETSI Telecomunicación, Universidad Politécnica de Madrid, 28040 Madrid, Spain
* Correspondence: a.gutierrez@upm.es

## Highlights

### What are the main findings?

- The paper demonstrates a tethered UAV platform capable of long-endurance (up to 50 h) autonomous aerial surveillance with reliable take-off and landing from moving vessels in realistic sea states.
- The study shows that YOLOv8-based RGB detection with BoT-SORT tracking provides robust vessel detection and multi-object tracking performance in real maritime conditions.

### What are the implications of the main findings?

- The integrated system offers port and coastal authorities a persistent, mobile "elevated sensor" that can close blind spots of fixed infrastructure and reduce dependence on costly manned aerial surveillance for smuggling detection.
- The work establishes a practical baseline for future enhancements in maritime AI, paving the way for next-generation autonomous maritime surveillance concepts.

## Abstract

Effective port and coastal surveillance require persistent monitoring, flexible deployment, and reliable target detection in dynamic maritime environments. This paper presents a system- and deployment-oriented autonomous tethered drone architecture, integrated with AI-based perception, for persistent maritime surveillance in ports and coastal areas. Mounted on a moving maritime platform and powered through a tether, the drone provides a persistent elevated viewpoint without the endurance limitations of conventional battery-powered Unmanned Aerial Vehicles (UAVs). The system combines maritime platform integration, tethered flight operation, fail-safe and safety mechanisms, and a distributed Artificial Intelligence (AI) pipeline for real-time object detection and tracking. The perception module is based on YOLOv8m for vessel detection and BoT-SORT for multi-object tracking, enabling continuous monitoring of maritime targets in realistic operational scenarios. Field trials conducted from moving vessels in maritime environments demonstrate autonomous take-off and landing, stable surveillance operation under realistic wind and wave conditions, and effective vessel detection and tracking on real image sequences. The results show the potential of AI-enabled tethered drone surveillance as a persistent and operationally relevant tool for maritime monitoring and security.

Check for updates

Academic Editor: Pablo Rodríguez-Gonzálvez

Received: 7 February 2026
Revised: 27 March 2026
Accepted: 31 March 2026
Published: 8 April 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) license.

**Keywords:** tethered drone; unmanned aerial vehicle; maritime surveillance; smuggling detection; port security; object detection; multi-object tracking; autonomous systems

## 1. Introduction

Maritime smuggling, human trafficking, illegal fishing and other clandestine activities in ports and coastal waters pose persistent risks to national security, economic stability and environmental protection [1]. Modern ports are highly complex, dynamic infrastructures where legitimate commercial operations, passenger movements and leisure activities occur alongside potential illicit behaviors that deliberately attempt to exploit blind spots in surveillance coverage. Conventional surveillance infrastructures typically rely on fixed-camera networks, thermal cameras, perimeter intrusion detection systems and software for behavior and object analysis [2]. While these systems provide continuous monitoring of critical facilities, their fixed viewpoints and line-of-sight limitations result in areas of reduced visibility, particularly behind large ships, cranes, warehouses and other port structures [3]. As smuggling tactics evolve and small, agile vessels are increasingly used to transfer contraband close to shore, these gaps in situational awareness become more critical.

Recent literature has highlighted the increasing need for alternative surveillance systems. Satellite earth observation, while valuable for large-scale maritime domain awareness, also has limitations for near-shore surveillance because revisiting time, latency, and spatial resolution often make it unsuited for persistent, fine-grained monitoring of port basins and coastal routes [4]. Reviews of maritime monitoring technologies emphasize that manned maritime systems are becoming key complementary assets for coastal surveillance, inspection, search-and-rescue, and maritime security tasks, as they can extend sensing beyond the limits of static infrastructure while reducing risk to personnel [5–7]. At the same time, surveys on maritime visual ship detection and aerial monitoring indicate that environmental variability, target scale, occlusion and real-time processing constraints remain major obstacles to robust operational deployment [8]. These observations reinforce the need for surveillance systems that combine mobility, persistence, and reliable perception in real maritime conditions. Although traditional airborne surveillance assets can partially complement fixed infrastructure by providing a flexible, elevated vantage point over wide maritime areas [9], manned aerial platforms are expensive, crew-dependent, and constrained in duration by duty cycles, which limits their suitability for continuous patrol. These factors motivate the exploration of unmanned and more cost-effective solutions capable of sustained operation.

In this context, unmanned surface vehicles (USVs) and unmanned aerial vehicles (UAVs) offer complementary strengths. USVs provide extended endurance, payload capacity, and safe operation in hazardous maritime environments [10], while UAVs offer elevated viewpoints and wide-area situational awareness [11]. However, conventional battery-powered UAVs are fundamentally constrained by endurance, often remaining airborne for only tens of minutes before requiring recovery and battery replacement or recharging [12]. This severely limits their usefulness for persistent aerial surveillance and increases mission complexity in maritime operations.

To overcome this limitation, recent research has begun to explore cooperative multi-platform systems in which aerial and surface vehicles operate together [13–15]. In such architectures, the surface vehicle can provide transport, energy, and communications support, while the aerial platform supplies the elevated perspective needed for wide-area observation. These cooperative concepts are particularly relevant for maritime security scenarios, where surveillance assets may need to relocate dynamically in response to vessel

movements or evolving operational demands. However, much of the existing literature focuses either on cooperative mission planning or on perception in isolation, with its associated autonomy limitations.

In this scenario, tethered UAV systems address this endurance limitation by supplying power through a cable from a ground or ship-based station, allowing flights that last hours instead of minutes [16]. In addition, the tether can serve as a robust data link, reducing reliance on wireless channels that may be congested or subject to interference in busy port environments [17]. When such a tethered aerial system is mounted on a mobile maritime platform such as a USV, the result is a persistent, relocatable "elevated sensor mast" that can follow dynamic targets, patrol port approaches and adapt to evolving threat patterns. The combination of USV and tethered UAV thus merges the endurance, stability and payload capacity of the surface platform with the wide area visibility and flexibility of an aerial asset.

This configuration is especially relevant in the maritime domain. However, previous studies on maritime UAV operation have shown that wind gusts, deck motion, turbulence, tether dynamics, and constrained landing geometries can significantly affect flight stability and safe recovery [18,19]. Tethered platforms are attractive because they support longer endurance and more robust power and data links, but they also require careful design of control, landing, and safety mechanisms. As a result, persistent tethered aerial surveillance from maritime carriers should be understood as a distinct systems problem, rather than simply an extension of conventional battery-powered UAV observation.

In parallel with these hardware developments, recent progress in computer vision and deep learning has greatly improved the ability to automatically detect, classify, and track objects in complex visual scenes [20]. Modern convolutional and transformer-based architectures achieve real-time performance with high accuracy and are increasingly used in maritime domain awareness applications [21]. Dedicated datasets captured from UAVs over water, including those developed for search-and-rescue scenarios, now enable training of detectors capable of recognizing small vessels, swimmers, and other maritime targets across a range of distances and viewing angles. Nevertheless, robust perception in real port and coastal environments remains difficult due to cluttered backgrounds, reflections, varying sea states, occlusions caused by infrastructure and large ships, horizon ambiguity, and rapidly changing illumination [22]. These effects are particularly problematic for small and distant vessels, which are often among the most relevant targets in surveillance applications.

Although prior maritime UAV studies have demonstrated the value of aerial imagery for vessel detection and coastal monitoring, most focus primarily on perception performance or short-duration flights using battery-powered platforms [23]. Fewer works address the system-level challenges of persistent deployment from moving maritime platforms, such as power continuity, tether management, fault recovery, safe autonomous landing, and communications robustness under real sea conditions [24]. Likewise, real-world validations are often limited in operational scope. This paper addresses these deployment-oriented gaps by combining persistent tethered aerial sensing, maritime platform integration, reliability mechanisms, and field-tested AI-based perception in a single operational framework.

This paper is framed under the SMAUG (Smart Maritime And Underwater Guardian) project (https://smaug-horizon.eu/, accessed on 12 February 2026) whose focus is on persistent maritime surveillance in ports and coastal zones with the specific goal of improving smuggling detection and general maritime domain awareness. The contribution of the paper is fourfold. First, we present the integration of a tethered dronebox with maritime carrier platforms, including mechanical installation, power distribution, communications, and operation from moving vessels. Second, we describe the reliability and safety architecture required for practical deployment in maritime environments, including layered health monitoring, automatic landing modes, tether and power supervision, com-

munication fault handling, and geo-barrier enforcement. Third, we develop and evaluate an AI-based detection and tracking pipeline for maritime targets using YOLOv8m and BoT-SORT within a distributed processing architecture. Fourth, we validate the complete system through field experiments in realistic maritime conditions, including autonomous take-off and landing from moving vessels, operation under wind and wave disturbances, and assessment of detection/tracking performance on real mission data. Crucially, this operational validation includes demonstrating the zero-shot cross-modality capability of the RGB-trained AI pipeline to successfully detect and track maritime targets using the drone's thermal (IR) sensor in open-water scenarios.

It is important to note that the present work demonstrates the operational use of RGB and thermal sensing in the surveillance platform, but it does not yet propose a full RGB-thermal fusion framework. Rather, the focus is on end-to-end system integration, reliable operation, and real-time maritime target perception under realistic deployment constraints. Therefore, the scientific novelty of this work lies not in the invention of an entirely new UAV platform or vision algorithm in isolation, but in the system-level integration, maritime-specific operationalization, and field validation of a persistent tethered UAV surveillance architecture, with AI-based detection and tracking serving as enabling components of the deployed system. In particular, the work combines persistent tethered aerial sensing, operation from moving maritime platforms, maritime-specific reliability and fail-safe mechanisms, and real-time AI-based vessel detection and tracking within a single validated system. This distinguishes the present study from prior maritime UAV surveillance works that typically focus either on short-endurance UAV imaging, fixed-platform observation, or perception benchmarking without full deployment-oriented system integration. Therefore, the key distinction of this work is that it treats persistent tethered UAV surveillance from moving maritime platforms as a new operational system problem, rather than as a straightforward combination of known aerial, maritime, and AI components. Nonetheless, Table 1 provides a simplified comparison between the proposed SMAUG system and the closest categories of existing maritime UAV presented. The table highlights the main aspects that distinguish the present contribution: persistent aerial operation, deployment from moving maritime platforms and integration of AI-based surveillance functions. This concise comparison helps clarify that the originality of SMAUG lies in bringing these elements together within a single operational framework.

**Table 1.** Comparison of SMAUG proposal with representative maritime UAV surveillance approaches.

| Approach | Main Focus | Persistent Operation | Moving Platform | AI-Based Surveillance |
|---|---|---|---|---|
| SeaDroneSee benchmark | Maritime UAV dataset and benchmark | No | Yes | Only benchmarking |
| Cooperative USV-UAV monitoring | Coastal monitoring with USV + UAV | Partial | Yes | Yes |
| USV-UAV extended coverage | Coastal monitoring with USV + UAV | Partial | Yes | Limited |
| Proposed SMAUG system | Maritime flight, landing, and tether operation | Yes | Yes | No |

To contextualize these contributions, this work explicitly evaluates the proposed system across four fundamental dimensions of modern maritime monitoring: maximizing surveillance coverage through dynamic, elevated viewpoints; overcoming traditional UAV limitations to achieve extended operational endurance (up to 50 h) via tethered power; reducing the operational costs and risks associated with manned aerial assets; and ensuring robust detection performance through quantified AI pipelines deployed in real-world

conditions. However, we note that, although the field deployments provide realistic operational validation, they do not include exhaustive ground-truth annotations; therefore, the field results should be interpreted as deployment-oriented performance evidence rather than as a full ground-truth-based benchmark evaluation. Therefore, this paper should be understood primarily as a system-integration and deployment study for persistent maritime surveillance, in which AI-based detection and tracking are treated as enabling components of the operational architecture rather than as the central methodological contribution.

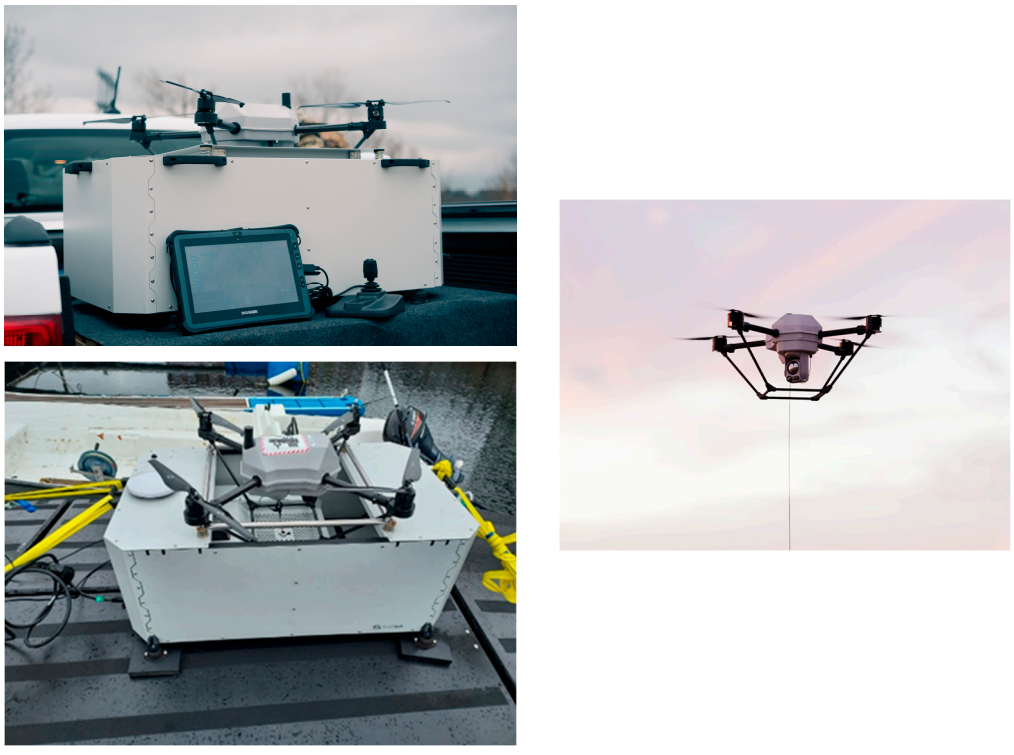
The remainder of this paper is structured as follows. Section 2 presents the materials and methods, distinguishing between the hardware platform (Section 2.1) and the AI algorithms (Section 2.2). Section 3 reports the results of field tests and algorithm evaluation. Section 4 discusses the findings with respect to system integration, algorithm performance and operational readiness. Section 5 concludes the paper and outlines recommendations for future work.

## 2. Materials and Methods

### *2.1. Hardware Platform: Tethered Dronebox System*

#### 2.1.1. System Architecture Overview

The UAV platform is composed of a quadrocopter drone (Khronos, Elistair, Dardily, France (https://elistair.com/solutions/tethered-dronebox-khronos/ accessed on 12 February 2026)) powered by its landing station (see Figure 1). The landing station is mainly composed of a power converter, a winch and a frame as a landing platform. The drone uses custom mode to always follow its station. Thus, if the station is in a vehicle in movement, the drone will follow it. Embedded camera is a Nextvision Raptor (Nextvision, Ra'anana, Israel), with a dual sensor EO/IR, able to zoom x80 (EO), x8 (IR). This concept transforms the UAV into a persistent aerial sensor node whose endurance is no longer dependent on onboard batteries but by the host vessel's power supply and mission constraints.



(**a**) (**b**)

**Figure 1.** (**a**) Dronebox and (**b**) drone.

The dronebox is designed to be mounted on a USV. As the USV moves through the water, the tethered drone maintains a position above the ground station, effectively acting as a mobile elevated observation point. The total system mass, including the drone and ground station, is around 35 kg, which falls well inside the payload capability of USVs. The compactness of the ground station allows installation on the deck without extensive structural modification.

#### 2.1.2. Unmanned Surface Vehicle Platforms

Two types of USVs are used as host platforms for the tethered dronebox (see Figure 2). The Mariner (Maritime Robotics, Trondheim, Norway, https://www.maritimerobotics.com/mariner, accessed on 12 February 2026) is a relatively large diesel-powered vessel with a length of approximately 6 m, a beam of 2 m and a height of 2.7 m. Its dry weight is about 2000 kg, and it offers a payload capacity of 500 kg, sufficient to carry the dronebox along with additional sensors if required. Its top speed reaches approximately 25 knots, and it can maintain survey speeds for around 50 h or loiter for up to four weeks. The vessel is designed to work in sea operations and transit, which makes it suitable for open water surveillance. The OtterX (Maritime Robotics, Trondheim, Norway, https://www.maritimerobotics.com/otter-x, accessed on 12 February 2026) is a smaller, electrically powered USV optimized for sheltered waters such as port basins and coastal inlets. It has dimensions of roughly 4.6 m $\times$ 2.2 m $\times$ 2.3 m, a dry weight of about 800 kg and a payload capacity around 200 kg. The top speed is about 8 knots, and survey endurance is approximately 12 h. Both USVs support a range of communication options, including direct radio links, mesh networks, 5G connections and satellite communications, allowing flexible integration with shore-based control centers and remote monitoring stations.



(**a**) (**b**)

**Figure 2.** (**a**) Mariner and (**b**) OtterX.

From an operational perspective, the Mariner is used for trials and missions that require longer range and endurance in more exposed sea conditions, while the OtterX is reserved for missions inside or near ports where sea conditions are milder and electric propulsion offers advantages in terms of noise and emissions. In both cases, the dronebox is mounted in such a way that the tether has a clear vertical path and is protected from interference with other structures on deck.

#### 2.1.3. Tethered Dronebox Technical Specifications

The airborne component of the dronebox is a multirotor UAV that weighs 3.2 kg. The maximum take-off weight is 3.85 kg, allowing for a payload of up to 650 g. The drone uses four folding 16-inch propellers, which also accept 17-inch variants, so that the arms can be folded for storage when the system is not in operation. This folding mechanism, combined with the compact station footprint, simplifies handling and deployment on the USV deck.

Operationally, the tethered UAV can remain airborne for up to 50 h when powered through the tether. Its maximum operational altitude is 60 m above the ground station, which is sufficient to provide a large field of view over typical port and nearshore envi-

ronments. The ascent and descent speed is limited to about 2 m/s to reduce mechanical stress on the tether and minimize dynamic load on the winch when the USV is moving. An automated follow-me mode allows the drone to maintain a stable position above the ground station while the USV moves at speeds of up to 30 km/h, provided that the wind speed at 60 m altitude does not exceed roughly 50 km/h.

In terms of environmental protection, the drone and station are rated to IP54, which provides adequate protection against spray and dust encountered in maritime environments. The operating temperature range is specified from −10 °C to +45 °C. The system can be deployed in under two minutes from stowed configuration to airborne operation, which makes it suitable for quick reaction scenarios.

#### 2.1.4. Power and Communications Integration

The dronebox must be powered by an external power supply. In each use case, the Khronos dronebox (Elistair, Dardily, France) is powered by the USV or the vessel. Thus, as long as the USV/vessel can power the station, the drone can fly. In case of a power cut, the drone will land automatically.

However, mounting a tethered aerial platform on a USV introduces specific challenges regarding power distribution and communication. The USV must supply enough power to meet both its own propulsion and navigation requirements and the demands of the dronebox. Propulsion is typically the dominant consumer, and its consumption varies with speed, maneuvers and environmental forces such as wind and current. The dronebox draws the highest power during take-off and initial ascent, as well as when compensating for gusts and vessel motion.

The integration design has been implemented to guarantee that the combined system can operate without overloading the USV's power system. Safety margins were defined to ensure that the drone always has access to sufficient power to maintain flight even if the vessel needs to execute power-intensive maneuvers. In cases where power availability could become critical, the power management strategy prioritizes the dronebox over non-essential loads to avoid an in-flight shutdown.

Communication requirements are heterogeneous. Low-bandwidth but low-latency channels are needed for control commands to the USV and the drone, while high-bandwidth channels are required to transfer video and sensor data from the drone to the ground station and command-and-control (C2). The tether itself can carry both power and data, providing a robust and secure link between the drone and the ground station. From the ground station, data can be relayed through the USV's communication systems, which may include radio links, 5G, mesh networks or satellite channels, depending on the deployment scenario and infrastructure availability.

#### 2.1.5. Operational Safety and Failsafe Systems

Because the dronebox operates over water and from moving platforms, safety considerations are central to the system design. The platform implements multiple layers of monitoring and failsafe logic that continuously evaluate the health of the system and trigger appropriate responses when abnormal conditions occur.

One key mechanism is the set of automatic landing modes. In many failure scenarios, the safest and most appropriate response is to command the drone to land on its ground station. When the system detects a fault that compromises safe continuation of flights, such as serious battery issues, critical communication failures or temperature excursions, it initiates an automatic landing procedure. If conditions improve while the drone is descending, the landing can be aborted autonomously, and the mission may continue from the stabilized altitude.

The system continuously monitors a range of fault categories. Examples include insufficient battery charge, battery overheating or poor battery health; loss of power supply to the ground station; power converter failures in either the drone or the station; communication interruptions between the drone, ground station and operator interface; sensor faults affecting ground station position estimation; motor saturation and propeller damage; tether blockage or excessive tether force; component temperatures outside acceptable ranges; insufficient tether traction and geo-barrier violations where the drone moves outside predefined safe zones. For each category, the system defines specific user actions, such as stopping the host vessel, clearing the area around the ground station or performing diagnostic procedures before resuming operations.

### *2.2. Artificial Intelligence: Detection and Tracking Algorithms*

#### 2.2.1. System Architecture Overview

The AI component of the system is responsible for detecting and tracking maritime objects in video streams acquired by the drone's sensors. The architecture is designed as a modular and distributed pipeline that processes dual-modal data from an RGB camera mounted on the drone's gimbal. (see Figure 3).

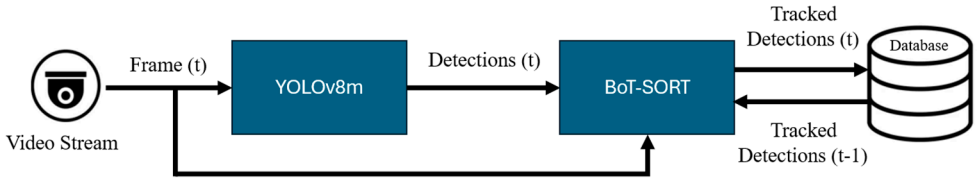



**Figure 3.** AI component architecture for object detection and tracking.

A processing layer continuously monitors this storage for new image frames or video segments. When new data arrives, the corresponding frames are retrieved, preprocessed and fed into the detection model. The detection outputs consist of bounding boxes, confidence scores and class labels for each frame. These detections are then passed to a multi-object tracking algorithm, in this case BoT-SORT [25], which combines motion modeling with spatial association to maintain consistent identities over time. The resulting tracks are used to generate high-level information about vessel trajectories and behaviors. Finally, the processed data, including annotated frames and structured detection/tracking results, is stored again or sent back to the operational interface for visualization and further analysis.

In typical operation, the cameras capture continuous video streams while the drone loiters over an area of interest or follows the USV along a patrol route. Rather than performing all computation onboard, the system uses a ground-based or remote processing infrastructure equipped with GPU resources. The sensor data is transmitted from the drone via the tether to the ground station and then through the USV's communication links to a dedicated server environment. There, the incoming data is stored in an object storage system such as MINIO, which organizes files, timestamps and associated metadata [26] (see Figure 4). This architecture offers several advantages: it decouples data acquisition from processing, allowing the system to scale horizontally by adding more processing nodes if necessary. It also facilitates integration with additional services, such as automated alerting, long-term storage or advanced analytics modules.

#### 2.2.2. Object Detection Dataset

Training effective detection models requires representative and well-annotated datasets. The object detection subset of SeaDronesSee (https://github.com/Ben93kie/SeaDronesSee, accessed on 12 February 2026) is used [27]. It contains more than 14,000 images, divided into training, validation and test splits, and includes nearly 400,000 annotated

bounding boxes. These high-resolution RGB images are captured from UAVs over water, with annotations for objects such as boats, swimmers, jet skis, life-saving appliances and buoys. The images cover a range of altitudes, from very low to several hundred meters, and include variations in camera gimbal angle, sea state and lighting conditions. Rich metadata is associated with each image, including GPS coordinates, altitude and timestamps.

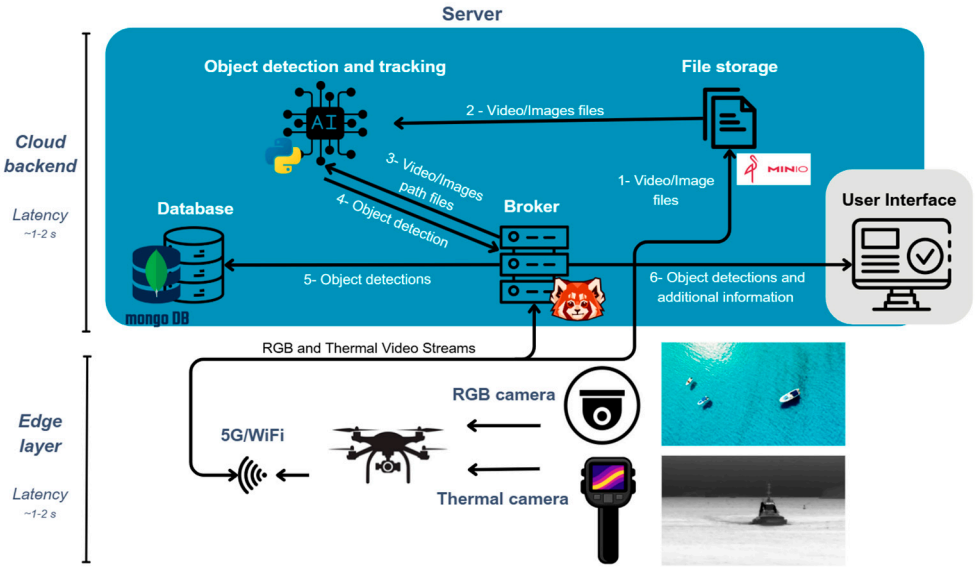



**Figure 4.** System detection architecture, with a MongoDB database, Red Panda Broker and Mino File Storage System.

In this paper, the primary target class for the SMAUG application is the "boat" class, but other classes, such as swimmers and life-saving appliances, are also of potential interest. The presence of varied maritime backgrounds and vessel types makes this dataset particularly suitable for training robust RGB detection models.

#### 2.2.3. Object Detection Model: YOLOv8m Architecture

The detection system is based on YOLOv8m [28] architecture. YOLOv8 is a modern one-stage object detector that improves on previous YOLO versions in terms of accuracy and efficiency. The "m" variant represents a balanced configuration with moderate computational requirements, making it suitable for real-time inference on contemporary GPUs while still delivering good detection performance. The architecture comprises a backbone network that extracts hierarchical visual features, a neck that aggregates features across multiple scales and a head that produces final bounding box and class predictions. Cross-Stage Partial (CSP) connections and C2f modules are used in the backbone to enhance gradient flow and reduce redundancy. In the neck, a combination of feature pyramid and path aggregation networks enables detection of objects at different scales. The head operates without predefined anchor boxes, instead predicting bounding boxes directly, which simplifies configuration and can improve generalization.

Based on an extensive review of current computer vision literature, the selection of the YOLOv8m architecture was driven by the strict operational requirements of tethered aerial surveillance: real-time inference speed, high accuracy on small objects, and deployment efficiency. It is well-documented that while two-stage detectors like Faster R-CNN [29] generally offer high localization precision, their computational overhead makes it difficult to sustain the >30 FPS required for smooth multi-object tracking. Conversely, compared to earlier one-stage iterations (such as YOLOv5 [30]), YOLOv8 introduces an anchor-free detection head. Literature shows this is particularly advantageous in maritime environments,

as it eliminates the need to pre-define anchor box sizes, allowing the model to generalize across extreme scale variations (e.g., a tiny swimmer versus a cargo ship)

Furthermore, we acknowledge that more recent state-of-the-art architectures, such as Real-Time Transformers (e.g., RT-DETR [31]) or subsequent YOLO iterations [32], offer highly competitive or even superior speed-to-accuracy trade-offs. However, the selection of YOLOv8m was ultimately dictated by its maturity, highly optimized deployment ecosystem (e.g., seamless TensorRT export for edge GPUs), and native, well-tested integration with multi-object tracking frameworks like BoT-SORT. For a system integration architecture focused on operational robustness and failsafe deployment rather than algorithmic benchmarking, YOLOv8m provided the most stable, supported, and reliable foundation to guarantee sustained real-time performance in the field.

To ensure rigorous benchmarking, fair comparability with other state-of-the-art methods, and to strictly prevent data leakage, we utilized the official predefined dataset splits provided by the SeaDronesSee creators. The model was trained exclusively on the official training set, while hyperparameter tuning and early stopping were monitored using the official validation set. Final detection metrics were computed on the designated test set, ensuring a standardized and transparent evaluation of the maritime target detection capabilities.

To guarantee transparency and repeatability, the training process was strictly defined. Training was carried out on an Intel Core i9-10900 CPU @ 2.80 GHz with 32 GB RAM and an NVIDIA RTX 4090 (24 GB VRAM) GPU. Input images were resized to a standard square resolution of 640 × 640 pixels. This dimension was selected to provide an optimal balance between preserving the visual features of critically small objects (such as swimmers or distant vessels) and maintaining efficient GPU memory utilization.

The model was trained for a maximum of 150 epochs. To mitigate the risk of overfitting—a common issue when dealing with inherently imbalanced class distributions in specialized maritime datasets—an early stopping mechanism was implemented with a patience of 50 epochs, monitoring the validation mAP at Intersection over Union (IoU) 0.5. Optimization was performed using Stochastic Gradient Descent (SGD) with momentum. A linear learning rate decay schedule was applied to ensure rapid initial convergence followed by stable fine-tuning in the later epochs. Standard YOLOv8 data augmentations (such as mosaic, horizontal flips, and HSV color jitter) were kept active to enhance the model's robustness against varying lighting and sea conditions. The complete set of hyperparameters is summarized in Table 2.

**Table 2.** Object Detection YOLOv8m Model Training Hyperparameters.

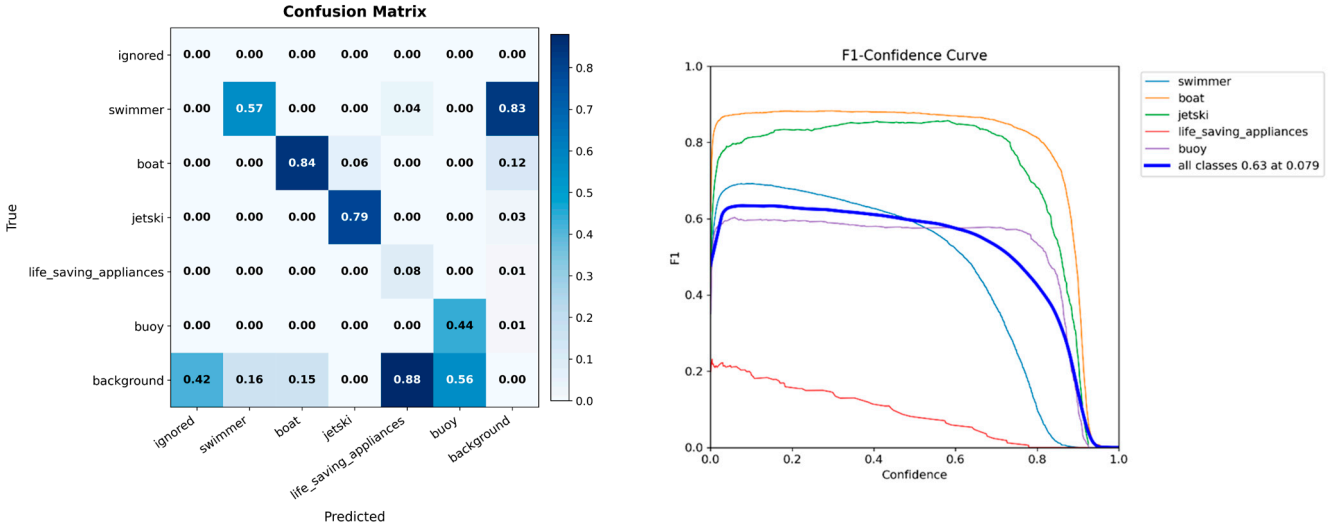
| Parameter | Value | Description |
|---|---|---|
| Input Resolution | 640 × 640 | Standardized image size for training |
| Max Epochs | 150 | Maximum number of training cycles |
| Early Stopping Patience | 50 | Epochs without validation mAP improvement before stopping |
| Batch Size | 32 | Number of images processed before model update |
| Optimizer | SGD | Stochastic Gradient Descent |
| Momentum | 0.937 | Accelerates SGD in the relevant direction |
| Weight Decay | 0.0005 | L2 regularization term to prevent overfitting |
| Initial Learning Rate (lr0) | 0.01 | Starting learning rate |
| Final Learning Rate (lrf) | 0.0001 | End learning rate (linear decay) |

For operational deployment and real-time inference during the field tests, the confidence threshold was empirically set to 0.4. This specific threshold was calibrated to aggressively filter out false positives frequently caused by dynamic sea clutter (e.g., breaking waves, sun glare, and whitecaps) while maintaining a high recall for actual targets.

Additionally, the Non-Maximum Suppression (NMS) IoU threshold was set to 0.5. This ensures that duplicate bounding boxes on the same target are suppressed, without accidentally merging valid detections of multiple vessels navigating in close proximity to one another.

#### 2.2.4. Detection Performance Evaluation

The RGB model trained on SeaDronesSee achieves good overall performance on the validation and test sets (see Figure 5a). For boats, which are the primary object of interest, recall is in the order of 84%, and precision at practical confidence thresholds is high. The mean average precision at Intersection over Union (IoU) 0.5 exceeds 0.64, indicating that the model can reliably localize and classify boats under a variety of conditions. Performance for other classes varies: jet skis are detected reasonably well, swimmers are more challenging due to their small size and low contrast against the water, and buoys and life-saving appliances show lower recall, partly because they are less frequent in the dataset and share visual characteristics with other objects or background structures. Complementary to the confusion matrices, the F1-confidence curve (see Figure 5b) indicates that the model maintains high F1 scores across a broad range of confidence thresholds for the most common classes.

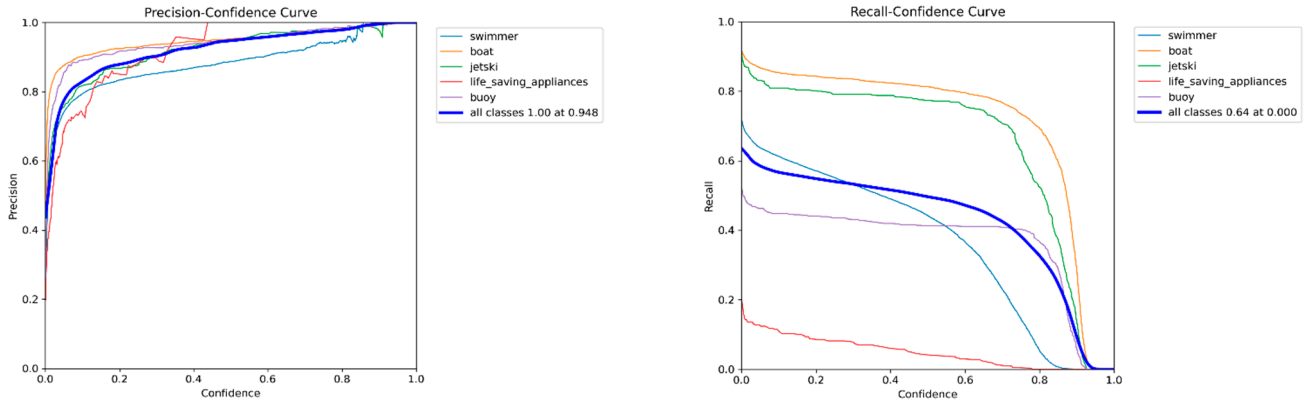



(**a**) (**b**)

**Figure 5.** (**a**) Confusion matrix for SeaDronesSea dataset evaluation with trained model. (**b**) F1-Confidence Curve for SeaDronesSea dataset evaluation with trained model.

Error analysis reveals that many misclassifications occur at large distances (>500 m), where vessels occupy only a small number of pixels and may blend into the horizon line. Some confusion between visually similar classes also occurs, such as between boats and jet skis, or between swimmers and background patterns like whitecaps. These observations point to potential benefits from targeted dataset augmentation for rare classes and specific scenarios such as horizon-edge detections.

The Precision-Confidence and Recall-Confidence curves (see Figure 6) confirm this interpretation. Precision remains high across classes as confidence increases, particularly for boat and buoy, which surpasses 0.9 in precision above a threshold of 0.5. However, recall for certain categories (especially life_saving_appliances and buoy) drops sharply at higher thresholds, suggesting a trade-off that must be carefully calibrated depending on the application requirements—whether minimizing false positives or maximizing detection coverage.




(**a**) (**b**)

**Figure 6.** (**a**) Precision-Confidence and (**b**) Recall-Confidence Curves for SeaDronesSea dataset evaluation with trained model.

Finally, the training evolution plots (see Figure 7) show a smooth and consistent convergence across loss components. Both bounding box regression (box_loss) and classification loss (cls_loss) decrease steadily, indicating stable training. The validation losses follow a similar trend, and the model's mAP metrics improve progressively, particularly mAP@0.5, which rises from around 0.45 at the start to over 0.64 by the end of training. This suggests the model continues to benefit from the dataset's diversity and has not been overfitted despite class imbalance.

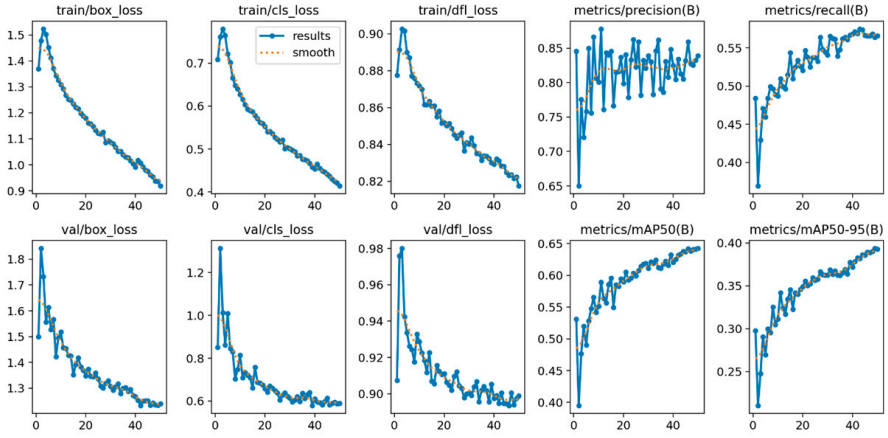
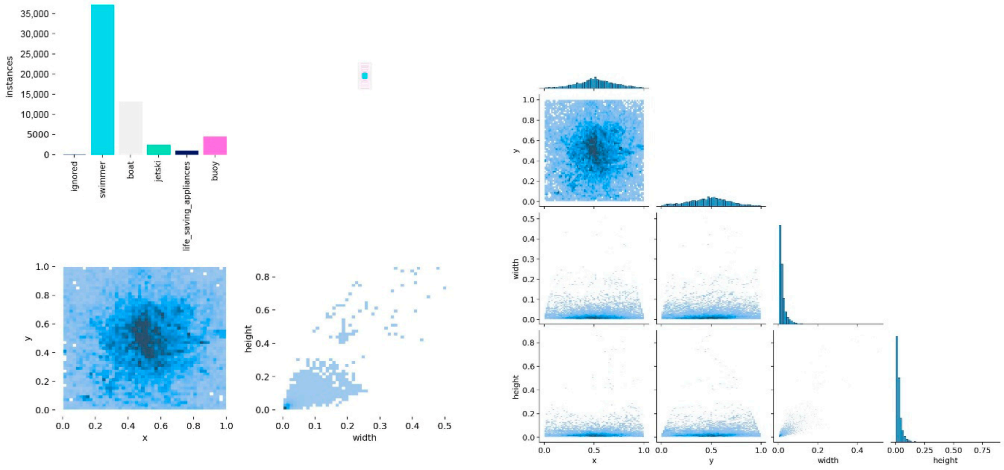



**Figure 7.** Training loss and mAP Curves per epoch for SeaDronesSea.

Visualizations of label distribution (see Figure 8) confirm the dataset is heavily skewed towards swimmer instances, which dominate the training set. Most annotations are concentrated near the image center, reflecting typical UAV framing behavior, and bounding box sizes cluster in the lower range, with most objects being small—another reason is that detection remains challenging for minor or partially occluded classes.

Overall, these results confirm that the model can perform accurate, real-time detection of key maritime objects, particularly boats, swimmers, and jet skis, under diverse conditions. Nevertheless, performance is partially limited by class imbalance and visual similarity between some categories. These aspects could be improved in future work through data augmentation, rebalancing strategies, or incorporating temporal information through tracking. Figure 9 shows some detection examples with the trained model.




(**a**) (**b**)

**Figure 8.** Labels distribution and correlogram for SeaDronesSea dataset. (**a**) Labels distribution and (**b**) correlogram.

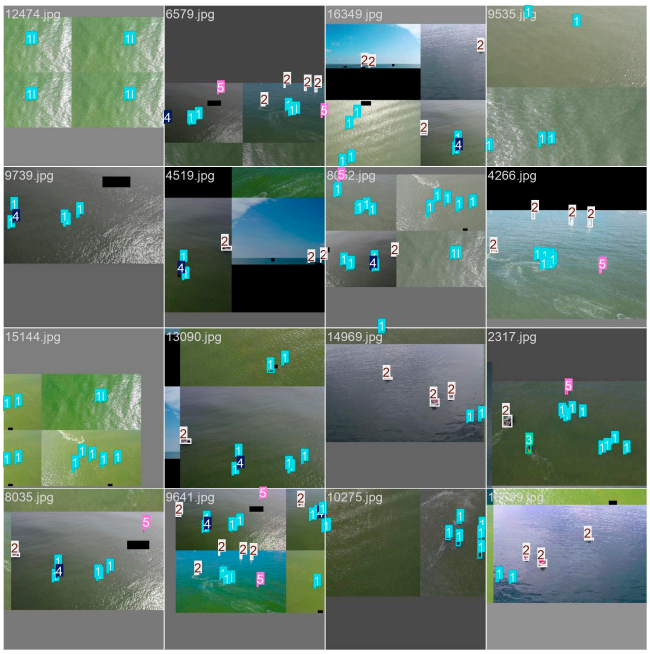



**Figure 9.** Detection examples with YOLOv8m trained on SeaDronesSea dataset [25]. Numbers and colors differentiate classes.

### 2.2.5. Multi-Object Tracking: BoT-SORT Algorithm

Detection alone is insufficient for many surveillance tasks because it does not maintain object identities over time. To address this, the system employs the BoT-SORT (Bag of Tricks SORT) [25], a tracking-by-detection algorithm that extends the SORT family of trackers. BoT-SORT combines a motion model based on a Kalman filter with data association techniques that compare predicted trajectories with current detections. For each object track, the Kalman filter maintains an estimate of the position, scale and velocity of the object's bounding box. At each new frame, the filter predicts the object's state based on the previous state and the assumed motion model. These predictions are then matched to current detections using spatial overlap (intersection-over-union) and, when available, appearance features. The matching procedure solves an assignment problem that pairs detections with existing tracks or initiates new tracks when no suitable matches exist.

BoT-SORT includes several refinements compared to earlier tracking algorithms. It supports more robust motion models that can better handle the dynamics of aerial imagery, where apparent object size and position may change rapidly due to camera motion. It also offers optional integration of re-identification (ReID) features (see Figure 10), which help maintain track continuity through occlusions or temporary loss of detections.

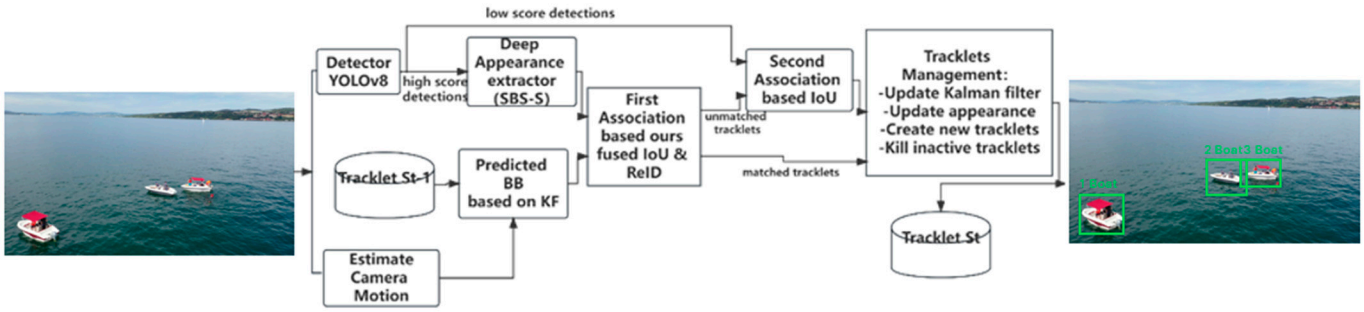



**Figure 10.** Overview of BoT-SORT-ReID tracker pipeline. Adapted with permission from Ref. [25]. Copyright 2022, Springer Nature.

Tracking performance is configured to prioritize stability and real-time operation. Thresholds for track initiation and termination and for association decisions are tuned to the expected motion patterns of vessels observed from a tethered drone. In practice, the combined YOLOv8m and BoT-SORT pipeline runs at more than 30 frames per second on GPU hardware, which is sufficient for real-time monitoring in typical deployment scenarios.

To validate the tracking capabilities of the proposed system, we conducted an evaluation on the Multi-Object Tracking (MOT) split of the SeaDronesSee dataset [27], considering a global performance metric across all classes. We conducted a systematic grid search over the SeaDronesSee validation set. The search space focused on critical BoT-SORT hyperparameters that dictate track initialization, termination, and spatial association, which are highly sensitive to maritime challenges such as sun glare, sea clutter, and abrupt camera motion. Specifically, we evaluated combinations of detection confidence thresholds ("track_high_thresh" in the range [0.4, 0.7], "track_low_thresh" in [0.05, 0.2]), track initiation thresholds ("new_track_thresh" in [0.5, 0.8]), and track memory limits ("track_buffer" in [30, 90] frames).

Following the standard MOT benchmark protocol [33], we assessed key metrics such as Multiple Object Tracking Accuracy (MOTA) and Identification F1 Score (IDF1). While state-of-the-art baselines like Tracktor++ [34] and FairMOT [35] establish high accuracy benchmarks (MOTA 71.9% and IDF1 80.5%, MOTA 36.5% and IDF1 43.8% respectively), often utilizing heavier backbones such as ResNet50 or DLA34, our YOLOv8m + BoT-SORT pipeline achieves competitive intermediate results in comparison with SeaDronesSee benchmark achieving MOTA 57.4% and IDF1 62.8%. Although our approach exhibits slightly lower MOTA and IDF1 scores compared to the top-performing Tracktor++ configuration reported in the original dataset paper, it significantly outperforms heavier models in terms of inference speed, 38 FPS vs. 22 (FairMOT) and 6 (Tracktor++) in our GPU hardware. This trade-off is intentional: by prioritizing computational efficiency, our system maintains robust tracking stability while ensuring real-time operation at over 30 FPS, a critical requirement for live aerial surveillance that many higher precision but computationally intensive trackers fail to meet.

After evaluating all parameter combinations in the grid search, the empirical results yielded a specific optimal configuration that maximizes both tracking accuracy and identity preservation. Interestingly, the best-performing values derived from the validation set align perfectly with the physical and environmental challenges of our maritime use case.

Specifically, the search revealed that the optimal primary detection confidence thresholds, "track_high_thresh" and "track_low_thresh", converged to 0.6 and 0.1, respectively. This wide margin in the two-tier filtering approach (BYTE logic) proved crucial in our tests: it successfully discards high-confidence false positives caused by sun glare or sea foam, while still associating low-confidence detections of vessels partially obscured by waves. Furthermore, the empirical results showed that the threshold to initiate a new track ("new_track_thresh") needed to be strictly set to 0.7. This relatively high value effectively prevents transient background anomalies, such as breaking whitecaps, from generating phantom trajectories.

To handle temporary occlusions caused by significant wave heights or sudden UAV gimbal movements, the "track_buffer" was found to be 60 frames. At an operating speed of 30 FPS, this provides a robust 2 s memory window to retain lost tracks before deletion. Finally, the spatial association limit ("match_thresh") converged to 0.8 (Intersection-over-Union). Alongside these thresholds, Camera Motion Compensation (CMC) was enabled to correct for the ego-motion and wind-induced sway of the tethered UAV, successfully aligning the Kalman filter predictions with the shifting background.

Regarding the optional integration of re-identification (ReID) features, we confirm that this module was enabled in our final system architecture. Rather than training a domain-specific extractor from scratch, we utilized the default lightweight ReID network provided by the BoT-SORT implementation (OSNet), functioning in a zero-shot transfer capacity. While typically pre-trained on urban or pedestrian datasets, the robust generic feature extraction of this model, capturing fundamental attributes like color distribution, structural edges, and textures, generalizes highly effectively to maritime vessels.

To explicitly quantify the specific contribution of the ReID module to identity preservation, we conducted an ablation study. Relying solely on the Kalman filter motion model, it yielded an IDF1 score of 57.5%. Enabling the default ReID appearance matching increased the IDF1 to 62.8%, representing an absolute gain of 5.3%. Although this improvement margin might appear moderate, it provides a clear quantitative enhancement in identity preservation without significantly degrading the real-time processing speed, as the default ReID network is highly lightweight. This relatively modest gain is intrinsically linked to the topological nature of the SeaDronesSee dataset. Unlike urban environments with dense crowds and frequent trajectory crossovers, where ReID modules traditionally excel at preventing identity switches, maritime traffic in this dataset is predominantly sparse. In our context, tracking interruptions are more frequently caused by missed detections (e.g., a vessel temporarily obscured by a wave crest or sea spray) rather than complex spatial overlaps between multiple vessels. Consequently, while the ReID module is highly effective at recovering identities after such temporal occlusions, the sparse nature of the scene limits the frequency of scenarios where appearance features are critically required, thus explaining the moderate but highly cost-effective 5.3% boost in IDF1.

Furthermore, to visually validate the discriminative power of the off-the-shelf ReID model on our specific maritime classes, we extracted the high-dimensional appearance embeddings from hundreds of diverse individual detections across all target classes in the validation set. As shown in Figure 11, these embeddings were projected into a 2D space using t-SNE (t-Distributed Stochastic Neighbor Embedding). The resulting visualization demonstrates clear clustering and separability not only among different maritime classes but also among distinct identities within those classes. This empirically justifies our use of the default ReID architecture, as it effectively extracts highly discriminative features without the computational overhead of additional domain-specific fine-tuning.

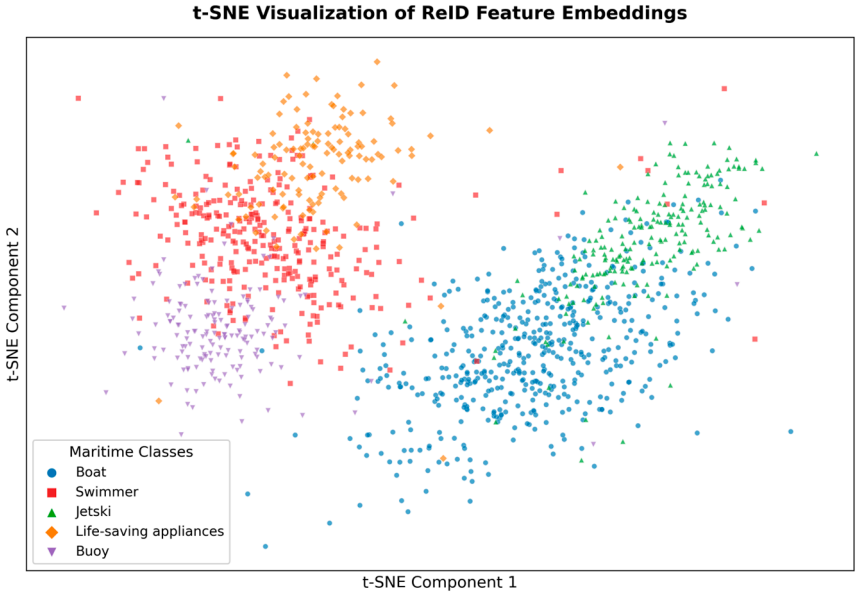



**Figure 11.** t-SNE visualization of ReID embeddings from the SeaDronesSee dataset. The natural clustering of maritime classes justifies using the default extractor without fine-tuning.

#### 2.2.6. System Output Format

The output of the detection and tracking pipeline is structured in a way that facilitates storage, visualization and downstream processing. For each processed frame, the system generates a record containing a frame identifier, timestamp, image dimensions and a list of detections. Each detection entry includes a track identifier, class identifier, class name, confidence score and bounding box coordinates. This information is encapsulated in JSON to be easily parsed by the SMAUG server.

Annotated images or videos with bounding boxes and track IDs are generated for visualization in operator interfaces. These visualizations aid human operators in understanding the situation and verifying algorithm performance. At the same time, the structured data enables automated analysis, such as counting vessel entries into a restricted area, measuring dwell times or identifying anomalous motion patterns that could indicate smuggling or other illicit activities.

## 3. Results

### *3.1. System Integration and Field Testing*

#### 3.1.1. Initial Testing in Trondheim

The first integrated field tests of the tethered dronebox mounted on a Mariner USV took place in Trondheim, Norway, in mid-May 2025. The primary objectives were to verify mechanical integration, test the power and communication subsystems and evaluate basic autonomous behaviors such as take-off, hovering and landing on a moving platform.

Weather conditions during the tests were challenging. On the first day, wind speeds reached approximately 60 km/h, accompanied by rain, which restricted flight operations to short test intervals. These tests were nevertheless valuable for verifying that the dronebox and tethering system remained secure under harsh conditions. Additional straps and protective routing of the tether were installed to reduce the risk of damage from wave-induced vessel motion (see Figure 12).



(**a**) (**b**)

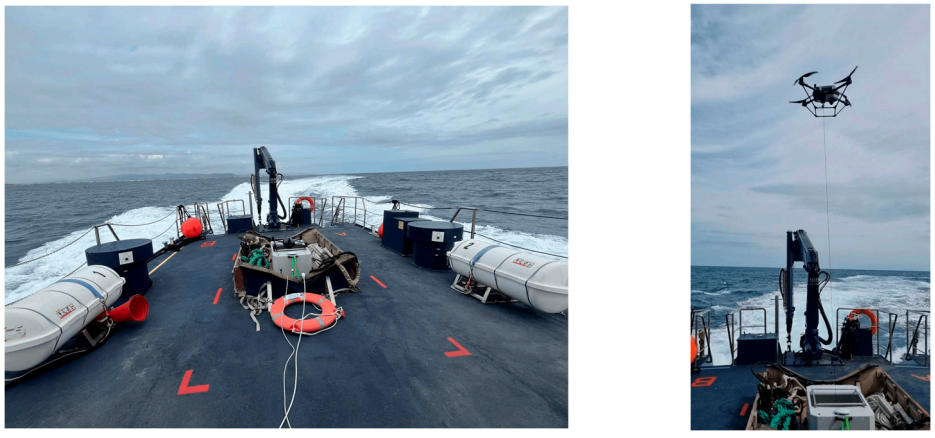
**Figure 12.** (**a**) Integration of the dronebox on Marine USV. (**b**) First flight of the dronebox integrated on the Mariner USV.

On the second day, the weather improved somewhat, with wind speeds closer to 45 km/h. This allowed more extensive testing. The team conducted multiple autonomous take-off and landing cycles, both with the vessel stationery and in motion. The USV performed maneuvers including straight-line runs and U-turns at speeds up to around 7 knots. The tethered drone maintained its position above the ground station and followed the vessel reliably. To further stress the system, the team generated waves of around 60 cm height by maneuvering the vessel. Throughout these maneuvers, the position estimation and control algorithms demonstrated robustness, although some oscillations were observed at maximum altitude during aggressive maneuvers.

These initial trials provided confidence in the mechanical robustness of the integrated system and in the basic performance of the tether management and positioning algorithms. They also highlighted areas where further tuning of control parameters and estimator gains would be beneficial to reduce oscillations and improve stability at higher altitudes and in gusty winds.

#### 3.1.2. Validation Testing in Valencia

A more comprehensive validation campaign was conducted in the port of Valencia, Spain, in September 2025. In this case, the dronebox was installed on a fast patrol boat (see Figure 13), and the focus was on demonstrating autonomous operation in realistic port and near-shore conditions, including video transmission to remote monitoring locations.



(**a**) (**b**)

**Figure 13.** (**a**) Positioning of the dronebox on the patrol boat. (**b**) Drone flight from the patrol boat.

During these tests, the vessel executed various maneuvers at speeds between 7 and 10 knots, including straight runs, turns and laps around predefined areas of interest. The drone performed multiple autonomous take-off and landing cycles while the vessel was moving. Wave heights up to approximately 1 m were encountered. Under these conditions, the drone maintained stable flight and accurate positioning relative to the ground station.

One of the objectives in Valencia was to test long-range video streaming via satellite internet. A Starlink antenna on the vessel provided external connectivity. Local recording and processing of video data on board remained stable. However, the tests revealed that antenna stability and electromagnetic compatibility issues significantly influenced link quality. Vessel motion and structural obstructions sometimes disrupted the link, which in turn affected real-time streaming.

Overall, the Valencia campaign confirmed that the tethered dronebox system can operate autonomously from a moving vessel in realistic maritime conditions and that it can deliver the persistent elevated viewpoint needed for surveillance tasks. It also underlined the importance of robust communication infrastructure for remote streaming, which remains sensitive to installation details and environmental factors.

### *3.2. Detection Algorithm Performance*

During the open-water field tests in Trondheim and Valencia, a total of approximately 3 h of continuous drone video (comprising approximately 2 h of RGB and 1 h of Thermal/IR footage) were collected across various real-world operational scenarios. It is important to emphasize that these recorded sequences were strictly reserved for operational evaluation. No images or data from these field deployments were used to fine-tune or re-train the underlying YOLOv8m or ReID models, thereby providing a true measure of the system's zero-shot generalization capabilities in completely unseen environments.

The YOLOv8m RGB detection model trained on the SeaDronesSee dataset was applied to video recordings from the Trondheim and Valencia field tests. These recordings included sequences where the drone observed vessels at various ranges and under different viewing angles and lighting conditions.

The first analysis was conducted on the videos recorded outside the port area. The proposed detection and tracking algorithms were applied to video sequences to evaluate their robustness in less structured, open-water environments. Representative frames are included to illustrate the performance: examples of successful detections as well as failure cases are shown on Figures 14 and 15. These visual results highlight the strengths and limitations of the system across modalities, providing a first comparative benchmark for real-world performance.

In the video sequences, the tests demonstrate that both detection and tracking generally perform correctly. The boat or small vessel is effectively identified across different positions and perspectives, with the tracking module able to maintain continuity over time. Some occasional losses occur when successive detections fail, leading to short tracking interruptions; however, these are limited and do not compromise the overall performance. Overall, the system provides reliable results at both close and longer distances, confirming its suitability for vessel detection and monitoring in real-world conditions.

In the first set of examples (see Figure 14), the system correctly identifies and follows the target boat during navigation. The detections are consistent across frames, and the tracker is able to maintain continuity even as the vessel changes position and distance relative to the drone. In a second case (Figure 14), the algorithms detect other vessels further away, including ships approaching the horizon line. Although these detections are less stable due to distance and resolution constraints, they demonstrate the model's potential to recognize multiple maritime targets in complex environments, which is par-

ticularly relevant for surveillance tasks. Interestingly, the RGB-trained detection model also shows an unexpected capability to detect vessels in thermal imagery (see Figure 15). While this cross-modality performance was not part of the original training or optimization process, it highlights the adaptability of the algorithms and opens opportunities for further exploration.



**Figure 14.** Detections of tracked boats with RGB drone camera. Correct classifications in green. Wrong classifications in red.

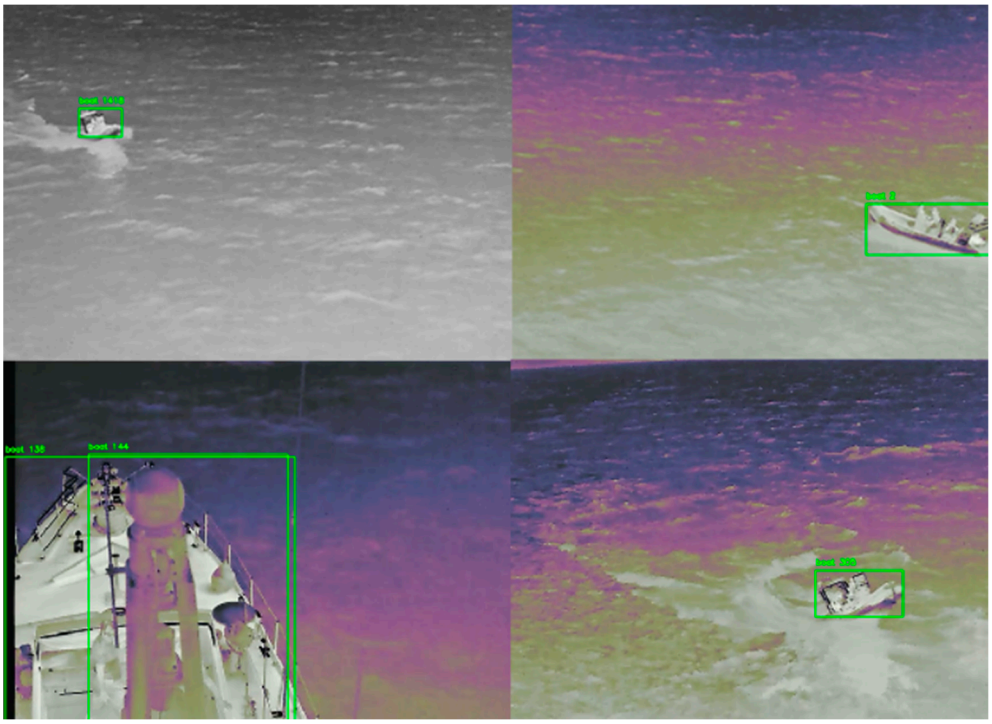


**Figure 15.** Detections of tracked boat with drone Thermal camera.

While the SeaDronesSee benchmark provides theoretical baseline for the AI algorithms, the open-water tests serve as the operational validation of the integrated architecture. To move beyond qualitative visual inspection and quantitatively assess the field deployment, we evaluated the system's operational metrics during the field trials. From a computational standpoint, the distributed architecture successfully handled the live video streams, main-

taining a stable inference and tracking rate of 36–39 FPS on the remote ground control server. There were zero critical failures or processing bottlenecks during 30 min of continuous airborne observation, validating the robustness of the data link through the tether.

Furthermore, to quantitatively estimate perception robustness in the wild without the need for exhaustive manual pixel-level annotation, we measured the Target Tracking Continuity (TTC) over several continuous operation sequences. Although TTC offers a practical measure of operational robustness, we acknowledge it is not a standard tracking metric (e.g., MOTA, IDF1). Computing standard metrics requires exhaustive frame-by-frame ground-truth annotation, which is unfeasible for hours of field footage. Consequently, TTC presents specific limitations: (1) it does not evaluate pixel-level localization precision (IoU); (2) it is a target-centric metric that does not penalize false positives elsewhere in the scene; and (3) it relies on human validation rather than automated benchmarking. Despite these limitations, TTC remains a highly effective indicator of the system's ability to maintain persistent situational awareness under severe maritime conditions.

In a representative 15 min continuous sequence (one RGB and one Thermal/IR) tracking a moving target vessel under real wind and wave disturbances, the YOLOv8m + BoT-SORT pipeline successfully maintained the bounding box and identity of the target for 85.5% of the visible frames. The few tracking interruptions (accounting for the remaining 14.5%) were primarily caused by severe temporary occlusions (e.g., heavy sea spray or extreme gimbal compensation movements) resulting in transient detection losses. In these instances, the ReID module successfully re-associated the original target ID within an average of 1.5 s once the vessel became visible again, preventing permanent identity switches. This operational quantification confirms that the system is highly reliable for persistent maritime surveillance.

Furthermore, an analysis of the detection outputs during these tracking sequences confirms the robustness of the zero-shot model in the wild. Consistent with typical YOLO performance on well-represented classes, the integrated system exhibited high certainty in its field predictions, maintaining a bounding box detection confidence score greater than 0.80 for over 90% of the successfully tracked frames in open water.

Similarly, to quantitatively assess the zero-shot cross-modality performance, we evaluated the TTC on the corresponding thermal (IR) video streams with same considerations as for the RGB images tests. In the thermal domain, the system achieved a slightly higher tracking continuity, maintaining the target identity for 88.3% of the visible frames. This increase in baseline stability is attributed to the thermal sensor's immunity to visible-spectrum artifacts, such as sun glare and shadows on the water surface, which occasionally disrupt RGB detections. However, the remaining 11.7% of tracking interruptions revealed a different operational dynamic: because the default ReID appearance extractor relies heavily on RGB-specific features (e.g., hull color and textures), re-association in the monochromatic thermal domain depends more heavily on the purely kinematic predictions of the Kalman filter. Consequently, while the thermal camera provides fewer initial detection drops, recovering a lost ID after a prolonged occlusion takes slightly longer compared to the RGB pipeline. This comparative quantification validates both modalities and highlights their complementary operational strengths. Table 3 summarizes the main field-trial conditions and outcomes for the Trondheim and Valencia deployments, together with the corresponding operational perception evaluation.

To provide deeper insight into the system's performance limits and explain the causes behind these tracking interruptions (the aforementioned TTC gaps), a qualitative error analysis of the field deployment footage revealed three primary failure modes: (1) Distance and Resolution Effects: Missed detections frequently occurred at extreme ranges (e.g., >500 m), where targets occupied only a few pixels and merged with the horizon line, causing

confidence to drop below the 0.4 operational threshold; (2) Environmental Occlusions: In higher sea states, breaking waves and heavy sea spray completely obscured small targets. While the RGB camera struggled with these visual occlusions, the Thermal (IR) sensor proved slightly more resilient, though not entirely immune to dense, cold sea spray; (3) Platform Dynamics: Operating from a moving USV in rough waters subjected the drone to sudden wind gusts or tether pulls. When these disturbances temporarily exceeded the gimbal's stabilization capabilities, severe motion blur caused the Kalman filter's linear motion predictions to fail, forcing the ReID module to re-establish the track once stabilized.

**Table 3.** Summary of field-validation conditions.

| Test | Conditions | Main Validated Capability | Main Result | Main Limitation |
|---|---|---|---|---|
| UAV flight | Wind up to 60 km/h; waves up to 60 cm; vessel up to 7 knots | Mechanical integration, autonomous take-off/landing, follow-me on moving USV | Successful multi-cycle autonomous operation | Oscillations at maximum altitude during aggressive maneuvers |
| | Port/nearshore environment; waves up to 1 m; vessel 7–10 knots | Stable autonomous operation from moving patrol boat; remote streaming trials | Stable flight and accurate positioning maintained | Streaming affected by antenna stability, EMC, motion, and obstructions |
| Field video evaluation | Approx. 3 h total video (2 h RGB, 1 h Thermal/IR) | Zero-shot operational evaluation on unseen field data | No retraining/fine-tuning on field data | Distant/horizon targets less stable |
| Runtime validation | Continuous airborne observation | Real-time end-to-end processing | 36–39 FPS, 0 critical failures in 30 min | Limited continuous test window |
| RGB tracking | Representative 15 min sequence | Robust tracking in real disturbances | TTC 85.5%, ReID recovery ~1.5 s, confidence > 0.80 in >90% of tracked frames | Interruptions from spray, occlusion, gimbal compensation |
| Thermal tracking | Representative 15 min sequence | Zero-shot cross-modality tracking | TTC 88.3% | Slower re-association after prolonged occlusion |

## 4. Discussion

Although the described system is motivated by smuggling-related maritime security scenarios, the present study evaluates only the enabling layers of the solution, namely, persistent sensing, vessel detection, and tracking. Dedicated behavior-analysis and anomaly detection modules capable of inferring smuggling-related events are left for future work. From a scientific standpoint, this manuscript is interpreted as a deployment-oriented autonomous systems paper rather than as a new AI-methods paper. Its main contribution is to show how persistent tethered aerial sensing, maritime vehicle integration, fail-safe operation, and real-time perception can be combined into a field-validated surveillance capability under realistic maritime constraints. The AI results should therefore be read primarily as evidence that the selected perception stack is sufficiently robust for deployment in the proposed architecture, not as a claim of state-of-the-art methodological novelty in object detection or tracking.

Nonetheless, the current AI pipeline demonstrates that state-of-the-art object detection models can be effectively applied to maritime surveillance tasks when trained on appropriate datasets. The model trained on SeaDronesSee delivers strong performance for vessel detection in a variety of conditions. However, the analysis also reveals important limitations, particularly in detecting small, distant objects and rare classes such as life-saving appliances. Addressing these issues will likely require augmentation of the training data

with additional examples and possibly the use of specialized loss functions or sampling strategies to mitigate class imbalance.

From a systems perspective, the interplay between detection and tracking is also important. Reliable detection is the foundation for stable tracking, but tracking can also mitigate some detection shortcomings by maintaining object identities across occasional missed detections. The chosen tracking algorithm, BoT-SORT, has demonstrated good performance in this regard, particularly when integrated into a pipeline that runs at sufficient frame rates.

### *4.1. Operational Readiness and Limitations*

The field tests provide evidence that the SMAUG platform is approaching operational readiness for certain classes of missions, particularly in ports and nearshore waters under moderate sea states. The system can autonomously deploy the drone, maintain persistent surveillance and recover the drone even while the host vessel is moving. It also provides real-time or near real-time visual information that can feed AI algorithms and human decision-making.

At the same time, several limitations remain. First, although the platform demonstrated autonomous take-off, hovering, and landing from moving vessels, flight stability is still sensitive to demanding maritime conditions. Position estimation and control at maximum altitude, particularly under high winds, gusts, or aggressive vessel maneuvers, exhibited oscillatory behavior that would be undesirable in production scenarios. These oscillations are likely caused by the combined effects of vessel motion, wind disturbances, tether dynamics, and the current limitations of the estimator and controller tuning. While they did not prevent operation during the reported trials, they reduce flight margin, may affect image quality for perception, and could limit deployment in harsher sea states. Further tuning of the control loops, improvements in state estimation, and tighter fusion of onboard and platform navigation sensors will be necessary to increase robustness and reduce tether-induced motion.

A second limitation concerns the environmental robustness of the perception pipeline. Detection performance remains more challenging at long ranges, where vessels occupy only a small number of pixels and may blend with the horizon or background clutter. This difficulty is amplified in adverse maritime conditions, including glare, reflections, haze, wave texture, and rapidly changing illumination. As a result, although the reported results confirm the feasibility of real-time detection and tracking, performance is not yet uniform across all operational scenarios. In particular, small and distant targets, as well as targets observed under unfavorable contrast conditions, remain a source of missed detections and less stable tracking. This indicates that the current AI models, while effective, still require additional maritime-specific data, improved augmentation strategies, and broader validation under diverse environmental conditions before they can be considered fully mature for operational deployment.

A third limitation is that the experimental validation is still limited in scale and scope. The field trials in Trondheim and Valencia demonstrated the feasibility of the integrated concept in realistic conditions, but they do not yet constitute a statistically broad operational assessment across seasons, ports, traffic densities, or sea states. The number of deployment environments remains limited, and the experiments were primarily designed to validate integration, autonomous operation, and vessel detection/tracking feasibility rather than providing exhaustive long-term benchmarking.

Another important limitation is related to communications infrastructure, which proved to be a critical dependency on the overall system. The experiments with satellite internet demonstrated that moving platforms impose additional constraints on antenna in-

stallation, pointing stability, and electromagnetic compatibility. Even when local recording and onboard processing remained functional, real-time remote transmission could degrade due to vessel motion, structural occlusion, or link instability. This is particularly relevant because persistent surveillance systems become substantially more valuable when their outputs can be transmitted reliably to remote operators or command centers. For robust operational use, especially in remote areas relying on satellite links, dedicated maritime-grade communication solutions and more resilient networking architectures may be necessary.

### *4.2. Comparison of Existing Surveillance Approaches*

The proposed SMAUG platform is intended as a complementary surveillance capability rather than a universal replacement for existing port and coastal monitoring systems. To better position its contribution, it is useful to compare it with three relevant alternatives: fixed camera surveillance, conventional battery-powered UAV deployment, and UAV-only operation without a mobile maritime carrier. The following comparison is based on operational characteristics and field observations from the present study rather than on controlled head-to-head benchmarking, which is left for future work.

#### 4.2.1. Fixed Camera Surveillance vs. Proposed System

Fixed camera systems are widely used in ports because they provide continuous monitoring, stable power supply, and relatively low recurring operational cost once installed. Their main strengths are persistent observation of predefined areas, straightforward integration with existing security infrastructure, and predictable imaging geometry. However, their surveillance coverage is inherently limited by installation location, field of view, and line-of-sight occlusions caused by ships, cranes, warehouses, and other port structures. As a result, blind spots remain difficult to eliminate completely, especially in complex and dynamic harbor layouts.

By contrast, the proposed tethered UAV system provides an elevated and relocatable viewpoint that can be repositioned according to the operational scenario. This increases effective surveillance coverage in areas that are partially hidden from shore-based cameras and enables observation of vessel movements in port approaches, coastal inlets, and nearshore transfer zones. In that sense, the main advantage over fixed systems is not raw persistence at a single point, but adaptive geometric coverage.

From an endurance perspective, fixed cameras remain superior to permanent monitoring of static zones, since they can operate continuously for very long periods with little interruption. Nevertheless, the tethered UAV narrows the endurance gap substantially compared to conventional UAVs by enabling flight times of many hours, which makes it suitable for persistent missions over selected areas of interest.

In terms of operational cost, fixed systems are generally cost-effective when the area of interest is stable and well covered by infrastructure. However, extending coverage often requires installing multiple cameras, masts, power supplies, and communications links. The proposed system may therefore become advantageous in scenarios where surveillance needs are dynamic, mobile, or concentrated in locations where new fixed infrastructure would be costly or impractical.

Regarding detection performance, fixed cameras benefit from stable viewpoints and can yield high image quality when targets pass through well-observed sectors. However, their effectiveness degrades when targets remain distant, appear in occluded corridors, or exploit geometric blind spots.

The proposed adjustable elevated platform improves target visibility precisely in such situations by reducing occlusion and providing more favorable observation angles. At the same time, the aerial platform introduces its own challenges, including motion,

wind-induced oscillations, and changing scales with altitude. Therefore, the proposed system should be understood as improving detectability in difficult geometries, rather than universally outperforming fixed cameras in all detection conditions.

#### 4.2.2. Battery-Powered UAV vs. Tethered UAV

Conventional battery-powered UAVs offer rapid deployment, flexible maneuverability, and broad availability. They are well suited for short inspections, event-driven monitoring, and missions where access to the area of interest is difficult. Their principal limitation in maritime surveillance is endurance. Typical multirotor platforms can remain airborne only for tens of minutes, after which they require re-turn, battery replacement, and relaunch. This significantly interrupts observation continuity and complicates long-duration missions such as sustained vessel monitoring or extended surveillance of a port approach.

The tethered UAV presented in this work addresses this limitation directly by supplying power through the cable from the host platform. In operational terms, this changes the role of the aerial vehicle from a short-duration scout to a persistent elevated sensor. The resulting increase in endurance is one of the central advantages of the proposed system, particularly for surveillance missions where continuous observation is more valuable than rapid free-flight maneuvering.

The tether also provides a robust physical link for power and data, which can improve communication reliability in electromagnetically complex port environments. In exchange, the tethered configuration imposes constraints on mobility, maximum operating radius, ascent/descent dynamics, and sensitivity to tether mechanics. A battery-powered UAV remains superior when long horizontal displacements, infrastructure bypassing, or aggressive pursuit maneuvers are required.

From a cost perspective, battery-powered UAVs may have lower initial system complexity, but their repeated recovery, battery handling, charging logistics, and operator workload increase over long missions. Tethered systems shift part of this burden toward a more complex launch and recovery architecture, but they reduce the operational interruptions associated with frequent battery cycling. Consequently, for persistent surveillance scenarios, tethered systems can offer a more favorable operational profile even if the hardware integration is more demanding.

With respect to detection performance, both battery-powered and tethered UAVs may use similar cameras and perception algorithms, so there is no intrinsic algorithmic advantage in the aerial platform alone. However, the tethered system can support more stable long-duration observation from a maintained elevated position, which can be beneficial for tracking continuity and repeated observation of the same targets over time. Conversely, a free-flying battery UAV may obtain better views temporarily by changing viewpoint more aggressively. The practical trade-off is therefore between persistence and observation stability on one side and freedom of reposition on the other.

#### 4.2.3. UAV-Only System vs. USV + UAV System

A UAV-only surveillance concept can be effective when the mission area is geographically limited or when operations are launched from a nearby shore base. Its main advantages are architectural simplicity and reduced dependence on a second autonomous platform. However, such systems are limited by the location of the launch site, the communication range to shore, and—if battery-powered—the same endurance constraints discussed above. Even when the UAV is tethered, a fixed shore-based deployment cannot easily follow moving targets or relocate to areas where surveillance demand shifts during the mission.

The addition of a USV changes the operational concept substantially. In the SMAUG architecture, the USV acts as a mobile power, communications, and launch-and-recovery platform for the tethered UAV. This expands surveillance coverage beyond what a shore-launched UAV can observe persistently, because the sensing node itself can be repositioned over water. The USV can approach port entrances, nearshore corridors, or temporary areas of interest, while the drone provides an elevated viewpoint above that moving base. This combination is particularly relevant in maritime security scenarios where suspicious activity may occur away from fixed infrastructure or move along the coastline.

In endurance terms, the USV + UAV architecture is also advantageous because the long-endurance surface platform sustains the aerial system, allowing the combined system to remain operational for extended periods. The USV effectively provides mobility without requiring the UAV itself to spend energy on long transits. This division of roles is one of the main strengths of the proposed concept.

The main drawback is increased system complexity and cost. Coordinating two autonomous platforms, integrating power and communications, and ensuring safe take-off and landing from a moving vessel are all significantly more demanding than operating a UAV alone. The field trials reported in this paper confirm that this complexity is manageable, but also show that communication infrastructure, vessel motion, and environmental loads remain important practical constraints.

In terms of detection performance, the USV + UAV combination can improve observation opportunities by placing the aerial sensor closer to the area of interest while still maintaining an elevated viewpoint. This can reduce target range and improve visibility compared with a more distant shore-based or static UAV deployment. Better observation geometry generally translates into improved practical detection conditions, especially for small vessels in cluttered maritime scenes. However, this benefit is operational rather than algorithmic; the perception gains arise mainly from improved viewpoint and persistent scene access, not from a different detector design.

### *4.3. Future Enhancement Roadmap*

Several directions for future improvement emerge from this work. On the hardware side, enhancements to the tether management system and position estimation algorithms could further reduce motion-induced oscillations, especially at higher altitudes and in more severe sea states. Integrating more powerful onboard computing hardware would open the possibility of executing detection and tracking algorithms on the drone or USV, reducing dependence on remote processing and external connectivity.

On the algorithmic side, collecting and annotating maritime thermal datasets is a clear priority. This could be done through dedicated data collection campaigns in different ports and coastal areas, capturing various vessel types, environmental conditions and ranges. Synthetic data generation techniques, including simulation and domain randomization, could complement real data and help fill gaps in coverage. Once such datasets are available, they can be used to fine-tune thermal detection models and to train domain-specific re-identification networks for enhanced tracking.

Integration with higher-level behavior analysis and anomaly detection modules is another important avenue. By analyzing trajectories, interaction patterns and deviations from normal behavior, the system could automatically flag suspicious activities such as loitering near restricted zones, rendezvous between vessels in unusual locations or repeated short visits to certain areas. Such capabilities would add considerable value to smuggling detection and broader maritime security tasks.

Table 4 provides a simplified operational comparison between the proposed SMAUG platform and the main alternative approaches used in maritime surveillance. Rather

than serving as a controlled benchmark, the table summarizes the relative trade-offs in coverage, endurance, observation conditions, and practical advantages or limitations. This helps clarify the specific application niche of the proposed tethered UAV–maritime carrier architecture.

**Table 4.** Comparison summary with alternative surveillance approaches.

| Approach | Coverage | Endurance | Main Strength | Main Limitation |
|---|---|---|---|---|
| Fixed Cameras | Static areas | Very high | Continuous monitoring | Blind spots |
| Battery UAV | Flexible local coverage | Low | Fast and maneuverable | Short flight time |
| Static tethered UAV | Elevated local coverage | High | Persistent observation | Fixed position |
| UAV-only | Limit by launch site | Low-Moderate | Simple deployment | Limited reach |
| USV-only | Mobile surface coverage | High | Long endurance | No aerial view |
| USV + battery UAV | Mobile + aerial coverage | Moderate | Better flexibility | Battery limited |
| Proposed Platform | Mobile elevated coverage | High | Persistence + mobility | More complex |

## 5. Conclusions

This paper has presented an integrated system for AI-driven tethered drone surveillance in ports and coastal areas. The system combines a tethered UAV, an autonomous or remotely operated surface vessel and advanced perception algorithms based on deep learning to deliver persistent, flexible maritime surveillance.

On the hardware side, the use of a tethered dronebox mounted on USVs has been shown to provide a stable and enduring aerial platform with flight times up to 50 h and operational altitudes of 60 m. The system can execute autonomous take-offs and landings from moving vessels and can operate in sea states up to around 1 m wave height during the current stage of validation. Integrated safety mechanisms and failsafe modes ensure controlled behavior under a wide range of failure conditions.

On the AI side, the adoption of YOLOv8m as detection system and BoT-SORT as tracking algorithm has enabled effective detection and tracking of maritime vessels in RGB video streams, as demonstrated through dataset-based evaluation and field test analysis. Field trials in Trondheim and Valencia have demonstrated the feasibility and robustness of the integrated platform under realistic operational conditions. They have also identified key challenges, particularly related to communication infrastructure and detection performance at long ranges, which guide ongoing and future work.

The field tests collectively demonstrate that the integrated platform is capable of autonomous persistent surveillance from moving maritime platforms in realistic environmental conditions. The tethered dronebox successfully executed repeated autonomous take-offs and landings from USVs and fast patrol boats, even in sea states with wave heights up to approximately 1 m. Positioning and tether management were robust enough to handle vessel maneuvers at operational speeds, though with some oscillation at higher altitudes. The AI-based detection and tracking pipeline performed well on data for vessel detection in most scenarios, especially at short and medium distances. The experiments also highlighted the sensitivity of long-range satellite communication links to installation and environmental factors, emphasizing the need for careful engineering of the communication subsystem in operational deployments. Extending the training with combined RGB and thermal datasets could strengthen performance and expand the operational scope of the system.

In operational terms, the system provides a powerful tool for port authorities, coast guards and other stakeholders engaged in maritime security and safety. It offers persistent, flexible and largely autonomous surveillance capabilities that complement existing fixed and manned systems. As technology evolves, particularly with improved thermal datasets,

refined control algorithms and more integrated behavior analysis, such systems can play an increasingly central role in combating smuggling, enhancing maritime situational awareness and supporting search and rescue operations.

**Author Contributions:** Conceptualization, A.B.-H. and S.T.; methodology, A.B.-H. and A.F.G.; software, A.B.-H. and A.F.G.; validation, T.H. and H.G.B.; formal analysis, A.F.G. and B.G.; investigation, A.B.-H. and A.F.G.; resources, Á.G., S.T., T.H. and H.G.B.; data curation, A.F.G. and A.B.-H.; writing—original draft preparation, Á.G.; writing—review and editing, A.B.-H., S.T. and A.F.G.; visualization, A.F.G.; supervision, Á.G., A.B.-H. and S.T.; project administration, Á.G.; funding acquisition, Á.G., S.T., T.H. and H.G.B. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work has been supported by the European Commission within the context of the project SMAUG, funded under EU Horizon Europe Grant Agreement 101121129. This work has been partially supported by Grant PID2023-146540OB-C42 funded by MCIN/AEI/10.13039/501100011033.

**Data Availability Statement:** Data available on request due to security restrictions.

**Conflicts of Interest:** Authors B.G. and S.T. were employed by the company Elistair. Author T.H. was employed by the company Maritime Robotics. The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## References

1. Perkovič, M.; Gucma, L.; Feuerstack, S. Maritime Security and Risk Assessments. *J. Mar. Sci. Eng.* **2024**, *12*, 988. [CrossRef]
2. Min, H. Developing a smart port architecture and essential elements in the era of Industry 4.0. *Marit. Econ. Logist.* **2022**, *24*, 189–207. [CrossRef]
3. Talpur, K.; Hasan, R.; Gocer, I.; Ahmad, S.; Bhuiyan, Z. AI in Maritime Security: Applications, Challenges, Future Directions, and Key Data Sources. *Information* **2025**, *16*, 658. [CrossRef]
4. Wang, J.; Zhou, K.; Xing, W.; Li, H.; Yang, Z. Applications, Evolutions, and Challenges of Drones in Maritime Transport. *J. Mar. Sci. Eng.* **2023**, *11*, 2056. [CrossRef]
5. Lv, Z.; Wang, X.; Wang, G.; Xing, X.; Lv, C.; Yu, F. Unmanned Surface Vessels in Marine Surveillance and Management: Advances in Communication, Navigation, Control, and Data-Driven Research. *J. Mar. Sci. Eng.* **2025**, *13*, 969. [CrossRef]
6. Li, J.; Zhang, G.; Jiang, C.; Zhang, W. A survey of maritime unmanned search system: Theory, applications and future directions. *Ocean. Eng.* **2023**, *285*, 115359. [CrossRef]
7. Zhang, B.; Liu, J.; Liu, R.W.; Huang, Y. Deep-learning-empowered visual ship detection and tracking: Literature review and future direction. *Eng. Appl. Artif. Intell.* **2025**, *141*, 109754. [CrossRef]
8. Moraes, M.T.; Alves, M.O.; Amario, M.; Pierott, R.M.R.; Najjar, M.K.; Freire, E.; Haddad, A.N. Marine monitoring using UAVs: A bibliometric and systematic review from 2022 to 2025 addressing macroalgae and organic waste in marine environment. *Estuar. Coast. Shelf Sci.* **2026**, *332*, 109770. [CrossRef]
9. Leyva-Mayorga, I.; Martinez-Gost, M.; Moretti, M.; Pérez-Neira, A.; Vázquez, M.Á.; Popovski, P.; Soret, B. Satellite Edge Computing for Real-Time and Very-High Resolution Earth Observation. *IEEE Trans. Commun.* **2023**, *71*, 6180–6194. [CrossRef]
10. Coito, J. Maritime autonomous surface ships: New possibilities—And challenges—In ocean law and policy. *Int. Law Stud.* **2021**, *97*, 19.
11. Kaup, M.; Pióro, K.; Deja, A. Use of unmanned aerial vehicles (UAVs) as a component of an advanced monitoring system for port areas. *Procedia Comput. Sci.* **2024**, *246*, 5142–5150. [CrossRef]
12. Jiao, S.; Zhang, G.; Zhou, M.; Li, G. A comprehensive review of research hotspots on battery management systems for UAVs. *IEEE Access* **2023**, *11*, 84636–84650. [CrossRef]
13. Wu, J.; Li, R.; Li, J.; Zou, M.; Huang, Z. Cooperative unmanned surface vehicles and unmanned aerial vehicles platform as a tool for coastal monitoring activities. *Ocean. Coast. Manag.* **2023**, *232*, 106421. [CrossRef]
14. Li, Y.; Li, S.; Zhang, Y.; Zhang, W.; Lu, H. Dynamic Route Planning for a USV-UAV Multi-Robot System in the Rendezvous Task with Obstacles. *J. Intell. Robot. Syst.* **2023**, *107*, 52. [CrossRef]
15. He, Q.; Liu, W.; Liu, T.-L.; Tian, Q. Robust coordinated path planning for unmanned aerial vehicles and unmanned surface vehicles in maritime monitoring with travel time uncertainty. *Transp. Res. Part B Methodol.* **2025**, *199*, 103284. [CrossRef]
16. Fattori, F.; Cocuzza, S. Tethered Drones: A Comprehensive Review of Technologies, Challenges, and Applications. *Drones* **2025**, *9*, 425. [CrossRef]

17. Marques, M.N.; Magalhães, S.A.; Dos Santos, F.N.; Mendonça, H.S. Tethered Unmanned Aerial Vehicles—A Systematic Review. *Robotics* **2023**, *12*, 117. [CrossRef]
18. Kownacki, C.; Ambroziak, L.; Ciężkowski, M.; Wolniakowski, A.; Romaniuk, S.; Kulesza, Z.; Bożko, A.; Ołdziej, D. Development and Evaluation of a Tethered Class C3 Hexacopter in Maritime Conditions on the Helipad of a Ferry. *Appl. Sci.* **2023**, *13*, 9396. [CrossRef]
19. Schuchardt, B.I.; Dautermann, T.; Donkels, A.; Krause, S.; Peinecke, N.; Schwoch, G. Maritime operation of an unmanned rotorcraft with tethered ship deck landing system. *CEAS Aeronaut. J.* **2021**, *12*, 3–11. [CrossRef]
20. Dai, Y.; Hu, Z.; Zhang, S.; Liu, L. A survey of detection-based video multi-object tracking. *Displays* **2022**, *75*, 102317. [CrossRef]
21. Nikouei, M.; Baroutian, B.; Nabavi, S.; Taraghi, F.; Aghaei, A.; Sajedi, A.; Moghaddam, M.E. Small Object Detection: A Comprehensive Survey on Challenges, Techniques and Real-World Applications. *arXiv* **2025**, arXiv:2503.20516. [CrossRef]
22. Tang, G.; Ni, J.; Zhao, Y.; Gu, Y.; Cao, W. A Survey of Object Detection for UAVs Based on Deep Learning. *Remote. Sens.* **2023**, *16*, 149. [CrossRef]
23. Ngo, T.B.; Ngo, L.; Thanh Nguyen, D.; Phi, A.V.; Perera, A.; Nguyen, A. UAV-Based Visual Detection and Tracking of Drowning Victims in Maritime Rescue Operations. *Drones* **2026**, *10*, 146. [CrossRef]
24. Chen, Z.; Luo, Z.; Wang, Z.; Huang, Z.; He, Y.; Wen, Z.; Ding, Y.; Xu, Z. A Novel Multi-Core Parallel Current Differential Sensing Approach for Tethered UAV Power Cable Break Detection. *Sensors* **2025**, *25*, 5112. [CrossRef]
25. Aharon, N.; Orfaig, R.; Bobrovsky, B. BoT-SORT: Robust Associations Multi-Pedestrian Tracking. *arXiv* **2022**, arXiv:2206.14651.
26. MinIO, Inc. *Object Storage and AI Report*; White Paper; MinIO: Palo Alto, CA, USA, 2024. Available online: https://resources.min.io/ (accessed on 2 February 2024).
27. Ngo, T.B.; Ngo, L.; Nguyen, D.T.; Phi, A.V.; Perera, A.; Nguyen, A. SeaDronesSee: A Maritime Benchmark for Detecting Humans in Open Water. In Proceedings of the 2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 3–8 January 2022; pp. 3686–3696.
28. Varghese, R.; M, S. YOLOv8: A Novel Object Detection Algorithm with Enhanced Performance and Robustness. In Proceedings of the 2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS), Chennai, India, 18–19 April 2024; pp. 1–6. [CrossRef]
29. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *39*, 1137–1149. [CrossRef]
30. Khanam, R.; Hussain, M. What is YOLOv5: A deep look into the internal features of the popular object detector. *arXiv* **2024**, arXiv:2407.20892. [CrossRef]
31. Wang, S.; Xia, C.; Lv, F.; Shi, Y. RT-DETRv3: Real-Time End-to-End Object Detection with Hierarchical Dense Positive Supervision. In Proceedings of the 2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Tucson, AZ, USA, 26 February–6 March 2025; pp. 1628–1636.
32. Chen, F.; Zhang, Y.; Fu, L.; Hua, R.; Zhang, Q.; Bi, S. A Comparative Review of the Next-Generation YOLO Models: YOLOv10 and YOLO11. *J. Comput. Sci. Artif. Intell.* **2025**, *3*. [CrossRef]
33. Milan, A.; Leal-Taix'e, L.; Reid, I.; Roth, S.; Schindler, K. Mot16: A benchmark for multi-object tracking. *arXiv* **2016**, arXiv:1603.00831. [CrossRef]
34. Bergmann, P.; Meinhardt, T.; Leal-Taixe, L. Tracking without bells and whistles. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019.
35. Zhang, Y.; Wang, C.; Wang, X.; Zeng, W.; Liu, W. Fairmot: On the fairness of detection and re-identification in multiple object tracking. *arXiv* **2004**, arXiv:2004.01888. [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.