

Review

Object Detection and Tracking in Maritime Environments in Case of Person-Overboard Scenarios: An Overview

Florian Hoehner ^{1,2,3,*} , Vincent Langenohl ², Suat Akyol ³, Ould el Moctar ¹  and Thomas E. Schellin ¹ 

¹ Institute for Sustainable and Autonomous Maritime Systems, University of Duisburg-Essen, 47057 Duisburg, Germany; ould.el-moctar@uni-due.de (O.e.M.); dr_thomas_schellin@yahoo.de (T.E.S.)

² Naval Optonics for Vessels and Underwater Systems, Hensoldt Optonics GmbH, 73447 Oberkochen, Germany; vincent.langenohl@hensoldt.net

³ Product Development and Research Department, Kreutz & Partner GmbH, 50733 Cologne, Germany; s.akyol@kreutz-partner.de

* Correspondence: florian.hoehner@stud.uni-due.de

Abstract: We examine the current state of the art and the related research on the automated detection and tracking of small objects—or persons—in the context of a person-overboard (POB) scenario and present the associated governing relationship between different technologies, platforms, and approaches as a system of systems. A novel phase model, structuring a POB scenario, comprises three phases: (1) detection, (2) search and track, and (3) rescue. Within these phases, we identify the central areas of responsibility and describe in detail the phases (1) and (2). We emphasize the importance of a high-level representation of different systems and their interactions to comprehensively represent the complexity and dynamics of POB scenarios. Our systematic classification and detailed description of the technologies and methods used provide valuable insights to support future regulatory and research activities. Our primary aim is to advance the development of corresponding technologies and standards.

Keywords: object detection and tracking; person-overboard; situation awareness; maritime search and rescue operation; unmanned aerial vehicle; unmanned surface vehicle



Citation: Hoehner, F.; Langenohl, V.; Akyol, S.; el Moctar, O.; Schellin, T.E. Object Detection and Tracking in Maritime Environments in Case of Person-Overboard Scenarios: An Overview. *J. Mar. Sci. Eng.* **2024**, *12*, 2038. <https://doi.org/10.3390/jmse12112038>

Academic Editor: Weicheng Cui

Received: 19 August 2024

Revised: 31 October 2024

Accepted: 5 November 2024

Published: 11 November 2024



Copyright: © 2024 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 (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The safety and integrity of crew and passengers on board ships are essential in modern seafaring. The number of passengers traveling on cruise ships is already back at pre-pandemic levels in 2023—three years after the COVID-19 pandemic. Forecasts show that passenger traffic will rise steadily over the next few years. By 2027, a global passenger volume of around 40 million is expected, which corresponds to an increase of over 30% compared to pre-pandemic levels. According to the Cruise Lines International Association [1], the cruise industry is one of the fastest-growing sectors in tourism, which is why shipping companies and safety authorities are already facing the challenge of ensuring the highest safety standards for crew and passengers.

There are numerous risks on board a ship that affect the safety of crew and passengers as well as the operation of the ship and the environment. Physical hazards include collisions, fires, explosions, leaks, and capsizing, which can lead to severe structural failures, health hazards, and environmental damage. Mechanical malfunctions, for example, due to defects in engines or control systems, also pose a significant hazard, as they may impair the ship's maneuverability. Health risks include disease outbreaks and injuries. Environmental hazards arise from possible oil and chemical spills as well as the spread of invasive species through ballast water. In addition, there are operational risks, such as navigational errors and extreme weather conditions, that may jeopardize the stability and safety of the ship [2].

A particular risk and probably the most devastating scenario for a person on board a ship, which has attracted increasing attention in the shipping industry in recent years,

is the “person-overboard scenario”—or POB scenario for short [3]. Despite initial efforts to enforce regulations [4] and define corresponding technological standards [5], there are repeated incidents of people disembarking unintentionally or intentionally, which resulted in their deaths.

The quick rescue of a person fallen overboard is crucial, as the chances of survival decrease by the minute [6]. The water temperature as well as the intrinsic dynamics of the sea, the danger of drowning, the possible injury caused by the fall, as well as drowsiness due to shock, panic, or the influence of alcohol, which cannot be ruled out, make an immediate rescue indispensable. To make matters worse, such incidents—especially in the case of intentional disembarkation—are often noticed far too late, which may make previous standards and rescue measures ineffective.

Current security precautions, such as patrols and technologies for monitoring (CCTV technology) of passengers, are not sufficient to capture or automatically detect the exact moment of a POB incident. If a POB incident is nevertheless noticed, the efficiency of the rescue measures may be negatively influenced by other factors, such as the crew’s lack of emergency experience, difficult communication, the extraordinary stress situation, adverse weather, and environmental conditions, as well as reduced visibility.

Especially from a long-term perspective and in view of ever-larger (passenger) ships with an increasing passenger capacity, increased and changing regulations, and internationally increasing passenger volumes, it is important in the future to enable automated and faster action in the event of a POB scenario and to increase situational awareness for the crew of each ship.

In summary, the constantly growing cruise market, the serious danger of a POB scenario, the annual deaths, the stricter regulations and international standards, and the associated research activities in recent years have led to an increased focus on the development of a corresponding technology.

To develop future regulations and technical standards for the POB scenario that ensure increased safety on board ships, it helps to understand and comprehend the technical relationships between subsystems as an overall system for the POB scenario. Accordingly, a holistic overview of the state of the art and the associated research is an important tool for future activities in this area.

However, previous research activities have focused only on individual aspects of a POB scenario: (1) detecting the person fallen overboard, (2) searching for and/or tracking the person in the water, and (3) rescuing this person. To the best of our knowledge, we are the first to present a comprehensive overview of these individual phases and the technologies involved by considering the integration of these three phases. The various aspects of the POB scenario must not be considered in isolation but in relation to their technological interaction with each other. Our aim is to merge the individual subsystems into a holistic overall system to arrive at a comprehensive picture of the state of the art and the related research in this context.

Thus, our primary aim is to introduce a holistic view—in the form of a three-stage phase model—for the “person-overboard scenario” and to describe the relationships as a “system of systems”. We present our phase model in a structured and simplified manner as follows:

Phase 1: Detect

Phase 2: Search and Track

Phase 3: Rescue

Figure 1 presents the POB’s three phases and their principal tasks in a simplified form to describe the relationship between each phase and the general scope of each phase.

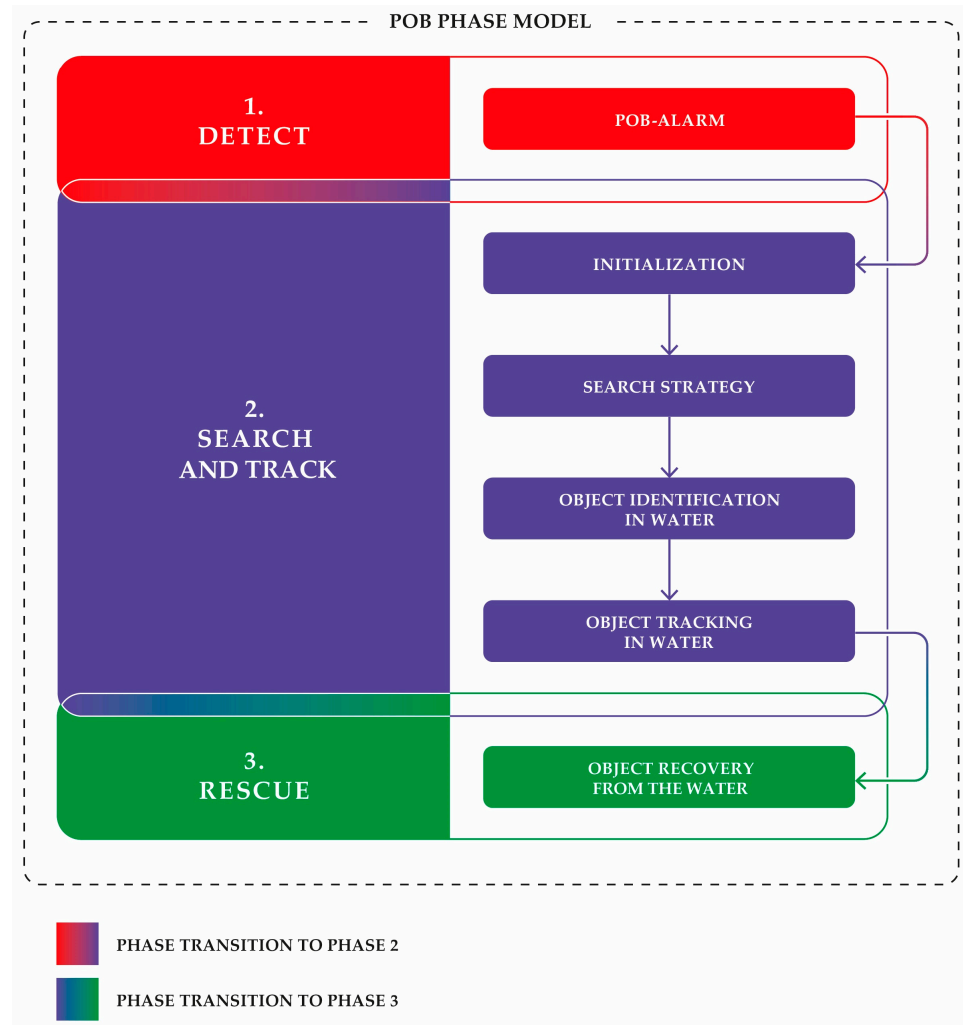


Figure 1. Representation of the POB phase model.

This study analyzes the state of the art as well as the associated challenges for the above-mentioned phases 1 and 2, taking into account possible weaknesses and contradictory approaches in the various systems. It also provides an outlook for phase 3. The structure of this study is as follows:

Section 1 presents the state of the art for Phase 1 (Detection) and Phase 2 (Search and Track), with a focus on the specific tasks within these phases of the POB phase model. To provide a comprehensive and structured overview, technologies and methods are described in detail, and comparisons between different technologies are drawn.

Section 2, which serves as the discussion in this study, addresses the challenges of the approaches described in Section 1. Following the tasks outlined in the POB phase model for Phases 1 and 2, this section highlights difficulties and opportunities for optimization, offering a structured overview of possible improvements and future research activities.

Finally, Section 3 provides the authors' concluding commentary on the approaches and systems that, based on the current state of the art, could serve as ideal solutions for future developments. Additionally, a brief outlook on Phase 3 (Rescue) is provided to support future regulatory and research activities within the context of the POB scenario in a holistic system.

2. Literature Review Approach

To conduct a systematic analysis of the existing research on automated detection and tracking of objects in maritime environments, particularly in the context of POB scenarios, an extensive literature review was performed across several scientific databases. The

primary objective of this review was to identify relevant studies that explicitly focus on the application of modern object detection methods in maritime rescue scenarios, providing a precise overview of the current state of research in this area.

The initial search was conducted in the IEEE Xplore and MDPI databases (time thresholds 2015–2024). For IEEE Xplore, the following search terms were used: (“Abstract”: maritime search and rescue mission) AND (“Abstract”: person detection) AND (“Abstract”: unmanned aerial vehicle) AND (“Abstract”: person-overboard) AND (“Abstract”: deep learning) OR (“Abstract”: maritime surveillance). This search yielded a total of 632 results, distributed across 457 conference papers, 153 journal articles, 14 magazine articles, 6 early-access articles, and 2 books.

Similarly, the MDPI database was queried using the terms: (“Abstract”: maritime) AND (“Abstract”: object detection) AND (“Keywords”: person-overboard) AND (“All fields”: person detection) OR (“Abstract”: aerial dataset) OR (“Abstract”: unmanned surface vehicle), which resulted in 423 articles and reviews.

A key selection criterion was that the methods described in the studies for detecting, searching for, and tracking a person-in-water did not involve additional technologies, such as transmitters, radio transponders, or similar aids. This approach ensures that the examined studies align with current regulatory and standardization requirements [4,5], as future regulations are likely to be based on these standards.

The selection process focused exclusively on studies that could be directly integrated with the structure and tasks of the POB phase model. This approach highlights the complexity of the overall process and was used as a criterion for selecting relevant studies. Studies that did not align with the POB phase model or did not meet the specific requirements of our research questions were excluded. After eliminating irrelevant and duplicate articles, a final selection of 51 papers remained, specifically addressing the challenges of automated detection, searching, and tracking of persons in maritime emergency scenarios. This deliberately narrow focus ensures that our review covers the most significant and relevant research findings in this area.

In addition to the systematic database search, we also reviewed the reference lists of previous scientific summaries, which will be discussed in more detail in the following chapter. These summaries provided valuable insights into additional relevant studies that were not captured by the initial keyword search.

3. Related Work, State of the Art, and an Overview of Papers on POB Scenarios

Numerous previous scientific studies summarize the state-of-the-art research on object detection and tracking of people in the water with the help of various detection sensors and highlight the associated challenges. Table 1 presents a corresponding overview of these scientific works.

Table 1. Provides an overview of latest scientific summaries related to the POB scenario. The references [7–11] in Table 1 represent key review papers that provide focused summaries on POB scenarios, covering detection, search, and tracking. These references were selected for their relevance and depth in addressing critical aspects of the POB scenario.

Ref.	Description	Year
[7]	Provides an overview of deep learning-based methods for object detection in maritime environments. For this purpose, various models are analyzed, in particular the YOLO series and datasets known at the time.	2021
[8]	Provides an overview of the identification of objects on the sea surface using electro-optical sensors and compares traditional with deep learning object detection methods. To do this, it compares advantages and disadvantages of different deep learning approaches to image recognition and structures previous activities into four basic categories: electro-optical sensors, traditional object recognition methods, deep learning methods, and maritime datasets.	2022

Table 1. Cont.

Ref.	Description	Year
[9]	Provides an overview of deep learning-based methods for small object detection (SOD) with a focus on maritime environments. In addition, it summarizes various datasets for the SOD.	2022
[10]	Focuses on the different search strategies using unmanned aerial vehicles (UAVs) for a POB scenario. It examines the applicability of UAVs and analyzes the development and strategies of UAV-based search and rescue operations.	2023
[11]	Presents the key challenges in object detection tasks for the use of an unmanned aerial vehicle (UAV) in maritime environments. The corresponding deep learning methods are compared and a corresponding UAV dataset is presented.	2024
Our present study	Classifies the state-of-the-art research in the context of a POB scenario in a holistic systematic (POB phase model) to represent the relationship between different technologies, platforms, tasks, and approaches as a “system of systems”. Specifically, this overarching presentation, with multitude of interacting systems and approaches, aims to support the development of future regulations and standardizations.	2024

4. Phase 1: Detect

The detection of a POB scenario begins—according to the phase model introduced—with the detect phase. An international standard already exists for this phase [5]. However, before presenting the state-of-the-art research for its implementation, its specific requirements are described.

Such a detection system (DS) must be able to automatically trigger a POB alarm based on the data collected by the sensor units without the need for manual triggering by the crew or for activating an additional emergency device. This ensures that any incident is noticed immediately, even if no crew member directly observes the incident or no appropriate emergency equipment is turned on. In addition, the detection zone of this DS must cover the entire periphery of the ship and must extend to at least eight meters beyond the ship’s periphery. This comprehensive coverage ensures that a person’s every movement, even when going overboard, is detected, regardless of this person’s position on board. This DS must generate both visual and audible alarms that remain active until the POB scenario is confirmed at the control center so that the incident will not go unnoticed and immediate rescue measures can be taken. In addition, verification data in the form of static images or video images must be provided within five seconds of a POB alarm being triggered to authenticate and confirm the incident by a human operator. Finally, the systems should also be able to detect and log false alarms to better understand the behavior of the DS under different environmental conditions and to improve the accuracy of future detections [5].

POB-Alarm

To capture such a POB scenario with the help of sensor units, state-of-the-art research has made available various methods based on using RGB video, thermal, radar, and LiDAR data. Each of these methods offers specific benefits and addresses the challenges of a POB scenario in different ways. Katsamenis et al. [12] analyzed RGB video data for the automated detection of POB scenarios. They presented an unsupervised deep learning model using spatiotemporal auto-encoders to process different image properties, such as appearance, gradient, and salience, simultaneously. As the use of RGB data alone is insufficient to comprehensively detect an anomaly, the system architecture includes an additional analysis method using three separate spatiotemporal auto-encoders. These auto-encoders analyze the RGB video data and extract representative vectors to more precisely detect an anomaly. This approach improves the accuracy of detecting an anomaly by integrating various aspects of image information.

Bakalos et al. [13] presented an alternative approach using thermal imaging devices in combination with spatiotemporal auto-encoders. They developed an unsupervised deep

learning model that relies on thermal data to trigger a POB alarm. Their model employs a convolutional auto-encoder that monitors the errors during image reconstruction. If the reconstruction error exceeds a certain threshold, a POB alarm is triggered. This approach is particularly useful as electro-optical systems often struggle with inadequate lighting, especially at night and/or under conditions of poor visibility. To train the model, a custom dataset of 320 test throws of a human-sized doll from a certain height was created and enriched with additional footage to simulate not only normal conditions but also negative events, such as non-human objects thrown overboard.

Tsekenis et al. [14] described a third approach to detect POB scenarios. In contrast to previous work, theirs is a machine learning-assisted approach using radar data. The system architecture consists of two radar sensors (R1 and R2) arranged at right angles to each other. The field of view in terms of maximum range, speed, solid angle, and distance, as well as speed and angle resolution, of both sensors is identical. The collected data are transmitted to a host processor via a Controller Area Network (CAN), which processes the data and uses it and the incoming information to decide whether the detected object is a human being. To increase the recognition accuracy for detecting a person, a machine learning algorithm based on a trained dataset is used. This pre-processing of the data includes synchronizing the sensor information, cleaning up anomalies, and generating relevant features.

Using LiDAR sensors, Advisory [15] presented another concept to detect a POB scenario. The LiDAR technology determines distances by emitting laser beams and detecting their reflections by means of time-of-flight (TOF) measurements. This process creates a point cloud by sending a laser beam in a controlled direction, measuring the distance to a point in the cloud, and storing this distance accordingly. A 1D LiDAR acts as a simple rangefinder, while a 2D LiDAR creates a linear scan, and a 3D LiDAR captures both horizontal and vertical points, creating a more comprehensive spatial picture [16]. For the concept they presented, the LiDAR sensor presumably detects interruptions in the reflected signals when a person enters its detection range. This change is detected by comparing the received signals and an expected pattern to identify deviations that indicate a POB scenario. In addition, the system estimates the size of the detected objects based on their distance, which increases detection accuracy. One advantage of such a system is the occurrence of a so-called timeout in the absence of reflections, which means that the LiDAR sensor no longer waits for a reflection. In short-range scenarios, where there is a clear reflection, the sensor continues to operate in normal operating mode to ensure reliable detection [15]. Published information on the system's architecture, the test environment, and the database is not unavailable for this concept.

The use of RGB video, thermal, radar, and LiDAR data thus offers a selection of heterogeneous technologies for the automatic detection of POB scenarios. The analysis of RGB video recordings allows for a detailed examination of visual properties, such as object textures. By using the above-mentioned deep learning models, various image properties, such as appearance, gradient, and salience, may be processed to identify precise anomalies. However, a disadvantage of this method is that its effectiveness depends heavily on the visibility. In haze, fog, rain, low-light conditions, such as at night, or strong sunlight and reflection, the detection accuracy is significantly affected. Thermal imaging equipment offers a valuable addition here. This enables reliable detection, even under poor visibility, as this detects the thermal radiation of objects. Therefore, this technology is particularly useful at night and in environments with poor lighting. One disadvantage of thermal imaging devices compared to RGB cameras is their limited resolution. For example, RGB cameras nowadays offer high resolutions of up to 7680×4320 pixels for 8K videos and even higher resolutions for photos, whereas thermal imaging devices achieve significantly lower resolutions. Long-wave infrared (LWIR) cameras typically offer a resolution of only 640 pixels, and medium-wave infrared (MWIR) cameras have a resolution of around 1280 pixels. This lower resolution means that the image detail and sharpness of thermal imaging devices are significantly lower than those of RGB cameras. They may also lose

accuracy under extreme weather conditions, such as heavy rain or snow, as these conditions can disrupt heat signatures.

Radar data, on the other hand, offers an additional dimension that is particularly useful under extreme environmental conditions. Radar sensors are able to detect movements and distances over a very long distance, regardless of light and weather conditions. This makes it a robust extension for detecting POB scenarios in harsh environments. However, a disadvantage of this method is the lower resolution compared to optical systems, which makes it difficult to detect smaller objects or fine movements. LiDAR sensors combine some of the advantages of the previous methods. One of the main advantages of LiDAR sensors is the high accuracy and resolution of the generated three-dimensional (3D) point clouds, which makes it possible to precisely detect and measure even small objects and subtle movements (distance, angle, and speed). However, one disadvantage of LiDAR sensors is their vulnerability to extreme weather conditions, such as direct sunlight or heavy rain or fog, which may affect the quality and thus the accuracy of the data.

Theoretically, the specific strengths of each method improve the ability to recognize the object and, possibly, compensate for the weaknesses of the other methods. To provide a comparative overview of the sensors used, Table 2 lists their characteristic properties and evaluates their performance. In summary, RGB video data provide detailed visual analyses; thermal data reliable results under low-visibility conditions; radar data robust detection under extreme environmental conditions; and LiDAR sensors precise 3D detection, even of small objects and movements.

Table 2. Comparative sensor characteristics were derived based on the findings presented in this chapter. The following comparisons were made by the authors at a high level to provide a simplified overview. For this reason, the evaluation categories were deliberately kept broad and relatively general: (—) very poor performance and/or very limited applicability; (−) poor performance and/or limited applicability, (+) good performance and/or good applicability; (++) superior performance and/or superior applicability.

Sensor	Object Texture	Fog or Rain	Darkness	Sun's Reflection	Range
RGB camera	++	—	—	—	+
Thermal camera	+	+	++	—	—
Radar	—	++	++	++	++
LiDAR	++	—	+	—	+

5. Phase 2: Search and Track

The Detect phase is followed by the Search and Track phase. The content of this second phase comprises four tasks: initialization, search strategy, object identification in the water, and object tracking in the water. In the following, we describe the state-of-the-art research. Various platforms are needed to accomplish the associated tasks: unmanned surface vehicles (USVs), unmanned aerial vehicles (UAVs), and a combination of UAVs and USVs.

Initialization is the first step and forms the basis for the subsequent measures. In this section, the essential parameters are defined by different detection sensors to ensure a targeted and efficient search. The search strategy focuses on the selection and implementation of optimal search patterns and algorithms that are specifically adapted to dynamic maritime conditions. This is followed by object identification in the water. Various detection sensors and detection algorithms are used to ensure that the person-in-water is quickly and reliably detected in the water. To perform object tracking in the water, the technologies for continuous tracking of the person's position are described. Figure 2 presents an overview of the research activities assigned to the tasks in phase 2.

In the following part, the state-of-the-art research is described in alignment with the tasks outlined in Figure 2 from Phase 2. Additionally, the known datasets relevant to this phase will be presented and discussed in detail at the end of this part.

5.1. Introduction to Image Processing

Continuous advances in deep learning have led to efficient solutions for detecting, classifying, and locating objects in image and video data. Of particular interest is the development of new neural network architectures, as their remarkable performance in certain applications is designed to mimic human perception.

Convolutional Neural Networks (CNNs) have established themselves as one of the most powerful methods for image and video processing. They are particularly suitable for tasks such as object recognition, classification, and segmentation. CNNs use a special architecture, consisting of a series of convolution and pooling layers to extract and process features from image data. This architecture allows CNNs to recognize complex patterns and structures in image data, making them ideal for applications in the maritime environment [17]. A key feature of CNNs is their ability to scale and adapt to different image sizes and resolutions. This is particularly useful in real-time scenarios where, for example, UAVs need to continuously process large amounts of data to make decisions quickly and accurately [18].

Region-based Convolutional Neural Networks (R-CNNs) are an advanced method, introduced by Girshick [19], that first identifies potential objects in an image by generating a series of regional suggestions. These regional suggestions are then passed through a CNN to extract features that are eventually used by a classifier to determine if the regions contain relevant objects. Following the introduction of the original R-CNN, several improvements have been developed to further increase efficiency and accuracy. Fast R-CNN [19] and Faster R-CNN [20] significantly reduce the computational time and improve the accuracy by introducing region proposal networks (RPNs) that generate region proposals directly in a CNN. In the case of a POB scenario, R-CNN-based approaches enable the accurate detection and tracking of people in the water. UAVs equipped with RGB cameras, thermal cameras, and other sensor systems can scan large areas and capture high-resolution images. These images are then processed by R-CNN algorithms to identify people in the water who have fallen overboard.

The YOLO (You Only Look Once) framework is also an outstanding model for real-time object detection in the context of a POB scenario. YOLO differs from traditional object detection approaches in its ability to integrate the entire detection pipeline in a single step—which is why this detection method is also called “One Stage Detection”. This allows for faster and more efficient recognition by dividing the image into grid cells and examining each cell to see if it contains an object [21]. YOLO’s speed and accuracy make it ideal for applications that require real-time performance, such as the POB scenario.

5.2. Introduction of USV/UAV Technology for POB Scenarios

From the state-of-the-art research, it is known that USVs can operate fully or partially autonomously on the water surface. By using navigational and communicational systems and other sensor technologies, USVs may be used to search for people in the water.

Compared to traditional manned vessels, unmanned USVs offer a variety of tactical and technical advantages. These include, for example, their compact size, high speed, and flexible maneuverability, eliminating the risk of human losses. These features enable USVs to handle a variety of maritime tasks in complex marine environments independently and without human control [22]. They are equipped with various sensors and communication systems, such as the Automatic Identification System (AIS), various cameras (RGB and/or thermal cameras), LiDAR systems, maritime radar, GPS, and an Inertial Measurement Unit (IMU). The interaction of these sensors enables comprehensive and precise detection of the environment, enabling continuous monitoring and real-time data transmission [23]. This makes USVs ideal for tasks such as search and rescue operations. Their use is particularly advantageous in dangerous or hard-to-reach areas and under adverse environmental conditions, where the use of manned ships might be too risky or inefficient.

In addition, the use of UAVs represents a promising technology for a wide range of applications in the maritime context, including search and rescue operations. Advances in

the computing power of UAVs, in particular through the integration of specialized artificial intelligence (AI) chips, allow them to not only collect data but also process these data in real-time. This supports timely and informed decisions in a POB scenario. UAVs also offer individual advantages that make them particularly suitable for searching and tracking people in the water in the event of a POB scenario. Probably the greatest advantage of a UAV is the ability to operate in the third dimension, i.e., at differing altitudes. While on a ship you only have a two-dimensional view from the side, UAVs can easily see over obstacles such as waves and other ships through a “hawk-eye” perspective. In addition, they are largely stabilized in the air and thus independent of sea swells. This is particularly difficult in rough sea areas, where strong currents, high waves, and turbulent waters prevail, which makes visibility and stability difficult for conventional observation methods [24]. UAVs are still capable of ensuring rapid response and mobilization in emergency situations. Thanks to their ability to quickly cover long distances and reach hard-to-reach areas, UAVs can quickly cover the search area and save valuable time, which can be crucial in critical situations—such as a POB scenario [10]. UAVs can also be equipped with a variety of sensors, including high-resolution RGB cameras, thermal cameras, and LiDAR systems [11]. This allows them to quickly cover large areas of the sea and collect visual information about the suspected rescue areas.

In contrast, USVs offer greater load capacities and the ability to pick up people who have fallen overboard, as well as more powerful computers to process the data collected. In addition, USVs have the advantage of a longer service life: they can remain in one position for a longer period of time without requiring significant energy replenishment, and they are also able to store more energy on board. In contrast, UAVs require a continuous supply of energy, which may significantly limit their flight and mission times [25]. The complimentary use of UAV and USV platforms is advantageous as it combines the advantages of both platforms and thus generates significant synergy effects. In particular, data are fused from multiple sources using UAVs and USVs in an integrative way. These multimodal approaches maximize the efficiency and precision of object detection and tracking by combining heterogeneous datasets and sensor data. Here, models for sensor data fusion are developed to achieve an improved spatial and temporal resolution of the person-in-water. The synergistic use of the different sensory and operational capabilities of UAVs and USVs should, therefore, enable extended and more reliable monitoring of maritime environments [22].

5.3. Initialization

The initialization process forms the data-based foundation for the subsequent search and tracking activities in a POB scenario. This involves determining the initial parameters to identify the likely areas of interest based on the last known position of the person who fell overboard, current weather conditions, and/or prevailing current conditions. Initialization also includes coordinating the movements between the platforms used to ensure optimal coverage of the search area.

Before the search and tracking activities can begin, environmental conditions, such as sea states, ocean currents, and winds, are analyzed. This is performed with the help of predictive models that simulate drift motion relative to the position of the person-in-water. In the system of Ramírez et al. [25], an artificial neural network (ANN) is used for this purpose, which has been trained with appropriate maps of ocean currents and winds. This ANN makes it possible to predict the movements of the person-in-water and to adjust the search area accordingly. During initialization, the UAVs and USVs are initialized with the collected data. The UAVs, which can operate at various altitudes and arrive at the search area faster due to their higher speed, immediately start collecting data by visual monitoring of the area. These data are continuously analyzed and transmitted to the USV, which is equipped with powerful computing units and can precisely determine the position of the person-in-water using a particulate filter (PF) consisting of an ANN and a sensor model.

The PF combines the forecasts of the ANN with the currently measured data to calculate the most likely positions of the person-in-water.

According to Feraru et al. [26], the initialization process already begins with the recording of the last known position of the person-in-water. This information is provided by the POB system on board the vessel, which usually consists of several sensor units (see Phase 1) that detect the event and determine the location via GNSS (Global Navigation Satellite System). Once the POB alert has been received, an additional request is made for up-to-date weather data, including wind and ocean current information. This information can then be used to determine areas of interest and to estimate the approximate position of the person-in-water. Leeway models, e.g., Allen et al. [27], consider wind speeds and surface currents to predict the likely position/drift of the person-in-water. Such models combine different wind vectors to determine the main wind direction, which is the main factor responsible for the drift of a person-in-water [10].



Figure 2. Research activities [10,22,23,25,26,28–49] assigned to tasks in phase 2.

5.4. Search Strategy

Developing an effective search strategy is also crucial to increase the likelihood of successfully locating the person-in-water. This section sheds light on the various search patterns and algorithms known for UAV, multi-UAV, and combined UAV-USV systems obtained from state-of-the-art research. The choice of search strategy takes into account the specific characteristics of the search area as well as the technical possibilities of the platforms used.

Dividing the search area, also known as an area of interest, is an important step in a POB scenario. Due to the influences of winds and currents, the position of the person-in-water changes over time. UAVs offer the opportunity to work cooperatively by using several drones at the same time to systematically search large areas. The UAVs communicate and coordinate their trajectories to avoid overlaps and optimize the coverage of the search area [28]. Heuristic algorithms inspired by natural behaviors, such as the particle swarm optimization algorithm (PSO) and other genetic algorithms, are used to optimize task distribution and path planning in multi-UAV systems. Such algorithms allow UAVs to dynamically adapt and change their behavior based on real-time conditions. This cooperative strategy improves coverage and increases the likelihood of quickly finding people who have fallen overboard. UAVs can share their data with each other and with a command center in real-time, which may further improve the coordination and efficiency of the search during a POB scenario [10,29].

Trong et al. [30] described the brute force search method as an essential approach to improve search and rescue operations in a maritime environment. This method is characterized by systematic and complete coverage of a search area, with UAVs following a predetermined grid pattern to ensure that each area is fully searched. During this search operation, the UAVs continuously collect visual data, which is analyzed in real-time to identify possible victims. Once a potential victim is detected, the UAVs adjust their search strategy and switch from systematic grid searches to targeted approaches to locate the victim more precisely.

Alternatively, the adaptive search strategy introduced by [31] shows a combination of advanced image processing and machine learning to identify areas of interest (ROI) in the captured video streams of the UAVs. These UAVs continuously collect and analyze visual data in real-time to dynamically adjust ROI to make search operations more dynamic. The adaptive search algorithms automatically prioritize the target areas according to the probability of finding people in the water there. This means that areas with high anomaly rates are monitored more intensively, while less likely zones are inspected less frequently. These prioritizations are continuously adjusted based on the continuously collected data from the UAVs.

Feraru et al. [26] presented another UAV system to assist in a POB scenario. The UAV considers the search area as an expanding ellipse, the center of which is based on the last known position of the person and whose periphery depends on the elapsed time and the current environmental conditions. The UAV's search strategy follows a pre-planned path in the form of a zigzag pattern, also known as the Boustrophedon movement. This search strategy ensures that the entire search area is covered and minimizes the chances of missing the person in the water.

The search strategy of a heterogeneous UAV-USV system, on the other hand, is based on cooperative control methods aimed at maximizing the efficiency and accuracy of search operations. A central element of the search strategy of Xue et al. [22] is the cooperative path planning and control of the movements of UAVs and USVs. It uses a "leader-follower" consensus strategy, where a UAV acts as the virtual leader and the USVs follow the leader's movements. Due to their "hawk-eye" perspective, UAVs offer precise monitoring of the USVs and quick identification of target objects. Another element of the search strategy is distributed consensus control. This allows USVs to share information with each other and coordinate their movements, even under suboptimal communication conditions. In addition, during search operations, the UAVs continuously collect visual data, which is

analyzed in real-time and forwarded to the USVs. This real-time data integration allows USVs to dynamically adjust their search paths and respond to changes in the environment.

In addition, Yang et al. [32] described a similar system based on the combination of UAVs and USVs using their mobile computing capabilities to optimize autonomous path planning in search operations during a POB scenario. These systems form a cognitive mobile computer network for cooperative search operations using the reinforcement learning (RL) method to plan search paths. The search area is modeled using the grid method, and a communication architecture based on UAVs and USVs is designed to support communication within the group. Path planning is conducted through the strategy iteration of the Markov decision-making process. Simulated results showed that path planning is successfully completed and that RL offers better throughput performance compared to traditional methods for the same number of UAVs.

Ramírez et al. [25] also presented a coordinated rescue system that combines UAVs and USVs to determine the position of people who have fallen overboard. This system uses the complementary capabilities of UAVs and USVs to ensure fast and precise location of people in the water. An essential element of the presented system is the use of artificial neural networks (ANNs) and particle filters (PFs) to predict and track a person's position in the water. The UAV uses these predictions to determine the person's position and begins the search using a controller that has been trained in search behavior. The USV, meanwhile, uses a particulate filter that integrates the measurement data collected by the UAV and USV to more accurately calculate the position of the person.

5.5. Object Identification in Water

Identifying objects—or a person—in the water is a significant challenge in a POB scenario. This section covers the different search and detection algorithms used on different platforms (USVs and/or UAVs) to identify people-in-water. The integration of advanced image processing and pattern recognition algorithms plays a central role in increasing the accuracy and reliability of identification.

The research activities for object identification in water mentioned here pursue the goal of optimizing the network model used so that an empirical improvement in the search and recognition function of the system can be determined. Each study faces specific challenges that are to be overcome through further developments and optimizations. In the following section, these studies are presented with their concrete challenges, the network model used, and the network optimization to present the state-of-the-art research for the identification of small objects regarding (algorithmic) optimizations.

Shibao et al. [33] investigated the scale-invariant identification of objects in maritime search and rescue operations using UAVs, which are influenced by changes in altitude and angle as well as large waves. To address this challenge, the YOLOv7-tiny network model is used and optimized by special preprocessing and optimization techniques. This includes scaling and perspective adjustment of the images based on the recorded heights and angles (preprocessing scaling) as well as the application of a horizontal and vertical attention scale (attention scaling) for the targeted adjustment of certain image areas. The automatic detection of people-in-water, influenced by different heights and viewing angles, is also investigated in [34]. The network model used in this case is YOLOv4. To optimize the network, multi-frame detection is used to reduce false positives. In addition, the grid zoom method is employed, in which the image is divided into a grid to detect people at high-resolution. Yang et al. [35] and Qingqing [29] investigated also the specific challenges of the scaling differences of objects at different heights as well as the detection accuracy and missing feature information using UAVs in maritime environments. The network model used is YOLO-BEV (Bird's-Eye View), which is based on YOLOv5. To optimize the detection performance, several adjustments are made. A feature fusion module (PAN+) is introduced to enable bidirectional feature fusion and to handle large-scale differences. In addition, a C2fSESA (Squeeze-and-Excitation Spatial Attention) module has been integrated to provide richer feature information. Finally, a light spatial pyramid pooling structure (RGSP) is

used to improve the generalization capability of the model. Coping with scale changes and the detection of smaller objects was also investigated in [36]. A YOLOv5s network model is used, with several optimizations made to improve the identification of objects in the water. An additional layer for the extraction of flat features has been added, as well as a new feature fusion plane and another prediction head, which is responsible for predicting the position and class of objects in an image as part of the neural network. Furthermore, the Conv module and the C3 module have been replaced by the Ghost module and the C3Ghost module to create a lighter network that is better suited for use on UAVs. These modules reduce the number of parameters and computational complexity, making the model more efficient while ensuring accuracy.

Among other things, Kiefer et al. [37] and Kiefer et al. [38] dealt with the challenges of object detection in maritime environments using UAVs and USVs. These specific challenges include the varying water texture, different lighting conditions, different object sizes, and real-time requirements. To overcome these challenges, they compared a number of YOLO variations based on YOLOv7 and/or YOLOv8 with test-time augmentations (TTA) and tested these against each other to improve the accuracy of UAV-controlled detection. USV-based object detection uses convolutional neural networks (CNNs) that improve the localization and classification of objects through adaptive model components. Rizk et al. [39] trained different YOLOv4 variations with different architectural specifications to optimize the performance for object identification in the water using UAVs.

Sneha and Surekha [40] dealt with the automatic detection of drowning persons using UAVs to improve safety and reduce response times during search and rescue operations. Their network model is based on a Conv2dLSTM architecture, which is a combination of CNNs and Long Short-Term Memory (LSTM). This architecture is specifically designed to process 3D data and detect drowning people in real-time. In addition, Lygouras et al. [41] also considered improving the recognition accuracy in the identification of drowning persons using a YOLOv3-tiny network model trained on a specially created dataset.

Zhang et al. [18] dealt with the development of a lightweight identification model for people-in-water to support search and rescue missions. The network model used is YOLOv7-FSB, which is based on YOLOv7. To optimize the efficiency and accuracy of the model, several improvements are implemented. One optimization is the reduction of redundant calculations and memory accesses using FSNet, which significantly increases the efficiency of the model. In addition, feature extraction capabilities are improved using channel reconstruction units. Finally, a bidirectional fusion of features is introduced to make the localization of people in aerial images more effective.

Furthermore, Kiefer and Zell [31] dealt with the rapid creation of Regions of Interest (RoI) in video streams for maritime UAVs. The network model used is based on an auto-encoder-based future frame prediction. To optimize the performance of the model, several improvements are implemented. A major optimization is the reduction of local noise through the Local Noise Remover (LNR), which helps to detect larger areas of error. In addition, averaging past error frames, known as error frame momentum, is used to stabilize temporary errors.

Zhang et al. [42] investigated the challenges of detecting people in maritime environments using UAV imagery under low-light conditions and water reflections. ABT-YOLOv7 is used as a network model, which is based on YOLOv7. To optimize the performance of this model, several adjustments are made. The introduced Asymptotic Feature Pyramid Network (AFPNet) makes it possible to better preserve characteristic information through direct interactions between different hierarchical levels. Small target detection is improved through adaptive computational assignments and context-aware mechanisms. In addition, Task-Specific Context Decoupling (TSCODE) separates the tasks of classification and localization to increase the accuracy of the model. In addition, Zhang et al. [43] investigated object detection in foggy environments to improve accuracy and reduce false detections in maritime environments. The SRC-YOLO model, which is based on YOLOv4-tiny, is used for this purpose. Several approaches are implemented to optimize detection performance.

First, the Single Scale Retinex (SSR) algorithm is used to pre-process the datasets to minimize interference from complex scenes. Furthermore, the detection range of the model is extended to improve the performance of detecting small objects. In addition, channel and spatial information are integrated to extend the model's attention to contextual information and object position points, thus increasing detection performance. To counteract the poor lighting conditions on UAVs, thermal cameras for object identification in the water using Faster R-CNN are employed based on a network model of Feraru et al. [26].

Douguet et al. [23] addressed the challenge of obstacle detection in maritime environments using USVs, with particular attention to factors such as sea disorder and background variability. A YOLOv7 network model is used to optimize detection performance. An important improvement is the combination of 3D coordinate points from the LiDAR sensor with image data to precisely determine the distances and azimuths of objects. This integration enables accurate and reliable detection of obstacles and people in real-time. In addition, Zust and Kristan [44] use USVs to distinguish between real objects and reflections or sunlight on the water surface, employing the network model WaSR-T. To improve segmentation accuracy, a temporal context module is used that extracts temporal information from a sequence of frames. This module helps to distinguish the dynamic appearances from reflections and real objects.

Trong et al. [30] investigated autonomous object identification at sea using UAVs under low-light conditions, employing a YOLOv7-tiny as a network model. To improve recognition performance in low-light conditions, image brightness is optimized and image noise is reduced. This is performed by applying the Self-Calibrated Illumination (SCI) algorithm, which quickly and flexibly improves image quality while minimizing noise. In addition, the X-Plane 11 software and the Ardupilot firmware are used to test realistic flight scenarios in a simulated environment and to evaluate the performance of the developed algorithms.

To improve object identification on UAVs, Kiefer et al. [45] investigated the use of memory maps for a more robust and interpretable representation of object positions. The network model used is YOLOv7-tiny. To optimize video object detection performance, temporal memory maps are created that use metadata to capture object positions in real-world coordinates. This method improves prediction accuracy by aggregating information over time and taking into account 3D geometry. In addition, GPS, compass, and IMU data are used to ensure the accurate localization of objects in 3D coordinates. Additionally, Rizk et al. [46] used other sensor units, such as RGB, infrared (IR), and near-infrared (NIR)—or, as in Gallego et al. [47], multispectral cameras—to improve detection performance for object identification in water from a top-view perspective. The YOLOv4 and YOLOv4-tiny network models ensure more robust detection of objects under different conditions through a wide range of sensor data. Goncalves and Dames [48] employed the YOLOv4-tiny network model to also optimize the automatic detection of rescue targets in maritime search and rescue missions. To improve the performance of the network, random transformations, such as sound change, image rotation, mirroring, and noise addition, are applied. These data augmentation techniques help to increase the robustness and accuracy of the model in detecting rescue targets.

To ensure fast object identification using UAVs, Sahana et al. [28] investigated the optimization of limited resource utilization using a YOLOv4 and MobileNetV3 network model. To optimize the performance of these models, specific adjustments are made. The MobileNetV3-Large and MobileNetV3-Small models are optimized for high and low resource utilization. These optimizations improve the efficiency and accuracy of object detection. Rizk et al. [49] also explored optimizing accuracy and real-time capability for UAVs, despite limited resources. The network models used here include YOLOv4 and YOLOv4-tiny. To optimize the performance of these models, several techniques are used, including FP16 and INT8 quantization, as well as pruning at the channel and layer level, to minimize memory and computational overhead.

5.6. Object Tracking in Water

After identifying the object in the water, continuous object tracking follows to monitor the position of the person-in-water until rescue. This section describes the technologies and methods used to track objects in the water.

In the context of a POB scenario, tracking people who have fallen into the water is of considerable importance for their rescue. Automated tracking of a person in a POB scenario presents a significant number of challenges because the person is in a dynamic and changing environment. These include, for example, the unknown properties of the water, the various environmental influences, the changing light and reflection conditions, the relatively small sizes of the person-in-water, and the partial or complete obscuration of this person by waves or foaming water over a certain period of time. To counteract these challenges, various tracking algorithms—referred to as trackers and listed in Table 3—were extracted from a study by Kiefer et al. [37] for use in a POB scenario.

Table 3. Overview of maritime tracking algorithms based on the findings presented in [37], providing a summarized overview of the tracking algorithms described in that source. The authors have condensed and categorized the information to offer a quick reference to the key characteristics and original sources of each tracking algorithm. This table serves as a high-level comparison to illustrate the core attributes of the algorithms, while detailed explanations and analysis can be found in the original source.

Tracker	Backbone–Network	Properties	Ref.
MoveSORT (based on DeepSORT)	YOLOv7	Optimized for tracking multiple objects in maritime UAV scenarios. ECC (Enhanced Correlation Coefficient Maximization) helps to estimate global rotation and translation between adjacent frames, which is especially useful for fast camera movements. The NSA Kalman filter improves robustness by adaptively adjusting noise covariance based on detection trustworthiness.	[50]
ByteTracker (based on ByteTrack)	YOLOX	High performance in detecting and tracking objects in maritime scenarios, especially by optimizing hyperparameters for smaller (and scattered) objects. Minimizing ID switching through a robust association scheme, resulting in more consistent tracking.	[51]
StrongerSORT (based on StrongSORT)	YOLOv7	High reliability in the association of tracklets (short sequences of object positions) to counteract the risk of occlusion by waves. Reduction of ID changes through a pre-trained, part-based re-identification model to identify body components.	[52]
OCSORT	YOLOX-XL	Precise tracking and low fragmentation in maritime environments to maintain a consistent trajectory and track the exact movement of the object.	[53]

Zhang et al. [18] complemented this consideration by presenting a lightweight model for detecting and tracking people in a POB scenario using a UAV. In their study, they applied YOLOv7 as a backbone network. Their model integrates the detection and tracking algorithms YOLOv7 and ByteTrack to enable real-time tracking. ByteTrack, a tracker listed in Table 3, improves tracking performance by optimizing the association of detections across different frames while minimizing ID changes. ID switching occurs when an object is incorrectly recognized as a new object, causing confusion and inconsistencies in tracking.

In summary, the trackers listed in Table 3 offer solutions with different benefits for tracking people in maritime environments. Kiefer et al. [37] demonstrated that various tracking algorithms are potentially useful in maritime environments for tracking persons-in-water. Integration of tracking algorithms, such as ByteTrack, DeepSORT, StrongerSORT, or OCSORT, and the development of specialized models described by Zhang et al. [18] provide the impetus for further advances in UAV-based tracking of persons-in-water.

Kiefer et al. [45] presented a novel method of object tracking in videos recorded on UAVs. Their presented method uses so-called “memory maps” to store and track the positions of objects in world coordinates. These maps not only store the positions of the objects in real-time but also take into account past predictions to improve the robustness and accuracy of predictions and the existence of objects. The basic idea of the method is based on the use of metadata provided by the UAVs, such as GPS data, compass information, and IMU measurements. These metadata make it possible to determine the relative position of the UAVs and the objects in 3D spatial coordinates. By accounting for the third dimension, they can create a temporal memory map that aggregates the predicted positions of objects, resulting in more robust predictions. An essential part of this method is the calculation of the relative and absolute coordinates of the objects. First, the relative coordinates from an object to the UAV are calculated using the projection of the camera and the pixel position of the object. These relative coordinates are then converted into absolute world coordinates using the UAV’s GPS data. These absolute coordinates are stored in a temporal memory map, which is continuously updated. The memory map not only stores current predictions but also integrates past predictions to increase the robustness of detection. To achieve this, an update of the map is performed at each time step, incorporating error frame momentum to reduce temporal instabilities. It also uses a local noise suppressor to ensure that only relevant anomalies are detected as objects. Through this innovative method, UAVs can not only track the positions of objects in real-time but also improve the detection and tracking of objects over longer periods of time and at different viewing angles and heights. The use of memory maps is groundbreaking as it not only reduces the likelihood of false alarms and significantly increases the accuracy of object detection and tracking, but also effectively limits the problematic areas and weaknesses exacerbated by the maritime environment, such as waves obstructing the view.

In addition, Kiefer et al. [31] described the development of an auto-encoder-based model to predict future frames for resource-constrained object tracking in water. This model is trained on sequences of images of the sea surface to learn to recognize normal patterns and movements of the water. The central idea is that the auto-encoder has difficulty reconstructing objects that differ in color, shape, and texture from the ocean surface. Such differences are identified as anomalies and used to track objects. The architecture of the model follows an encoder-decoder design with small channel dimensions for the filters to minimize computational overhead and enable real-time processing on embedded GPUs.

Kiefer et al. [38] also described the use of a multi-object tracking system (MOT) for maritime applications. For object tracking in water, the system relies on the BYTE Association mechanism. This algorithm links detected objects across consecutive frames and takes into account the camera’s movements to increase the accuracy of tracking. An important component of this system is a hierarchical clustering procedure based on the visual characteristics of the objects. This technique helps to keep object IDs stable over time by fusing and adjusting tracks that are interrupted due to temporary object losses. This minimizes tracking errors and improves the consistency of tracking.

5.7. Published Datasets for POB Scenarios

Available datasets are crucial for the development and optimization of image processing algorithms used in POB scenarios. As the entire field of deep learning is based on learning from existing data, datasets are necessary for training the algorithms to generate intelligent behavior in the first place. Without datasets, it would be impossible to develop algorithms that work reliably and efficiently in real-world applications. The creation and validation of large and diversified datasets covering a variety of realistic scenarios and environmental conditions are of great importance to ensure the performance and robustness of the algorithms. By using annotated image and video data, obtained under different light and weather conditions as well as in different bodies of water and different acquisition methods, the algorithms can be trained for use in a POB scenario. By using annotated image and video data taken under different light and weather conditions as well as in

different bodies of water and different acquisition methods, the algorithms can be trained for use in a POB scenario. This ensures that they function reliably in real use.

Rekavandi et al. [9] already provided a comprehensive overview of generic and maritime datasets for the identification of small objects. These datasets comprise three categories: (i) maritime datasets with persons, (ii) maritime datasets without persons, and (iii) non-maritime datasets with persons.

Maritime records of people contain images and/or videos that capture human activity of people at sea. These data are particularly important for search and rescue operations, as they allow the identification of people during a POB scenario. Examples of this are the SeaDronesSee dataset of Varga et al. [54] and the SeaDroneSee Object Detection V2 dataset of Kiefer et al. [37], which include UAV-based recordings of people (swimmers) in open waters. Another example in this category is the MOBDrone dataset of Cafarelli et al. [55], which also contains UAV images near the coast and captures objects, such as people, boats, and life belts. Research activities of, e.g., Ruiz-Ponce [56] have already indicated in the past that there are only a few available maritime datasets that include image and/or video material of people (swimmers).

Maritime datasets without people focus on objects such as ships or buoys on the water's surface. These data play a crucial role in monitoring and managing shipping to avoid collisions. Examples of such datasets are the SeaShip dataset of Shao et al. [57], which contains images of six kinds of ships, namely, cargo ships and fishing boats, and was compiled from video sections along the coast of China.

Non-maritime records of individuals encompass a wide range of applications that are not related to the maritime environment. These include records for search and rescue operations of individuals in a non-maritime environment. An example of this is the HERIDAL dataset of Bozic-Stulic et al. [58], which was specially designed for search and rescue operations using UAVs in mountainous regions.

The following describes the most relevant public maritime datasets for a POB scenario—i.e., (i) maritime dataset with people—extracted from the state-of-the-art research.

5.7.1. The SeaDroneSee Dataset

The SeaDroneSee dataset of Varga et al. [54] aims to bridge the gap between land-based and maritime visual systems, especially for maritime search and rescue missions. It includes over 54,000 images with 400,000 annotated instances, taken from heights of 5 to 260 m and angles of view from 0 to 90 degrees. The images are from UAVs and include both RGB and multispectral images. This dataset contains meta-information, such as altitude, camera angle, and speed, which enables the development of multimodal systems. The annotations include swimmers, swimmers with life jackets, people on boats (with and without life jackets), and boats alone.

5.7.2. The SeaDroneSee Object Detection V2 Dataset

The SeaDroneSee Object Detection v2 (S-ODv2) dataset of Kiefer et al. [37] extends the original SeaDroneSee dataset of Varga et al. [54] to include more classes and more extensive recordings. The dataset consists of 14,227 RGB images taken from heights of 5 to 260 m and viewing angles of 0 to 90 degrees. The images, which include training, validation, and test data, were taken with six different camera systems. This extended dataset also contains the corresponding meta-information, such as altitude, camera angle, and speed. The annotations include categories, such as floats, boats, jet skis, buoys, and rescue equipment (life jackets/harnesses). A special feature of this dataset is the high-resolution and the variety of images, which facilitate easy maritime object recognition under different conditions.

5.7.3. The SeaDroneSee-MOT Dataset

The SeaDroneSee-MOT (Multi-Object Tracking) dataset of Kiefer et al. [38] expands the previous SeaDroneSee dataset of Varga et al. [54] for maritime object tracking by

introducing re-identification functions. This dataset comprises 54,105 frames, divided into 21 training clips, 17 validation clips, and 19 test clips. The footage comes from UAVs and covers various scenarios of a maritime search and rescue operation. The dataset contains meta-information, such as UAV height, gimbal angle, and GPS coordinates, which enable tracking and re-identification of objects. The annotations include boats and floats, which are recorded in a uniform class system. This extension allows tracking and re-identifying of objects, even if they briefly disappear from view or become obscured.

5.7.4. The MOBDrone Dataset

The MOBDrone dataset of Cafarelli et al. [55] was developed to support the identification and rescue of Man OverBoard (MOB) scenarios. This dataset contains over 126,000 images taken from UAV footage taken in various maritime scenarios. These images were taken from heights of 10 to 60 m and manually annotated to accurately locate more than 180,000 objects, including over 113,000 people. The annotations comprise five categories: people, boats, pieces of wood, lifebuoys, and surfboards. The dataset was subdivided into 66 video clips covering different lighting and height conditions to reflect the variety of real-world rescue scenarios. A special emphasis is placed on capturing people in conscious, semi-conscious, and unconscious states.

5.7.5. The Open Water Dataset

The Open Water dataset of Kiefer et al. [37] was developed to support the automatic detection of anomalies in the maritime environment using UAVs. The dataset includes over 100,000 frames from 60 min of high-resolution (4K) video footage shot on different days, from different heights (5 to 120 m) and angles (0 to 90 degrees). This dataset was manually filtered to include frames with no anomalies, removing objects such as people, boats, lifebuoys, and buoys. This dataset provides additional metadata, such as UAV altitude, camera gimbal angle, time, and GPS coordinates, to enable real-world condition analysis. Additionally, the dataset contains over 20 min of annotated recordings with bounding boxes for the same classes as in the SeaDronesSee dataset to serve as anomalies.

5.7.6. The AFO Dataset

The AFO dataset of Gasienica-Jaszkowy et al. [59] was developed specifically for maritime search and rescue purposes. This dataset consists of aerial photographs from UAVs that include 40,000 hand-annotated people and objects floating in the water. The particular challenge of this dataset is the detection of many small objects that are difficult to identify.

5.7.7. The MaSTr1325 Dataset

The MaSTr1325 dataset of Bovcon et al. [60] is designed to improve semantic segmentation and obstacle detection on USVs. The dataset contains 1325 per-pixel annotated images taken over a two-year period under various realistic maritime conditions. The images were captured with a small USV in coastal waters and cover a wide range of weather conditions and times of day. The annotations comprise three main categories: sea, sky, and obstacles.

5.7.8. The MODS Dataset

The MODS dataset of Bovcon et al. [61] is designed to improve object detection and obstacle segmentation on USVs. The dataset includes about 81,000 stereo images taken while driving a USV. The images contain more than 60,000 annotated objects, such as boats, floats, and other maritime obstacles. These data were taken under different weather conditions and times of the day and covered a variety of realistic scenarios. The dataset provides additional metadata, such as IMU and GPS coordinates, to enable an accurate assessment of obstacle detection.

5.7.9. The WSODD Dataset

The WSODD dataset of Zhou et al. [62] was developed to improve object identification on water surfaces and includes 7467 images taken in different water environments and under climate conditions at different recording times. This dataset contains 21,911 instances classified into 14 common object categories, including boats, ships, balls, bridges, rocks, people, garbage, masts, buoys, platforms, harbors, trees, grass, and animals. In addition, the dataset includes certain metadata, such as GPS coordinates and weather conditions.

5.7.10. Synthetic Dataset Generation

In addition, the studies of Kiefer et al. [63] and Yun et al. [64] describe methods to synthetically generate a category (i) maritime dataset with persons. Game engines, such as Grand Theft Auto V of Games [65] and ARMA3 of Interactive [66], were used to generate realistic training data for object detection in various UAV application scenarios. Synthetic data were generated for this scenario to account for realistic conditions, such as different altitudes, camera angles, and times of day. The use of synthetic datasets offers several key advantages. For one, it improves data availability, especially in cases where real-world data are difficult to access or expensive. In addition, synthetic data allow for greater variability and diversity by simulating a wide range of scenarios, lighting, perspectives, and objects that may not be sufficiently represented in real-world datasets. Another advantage is the automatic and precise creation of annotations and labels, which significantly increases the efficiency and accuracy of algorithm development. However, the use of synthetic data is also subject to negative influences. Thus, synthetic data are unable to fully capture all the subtleties and variations of the real-world, which may mean that models trained solely on synthetic data provide insufficient results when applied to real data [67].

Kiefer et al. [63] investigated the use of synthetic data to improve object identification on UAVs in a maritime environment. Their focus was on the generation and use of synthetic data created with the DeepGTAV platform to cover various application scenarios. The synthetic dataset comprises several high-resolution images that simulate various maritime and terrestrial environments. By using DeepGTAV, metadata such as UAV height, camera gimbal angle, and timestamp could be accurately captured and integrated into the synthetic images.

Yun et al. [64] also dealt with the detection of small targets in search and rescue operations using synthetic data generation. To generate synthetic data, they used the 3D game engine ARMA3, which generates realistic images of search and rescue scenarios in different spatial perspectives, at different times of the day, and with different backgrounds. These synthetic data were used to augment the training data of the target recognition system. Their study showed that supplementing the training data with synthetic images improves recognition accuracy and reduces training costs and times compared to collecting real-world data.

6. Discussion

Here, the challenges of the aforementioned approaches from Phase 1 (Detect) and Phase 2 (Search and Track) will be considered to identify the possible potential for improvement. The various aspects that must be taken into account in the implementation of the respective tasks within the two phases are examined. By analyzing the existing challenges, potential ways to further optimize the systems in the future will be shown.

6.1. POB-Alarm

According to Tsekenis et al. [14], one of the main challenges in the implementation of machine learning-based surveillance systems in maritime environments is the integration and interoperability of different sensor technologies, such as RGB cameras and radar and thermal images. The combination of these technologies requires precise tuning and significantly increases the complexity of the entire system.

Another key issue is the robustness of the systems to different environmental conditions and lighting scenarios. To work reliably under different conditions, the systems must include a range of different sensor technologies. High false alarm rates and detection losses significantly affect the reliability of these systems. Another problem is that the training data often only includes positive events, which makes it difficult to balance false positives with detection probabilities. For example, false alarms that occur several nights in a row can severely affect acceptance and trust in the system. In addition, the development of multi-tiered systems that could potentially mitigate these problems has not yet been sufficiently researched. Another important aspect is the lack of sufficient and high-quality datasets for training the models. Creating realistic and diversified datasets is not only time-consuming but also resource-intensive. This challenge is particularly serious because the models must be trained on a broad database to function reliably [12,13].

6.2. Initialization

One of the most relevant challenges is modeling the movement uncertainty of persons-in-water. These movements are strongly influenced by winds and currents, which increases the uncertainty of their position over time, resulting in an increasing search area [26].

According to Ramirez et al. [25] and Feraru et al. [26], the integration of weather and current data to estimate the drift speed and direction of people who have fallen overboard is another complex challenge. These estimates are based on linear regressions and must account for the uncertainties in the models. The dynamic and changing nature of the maritime environment impedes the combined modeling of these factors.

In addition to external environmental factors, the person's own movement and his ability to swim also play a significant role, as living individuals can actively try to stay on the surface and move around. Further uncertainties arise from unknown factors, such as the presence of buoyancy agents and the wind-prone height of the person in the water. These variables make precise modeling difficult and require extensive consideration in the corresponding models.

6.3. Search Strategy

One of the greatest challenges is the time constraints, as the long duration of search operations significantly reduces the victims' chances of survival. The efficiency and speed of the search efforts are therefore crucial, as the probability of survival depends directly on timely detection. According to Trong et al. [30], a significant challenge is the continuous optimization of search strategies. The flight paths and search patterns must be designed in such a way that the efficiency of the search operation is maximized. This requires the use of path-planning algorithms that can dynamically adapt to the changing conditions during the search. Xue and Wu [22] and Yang et al. [32] highlight that autonomous and cooperative search strategies between multiple UAVs and USVs also help to maximize search area coverage and increase detection rates.

Effective coordination and communication between UAVs and USVs is also a major challenge. This requires robust communication networks that function reliably even in remote or difficult areas. These networks must be flexible and quickly recoverable to compensate for node failures and ensure low latency to ensure real-time communication and coordination [32].

Another critical problem is the temporary submersion of the person-in-water. If an area of interest has already been searched, but the person has just disappeared or was obscured by a wave, it can be easily missed. These additional factors add complexity and require even more precise and repeated searches of areas of interest to ensure that no person will be missed in the water.

6.4. Object Identification in Water

One of the greatest challenges in object identification in water is the varying texture and reflections on the surface of the water. These factors can significantly affect the

ability to accurately detect objects, leading to false alarms and misclassifications. Zust and Kristan [44], for example, highlight the problems that arise from reflections and glitter effects. Kiefer et al. [38] described similar difficulties. In addition, different lighting conditions pose another major challenge. Strong sunlight, twilight, and night conditions can greatly affect the performance of the detection algorithms. While extensive research has been conducted in the field of visual (VIS) image processing, infrared (IR) image processing has not been adequately researched. However, IR technology could potentially help improve detection under difficult lighting conditions. Trong et al. [30] discussed the problems of changing light conditions and their influence on detection accuracy. Alao Bai et al. [36] emphasized the importance of models that must work robustly under different lighting conditions.

According to Shibao et al. [33] and Taipalmaa et al. [34], the identification of small objects—or people—and the associated management of scale variations, different heights, and angles of view are also significant challenges. Small objects, often only a few pixels in size, require high-precision detection algorithms. Research by Bai et al. [36] and Kiefer et al. [38] showed that specialized network architectures are necessary to meet these challenges. Rizk et al. [49] have also investigated the difficulty of detecting small objects and coping with scale variations. In addition, difficult visibility conditions, such as fog or heavy rain, further exacerbate the problem. Zhang et al. [43] emphasized the negative effects of such conditions on recognition accuracy. Sahana et al. [28] also discussed the importance of robust techniques for coping with adverse visibility conditions.

To develop deep learning models for object identification and tracking in maritime environments, Ruiz-Ponce et al. [56] addressed non-standardized datasets. However, their variability and inconsistency reduce performance while they rely on freely available and sufficiently large data. Although extensive and high-quality datasets exist, such as SeaDronesSee of Varga et al. [54], other datasets often show a significant class imbalance. This means that some classes of objects (i.e., people or swimmers) are severely underrepresented, resulting in a bias of the results and a reduced generalization ability of the models. The limited availability and disparity in existing datasets make it difficult for researchers and developers to train models that work reliably and efficiently in real-world applications. In addition, Sahana et al. [28] also mentioned the need for large and well-annotated datasets to effectively train the algorithms, and Kiefer et al. [38] emphasized the importance of dataset quality.

Real-time requirements and limited computing resources are particularly relevant for applications on mobile and embedded devices such as UAVs. Zhang et al. [18] and Rizk et al. [46] underscored the importance of optimizing algorithms for edge devices to achieve high recognition accuracy with minimal computational effort. Trong et al. [30] and Bai et al. [36] also addressed the challenges of real-time processing and resource conservation.

In USVs, on the other hand, the synchronization and fusion of data from different sensor units requires complex and necessary processes to ensure reliable object identification in the water. Douguet et al. [23] described the integration of GPS, LiDAR, radar, and cameras to accurately determine object positions from a USV.

6.5. Object Tracking in Water

Kiefer et al. [37] described the tracking of small, partially obscured objects that change their appearance due to their movement and the degree of occlusion by water as a complex challenge because the constantly changing water surface can lead to frequent false alarms and make it difficult to determine the actual position of the object. In addition, Kiefer et al. [38] and Kiefer et al. [45] emphasized the dynamic and often unpredictable texture of the water surface, combined with strong light reflections, as an additional problem for object tracking in water. Kiefer and Zell [31] illustrated the negative influence of variable lighting conditions on image acquisition and processing as they can lead to underexposed or overexposed areas in the images.

Zhang et al. [18] stressed that the small size of the objects, combined with the corresponding distances, poses a significant challenge in object tracking in the event of a POB scenario. Persons-in-water often appear too small in the footage of UAVs, which may have to operate from heights of more than 50 m, making tracking difficult. In addition, Kiefer et al. [38] and Kiefer et al. [45] pointed out that POB scenarios often require tracking multiple objects at the same time. This multi-goal tracking leads to further challenges, especially in the form of ID changes. These issues affect the accuracy and reliability of object tracking.

7. Outlook for Phase 3 and Future Work

Section 3 provides a final outlook on Phase 3 (Rescue) to support future regulatory and research activities in the context of the POB scenario, considered a holistic system.

Technological advances have led to the development of fully and semi-autonomous systems that can reduce reaction time and significantly increase the chances of survival of people who have fallen overboard. An example is the AGaPaS project of Kurowski et al. [68], where a semi-autonomous rescue system based on a USV was developed. This system includes a remote-controlled rescue vehicle capable of locating and rescuing a person-in-water.

The greatest challenges in the development of such systems lie in the precise localization and safe rescue of persons-in-water under variable and often extreme environmental conditions. Future research activities should focus on improving the autonomy and reliability of these systems. This includes the integration of AI algorithms for the real-time recognition and processing of environmental data as well as the development of robust communication systems that function reliably even under difficult environmental conditions [69].

To provide a final assessment of the state-of-the-art presented in this work, a brief commentary is included on which approaches from this scientific investigation are most suitable, in combination, for a future solution addressing Phases 1 and 2 of the POB phase model. For the detection of a person-overboard, the approach from Katsamenis et al. [12], which utilizes RGB data in conjunction with spatiotemporal auto-encoders to process various image features such as appearance, gradient, and saliency simultaneously, offers a high degree of object texture and enables reliable detection. Additionally, radar, as described by Tsekenis et al. [14], provides some of the most valuable characteristic features for this task. In combination, these two technologies present a complementary solution with the potential to reinforce each other and significantly enhance detection efficiency.

For Phase 2, search and track, the combination of various technologies and platforms appears particularly promising. In particular, the integration of unmanned surface vehicles (USVs) and unmanned aerial vehicles (UAVs) provides a synergistic foundation to address the complex challenges of this phase. By theoretically combining the approaches of Ramírez et al. [25] and Kiefer and Zell [31], the strengths of both approaches could be effectively linked to create an integrated, high-performance system according to every task in Phase 2.

The approach of Ramírez et al. [25] relies on the dynamic adaptation of the search area based on environmental factors, such as ocean currents and wind, which are analyzed and predicted by an artificial neural network. This predictive method enables the precise forecasting of the movements of a person-overboard and the real-time adjustment of the search area. UAVs could efficiently monitor large areas, while USVs could take over tracking in closer proximity. A key strength of this combination lies in the “Search Strategy” task, where both approaches could complement each other and create a highly effective combination between the dynamic adjustment of the search area and the real-time tracking capabilities of the platforms.

This is where the approach of Kiefer and Zell [31] connects, focusing on the rapid identification of objects in the water. By generating regions of interest in video streams captured by the UAVs, suspicious objects can be automatically identified. These data could then be relayed in real-time to the USVs to ensure the precise tracking of the identified

objects. UAVs thus serve as eyes in the sky, while USVs are responsible for detailed tracking of the water's surface.

The combination of both approaches creates a cohesive and effective solution. The neural network from Ramírez et al. [25] provides the foundation for an intelligent and dynamic search strategy, which, when paired with Kiefer and Zell's [31] rapid and automated object detection, enables highly efficient identification and tracking. While UAVs efficiently scan the search area and detect potential objects, USVs ensure targeted tracking. This close collaboration between both platforms, supported by predictive models and real-time data processing, would significantly increase success rates in Phase 2 and reduce the time to rescue.

In addition, future research activities should also focus on the integration of the above-mentioned subsystems into a coherent overall system according to the POB phase model introduced here. This requires the development of standardized interfaces and communication protocols that ensure seamless interaction between the various subsystems as well as the existing onboard safety systems.

Another focus of future research should be on the further development and optimization of the handling and user-friendliness of rescue systems to enable their use in real rescue scenarios. In particular, the use of USVs is often a state-of-the-art technical stand-alone solution [70–72]. Therefore, these systems should be designed and standardized in such a way that they can be stationed on passenger ships in the future, are functional, and can be mobilized efficiently for real use in a POB scenario.

Therefore, future research activities should investigate cost-effective and practicable applications of such systems that could be meaningfully integrated along the POB phase model. The goal should be to implement holistic standardization projects for the POB scenario and to harmonize the interacting technologies in such a way that these systems can be made economically accessible to users. This can further improve situational awareness and safety on board ships.

Author Contributions: Conceptualization, F.H. and O.e.M.; literature search, F.H.; validation, V.L., S.A. and O.e.M.; formal analysis, F.H. and V.L.; writing—original draft preparation, F.H.; writing—review and editing, F.H., V.L., O.e.M. and T.E.S.; visualization, F.H.; project administration, F.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: Author Vincent Langenohl was employed by the Hensoldt Optronics Co., Ltd.; Suat Akyol was employed by Kreutz & Partner Co., Ltd. The remaining authors declare 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. Cruise Lines International Association, CLIA. *State of The Cruise Industry 2023*; Cruise Lines International Association, CLIA: Washington, DC, USA, 2023; p. 54. Available online: https://cruising.org/-/media/clia-media/research/2023/2023-clia-state-of-the-cruise-industry-report_low-res.ashx (accessed on 2 July 2024).
2. Dominguez-Péry, C.; Vuddaraju, L.N.R.; Corbett-Etchevers, I.; Tassabehji, R. Reducing maritime accidents in ships by tackling human error: A bibliometric review and research agenda. *J. Shipp. Trade* **2021**, *6*, 20. [CrossRef]
3. Richard Blumenthal, E.J.M.; Matsui, D.; Doggett, L. *Report to US Coast Guard*; Congress of the United States: Washington, DC, USA, 2023; p. 2.
4. Christensen, E.P. *Cruise Vessel Security and Safety Act*; U.S.C. Guard, Ed.; U.S. Department of Homeland Security: Washington, DC, USA, 2011.
5. *ISO 21195:2020*; Ships and Marine Technology—Systems for the Detection of Persons While Going Overboard from Ships (Man Overboard Detection). ISO: Geneva, Switzerland, 2020; p. 22.
6. Heggie, T.W.; Burton-Heggie, T. Death at Sea: Passenger and Crew Mortality on Cruise Ships. *Int. J. Travel Med. Glob. Health* **2020**, *8*, 146–151. [CrossRef]

7. Zhang, R.; Li, S.; Ji, G.; Zhao, X.; Li, J.; Pan, M. Survey on Deep Learning-Based Marine Object Detection. *J. Adv. Transp.* **2021**, *2021*, 1–18. [\[CrossRef\]](#)
8. Lyu, H.; Shao, Z.; Cheng, T.; Yin, Y.; Gao, X. Sea-Surface Object Detection Based on Electro-Optical Sensors: A Review. *IEEE Intell. Transp. Syst. Mag.* **2022**, *15*, 190–216. [\[CrossRef\]](#)
9. Rekavandi, A.M.; Xu, L.; Boussaid, F.; Seghouane, A.K.; Hoefs, S.; Bennamoun, M. A Guide to Image and Video based Small Object Detection using Deep Learning: Case Study of Maritime Surveillance. *arXiv* **2022**, arXiv:2207.12926. [\[CrossRef\]](#)
10. Liu, S.; Li, X. A Survey on Man Overboard Accident Search and Rescue Technology by Unmanned Aerial Vehicle. In Proceedings of the 2023 35th Chinese Control and Decision Conference (CCDC), Yichang, China, 20–22 May 2023; pp. 1062–1067. [\[CrossRef\]](#)
11. Zhao, C.; Liu, R.W.; Qu, J.; Gao, R. Deep learning-based object detection in maritime unmanned aerial vehicle imagery: Review and experimental comparisons. *Eng. Appl. Artif. Intell.* **2024**, *128*, 107513. [\[CrossRef\]](#)
12. Katsamenis, I.; Bakalos, N.; Karolou, E.E.; Doulamis, A.; Doulamis, N. Fall Detection Using Multi-Property Spatiotemporal Autoencoders in Maritime Environments. *Technologies* **2022**, *10*, 47. [\[CrossRef\]](#)
13. Bakalos, N.; Katsamenis, I.; Karolou, E.E.; Doulamis, N. Unsupervised Man Overboard Detection Using Thermal Imagery and Spatiotemporal Autoencoders. In *Novelties in Intelligent Digital Systems*; IOS Press: Amsterdam, The Netherlands, 2021. [\[CrossRef\]](#)
14. Tsekenis, V.; Armeniakos, C.K.; Nikolaidis, V.; Bithas, P.S.; Kanatas, A.G. Machine Learning-Assisted Man Overboard Detection Using Radars. *Electronics* **2021**, *10*, 1345. [\[CrossRef\]](#)
15. Advisory, I.S. ProEye DTR. Available online: <https://intsa.no/en/> (accessed on 24 July 2024).
16. Jia, D.; Hermans, A.; Leibe, B. 2D vs. 3D LiDAR-based Person Detection on Mobile Robots. In Proceedings of the 2022 IEEE/RISJ International Conference on Intelligent Robots and Systems (IROS), Kyoto, Japan, 23–27 October 2022; pp. 3604–3611.
17. Krizhevsky, A.; Sutskever, I.; Hinton, G. ImageNet Classification with Deep Convolutional Neural Networks. In Proceedings of the 25th International Conference on Neural Information Processing Systems, Lake Tahoe, NV, USA, 3–6 December 2012. [\[CrossRef\]](#)
18. Zhang, Y.; Tao, Q.; Yin, Y. A Lightweight Man-Overboard Detection and Tracking Model Using Aerial Images for Maritime Search and Rescue. *Remote Sens.* **2024**, *16*, 165. [\[CrossRef\]](#)
19. Girshick, R. Fast R-CNN. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015. [\[CrossRef\]](#)
20. 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.* **2017**, *39*, 1137–1149. [\[CrossRef\]](#)
21. Redmon, J.; Divvala, S.; Girshick, P.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016. [\[CrossRef\]](#)
22. Xue, K.; Wu, T. Distributed Consensus of USVs under Heterogeneous UAV-USV Multi-Agent Systems Cooperative Control Scheme. *J. Mar. Sci. Eng.* **2021**, *9*, 1314. [\[CrossRef\]](#)
23. Douguet, R.; Dominique, H.; Laurent, J. Multimodal perception for obstacle detection for flying boats—Unmanned Surface Vehicle (USV). In Proceedings of the OCEANS 2023—Limerick, Limerick, Ireland, 5–8 June 2023; pp. 1–8. [\[CrossRef\]](#)
24. 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. [\[CrossRef\]](#)
25. Ramirez, F.F.; Benitez, D.S.; Portas, E.B.; Orozco, J.A.L. Coordinated sea rescue system based on unmanned air vehicles and surface vessels. In Proceedings of the OCEANS 2011 IEEE—Spain, Santander, Spain, 6–9 June 2011; pp. 1–10. [\[CrossRef\]](#)
26. Feraru, V.A.; Andersen, R.E.; Boukas, E. Towards an Autonomous UAV-based System to Assist Search and Rescue Operations in Man Overboard Incidents. In Proceedings of the 2020 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR), Abu Dhabi, United Arab Emirates, 4–6 November 2020. [\[CrossRef\]](#)
27. Allen, A.; Robe, R.; Morton, E. *The Leeway of Persons-In-Water and Three Small Craft*; U.S. Coast Guard Research and Development Center: Groton, CT, USA, 1999; p. 151.
28. Sahana, M.P.; Sengar, A.; Dubey, A.; Agrawal, U. Person Detection in Maritime Search And Rescue Operations. *Int. Res. J. Eng. Technol.* **2022**, *9*, 1383–1388.
29. Li, Q.; Taipalmaa, J.; Queralta, J.P.; Gia, T.N.; Gabbouj, M.; Tenhunen, H.; Raitoharju, J.; Westerlund, T. Towards Active Vision with UAVs in Marine Search and Rescue: Analyzing Human Detection at Variable Altitudes. In Proceedings of the 2020 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR), Abu Dhabi, United Arab Emirates, 4–6 November 2020. [\[CrossRef\]](#)
30. Do Trong, T.; Khai, V.D.; Duy, T.B.; Van, M.V. A Scheme of Autonomous Victim Search at Sea Based on Deep Learning Technique Using Cooperative Networked UAVs. In Proceedings of the 2023 12th International Conference on Control, Automation and Information Sciences (ICCAIS), Hanoi, Vietnam, 27–29 November 2023; pp. 206–211. [\[CrossRef\]](#)
31. Kiefer, B.; Zell, A. Fast Region of Interest Proposals on Maritime UAVs. In Proceedings of the 2023 IEEE International Conference on Robotics and Automation (ICRA), London, UK, 29 May–2 June 2023. [\[CrossRef\]](#)
32. Yang, T.; Jiang, Z.; Sun, R.; Cheng, N.; Feng, H. Maritime Search and Rescue Based on Group Mobile Computing for Unmanned Aerial Vehicles and Unmanned Surface Vehicles. *IEEE Trans. Ind. Inform.* **2020**, *16*, 7700–7708. [\[CrossRef\]](#)
33. Li, S.; Li, C.; Wang, Z.; Jia, Z.; Zhu, J.; Cui, X.; Liu, J. Pas: A scale-invariant approach to maritime search and rescue object detection using preprocessing and attention scaling. *Intell. Serv. Robot.* **2024**, *17*, 465–477. [\[CrossRef\]](#)

34. Taipalmaa, J.; Raitoharju, J.; Queralta, J.P.; Westerlund, T.; Gabbouj, M. On Automatic Person-in-Water Detection for Marine Search and Rescue Operations. *IEEE Access* **2024**, *12*, 52428–52438. [\[CrossRef\]](#)
35. Yang, Z.; Yin, Y.; Jing, Q.; Shao, Z. A High-Precision Detection Model of Small Objects in Maritime UAV Perspective Based on Improved YOLOv5. *J. Mar. Sci. Eng.* **2023**, *11*, 1680. [\[CrossRef\]](#)
36. Bai, J.; Dai, J.; Wang, Z.; Yang, S. A detection method of the rescue targets in the marine casualty based on improved YOLOv5s. *Front. Neurorobotics* **2022**, *16*, 1053124. [\[CrossRef\]](#)
37. Kiefer, B.; Kristan, M.; Perš, J.; Žust, L.; Poiesi, F.; Andrade, F.; Bernardino, A.; Dawkins, M.; Raitoharju, J.; Quan, Y.; et al. 1st Workshop on Maritime Computer Vision (MaCVi) 2023: Challenge Results. In Proceedings of the 2023 IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (WACVW), Waikoloa, HI, USA, 3–7 January 2023. [\[CrossRef\]](#)
38. Kiefer, B.; Žust, L.; Kristan, M.; Perš, J.; Teršek, M.; Wiliem, A.; Messmer, M.; Yang, C.Y.; Huang, H.W.; Jiang, Z.; et al. 2nd Workshop on Maritime Computer Vision (MaCVi) 2024: Challenge Results. In Proceedings of the 2024 IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (WACVW), Waikoloa, HI, USA, 1–6 January 2024. [\[CrossRef\]](#)
39. Rizk, M.; Slim, F.; Baghdadi, A.; Diguët, J.P. Towards Real-Time Human Detection in Maritime Environment Using Embedded Deep Learning. In *Advances in System-Integrated Intelligence, Proceedings of the 6th International Conference on System-Integrated Intelligence (SysInt 2022), Genova, Italy, 7–9 September 2022*; Springer: Cham, Switzerland, 2022; pp. 583–593. [\[CrossRef\]](#)
40. Sneha, S.; Surekha, P. Maritime Human Drone Teaming For Search and Rescue Operations. In Proceedings of the 2023 IEEE Technology & Engineering Management Conference—Asia Pacific (TEMSCON-ASPAC), Bengaluru, India, 14–16 December 2023; pp. 1–7. [\[CrossRef\]](#)
41. Lygouras, E.; Santavas, N.; Taitzoglou, A.; Tarchanidis, K.; Mitropoulos, A.; Gasteratos, A. Unsupervised Human Detection with an Embedded Vision System on a Fully Autonomous UAV for Search and Rescue Operations. *Sensors* **2019**, *19*, 3542. [\[CrossRef\]](#)
42. Zhang, Y.; Yin, Y.; Shao, Z. An Enhanced Target Detection Algorithm for Maritime Search and Rescue Based on Aerial Images. *Remote Sens.* **2023**, *15*, 4818. [\[CrossRef\]](#)
43. Zhang, Y.; Ge, H.; Lin, Q.; Zhang, M.; Sun, Q. Research of Maritime Object Detection Method in Foggy Environment Based on Improved Model SRC-YOLO. *Sensors* **2022**, *22*, 7786. [\[CrossRef\]](#) [\[PubMed\]](#)
44. Žust, L.; Kristan, M. Temporal Context for Robust Maritime Obstacle Detection. In Proceedings of the 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Kyoto, Japan, 23–27 October 2022. [\[CrossRef\]](#)
45. Kiefer, B.; Quan, Y.; Zell, A. Memory Maps for Video Object Detection and Tracking on UAVs. In Proceedings of the 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Detroit, MI, USA, 1–5 October 2023. [\[CrossRef\]](#)
46. Sharafaldeen, J.; Rizk, M.; Heller, D.; Baghdadi, A.; Diguët, J.P. Marine Object Detection Based on Top-View Scenes Using Deep Learning on Edge Devices. In Proceedings of the 2022 International Conference on Smart Systems and Power Management (IC2SPM), Beirut, Lebanon, 10–12 November 2022. [\[CrossRef\]](#)
47. Gallego, A.J.; Pertusa, A.; Gil, P.; Fisher, R.B. Detection of bodies in maritime rescue operations using unmanned aerial vehicles with multispectral cameras. *J. Field Robot.* **2019**, *36*, 782–796. [\[CrossRef\]](#)
48. Gonçalves, L.; Damas, B. Automatic detection of rescue targets in maritime search and rescue missions using UAVs. In Proceedings of the 2022 International Conference on Unmanned Aircraft Systems (ICUAS), Dubrovnik, Croatia, 21–24 June 2022. [\[CrossRef\]](#)
49. Rizk, M.; Heller, D.; Douguet, R.; Baghdadi, A.; Diguët, J.P. Optimization of Deep-Learning Detection of Humans in Marine Environment on Edge Devices. In Proceedings of the 2022 29th IEEE International Conference on Electronics, Circuits and Systems (ICECS), Glasgow, UK, 24–26 October 2022. [\[CrossRef\]](#)
50. Wojke, N.; Bewley, A.; Paulus, D. Simple online and realtime tracking with a deep association metric. In Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017. [\[CrossRef\]](#)
51. Zhang, Y.; Sun, P.; Jiang, Y.; Yu, D.; Weng, F.; Yuan, Z.; Luo, P.; Liu, W.; Wang, X. ByteTrack: Multi-Object Tracking by Associating Every Detection Box. In *Computer Vision—ECCV 2022, Proceedings of the 17th European Conference, Tel Aviv, Israel, 23–27 October 2022*; Springer: Cham, Switzerland, 2021.
52. Du, Y.; Zhao, Z.; Song, Y.; Zhao, Y.; Su, F.; Gong, T.; Meng, H. StrongSORT: Make DeepSORT Great Again. *IEEE Trans. Multimed.* **2023**, *25*, 8725–8737. [\[CrossRef\]](#)
53. Cao, J.; Pang, J.; Weng, X.; Khirodkar, R.; Kitani, K. Observation-Centric SORT: Rethinking SORT for Robust Multi-Object Tracking. In Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 17–24 June 2023. [\[CrossRef\]](#)
54. Varga, L.A.; Kiefer, B.; Messmer, M.; Zell, 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. [\[CrossRef\]](#)
55. Cafarelli, D.; Ciampi, L.; Vadicamo, L.; Gennaro, C.; Berton, A.; Paterni, M.; Benvenuti, C.; Passera, M.; Falchi, F. MOBDronE: A Drone Video Dataset for Man OverBoard Rescue. In *Image Analysis and Processing—ICIAP 2022, Proceedings of the 21st International Conference, Lecce, Italy, 23–27 May 2022*; Springer: Cham, Switzerland, 2022.
56. Ruiz-Ponce, P.; Ortiz-Perez, D.; Garcia-Rodriguez, J.; Kiefer, B. POSEIDON: A Data Augmentation Tool for Small Object Detection Datasets in Maritime Environments. *Sensors* **2023**, *23*, 3691. [\[CrossRef\]](#)
57. Shao, Z.; Wu, W.; Wang, Z.; Du, W.; Li, C. SeaShips: A Large-Scale Precisely Annotated Dataset for Ship Detection. *IEEE Trans. Multimed.* **2018**, *20*, 2593–2604. [\[CrossRef\]](#)

58. Božić-Štulić, D.; Marusic, Z.; Gotovac, S. Deep Learning Approach in Aerial Imagery for Supporting Land Search and Rescue Missions. *Int. J. Comput. Vis.* **2019**, *127*, 1256–1278. [\[CrossRef\]](#)
59. Gašienica-Józkowy, J.; Knapik, M.; Cyganek, B. An ensemble deep learning method with optimized weights for drone-based water rescue and surveillance. *Integr. Comput.-Aided Eng.* **2021**, *28*, 221–235. [\[CrossRef\]](#)
60. Bovcon, B.; Muhovič, J.; Perš, J.; Kristan, M. The MaSTr1325 dataset for training deep USV obstacle detection models. In Proceedings of the 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Macau, China, 3–8 November 2019. [\[CrossRef\]](#)
61. Bovcon, B.; Muhovič, J.; Vranac, D.; Mozetič, D.; Perš, J.; Kristan, M. MODS—A USV-Oriented Object Detection and Obstacle Segmentation Benchmark. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 13403–13418. [\[CrossRef\]](#)
62. Zhou, Z.; Sun, J.; Yu, J.; Liu, K.; Duan, J.; Chen, L.; Chen, C.P. An Image-Based Benchmark Dataset and a Novel Object Detector for Water Surface Object Detection. *Front. Neurobot.* **2021**, *15*, 723336. [\[CrossRef\]](#)
63. Kiefer, B.; Ott, D.; Zell, A. Leveraging Synthetic Data in Object Detection on Unmanned Aerial Vehicles. In Proceedings of the 2022 26th International Conference on Pattern Recognition (ICPR), Montreal, QC, Canada, 21–25 August 2022; pp. 3564–3571. [\[CrossRef\]](#)
64. Yun, K.; Nguyen, L.; Nguyen, T.; Kim, D.; Eldin, S.; Huyen, A.; Lu, T.; Chow, E. Small target detection for search and rescue operations using distributed deep learning and synthetic data generation. In *Pattern Recognition and Tracking XXX*; SPIE: Bellingham, WA, USA, 2019. [\[CrossRef\]](#)
65. Games, R. Grand Theft Auto V. Available online: <https://www.rockstargames.com/de/gta-v> (accessed on 10 July 2024).
66. Interactive, B. ARMA3. Available online: <https://arma3.com> (accessed on 10 July 2024).
67. Khan Mohammadi, M.; Schneidereit, T.; Mansouri Yarahmadi, A.; Breuß, M. Investigating Training Datasets of Real and Synthetic Images for Outdoor Swimmer Localisation with YOLO. *AI* **2024**, *5*, 576–593. [\[CrossRef\]](#)
68. Kurowski, M.; Korte, H.; Lampe, B. AGaPaS—a new approach for Search-and-Rescue-Operations at sea. *IFAC Proc. Vol.* **2012**, *45*, 73–78. [\[CrossRef\]](#)
69. Krieg, E.; Weber, D.; Renkewitz, H.; Thomas, S. Baywatch 2.0: Rescuing drowning persons with an underwater robotic lifeguard. In Proceedings of the OCEANS 2022, Hampton Roads, Hampton Roads, VA, USA, 17–20 October 2022. [\[CrossRef\]](#)
70. Feng, Z.; Pan, Z.; Chen, W.; Liu, Y.; Leng, J. USV Application Scenario Expansion Based on Motion Control, Path Following and Velocity Planning. *Machines* **2022**, *10*, 310. [\[CrossRef\]](#)
71. Nunes, D.; Fortuna, J.; Damas, B.; Ventura, R. Real-time Vision Based Obstacle Detection in Maritime Environments. In Proceedings of the 2022 IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC), Santa Maria da Feira, Portugal, 29–30 April 2022. [\[CrossRef\]](#)
72. Wang, M.; Wang, Q.; Wang, Y.; Li, Z.; Shao, S.; Zhou, Y. Fusion Detection Algorithm of Maritime Radar and Electro-Optical Pod for Complex Sea Conditions. In Proceedings of the 2023 5th International Conference on Intelligent Control, Measurement and Signal Processing (ICMSP), Chengdu, China, 19–21 May 2023. [\[CrossRef\]](#)

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.