Review

Multi-Modal Perception and Fusion for Maritime Autonomy: A Survey

Chunliu Wang *, Yutong Xu, Daihui Zhang

Posted Date: 14 October 2025

doi: 10.20944/preprints202510.0977.v1

Keywords: maritime autonomy; autonomous ships; multimodal perception; sensor fusion; vessel detection; obstacle avoidance; trajectory prediction; multimodal maritime data

Article

Multi-Modal Perception and Fusion for Maritime Autonomy: A Survey

Chunliu Wang [1,*], Yutong Xu [2] and Daihui Zhang [3]

[1] College of Artificial Intelligence, Dalian Maritime University, Dalian 116026, China
[2] Ocean Engineering College, Dalian Maritime University, Dalian 116026, China
[3] Navigation College, Dalian Maritime University, Dalian 116026, China
[*] Correspondence: chunliuwang@dlmu.edu.cn

Abstract

With the rapid progress of deep learning and the increasing availability of maritime sensing and communication technologies, autonomous ships are emerging as a pivotal direction for the future of intelligent marine operations. In autonomous maritime systems, sensors are the foundation of environmental perception, and their cooperative performance directly affects the safety and reliability of navigation. This survey focuses on recent advances in multi-modal perception and fusion strategies for maritime autonomy. The sensing modalities considered include radar, EO/IR cameras, LiDAR (for near-field perception), sonar, INS, AIS, and satellite-based observations. We analyze the strengths and limitations of these sensors and highlight the necessity of multi-modal fusion under complex maritime conditions, such as adverse weather, dynamic sea states, and non-cooperative targets. Based on recent studies, fusion approaches are categorized into early fusion, feature-level fusion and decision-level fusion, with applications in tasks such as vessel detection, obstacle avoidance, and trajectory prediction. Finally, we discuss the current limitations of multimodal fusion in maritime autonomy—such as asynchronous data streams, sensor misalignment, and limited public datasets—and suggest future research directions toward robust, scalable, and real-time fusion frameworks for autonomous ship operations.

Keywords: maritime autonomy; autonomous ships; multimodal perception; sensor fusion; vessel detection; obstacle avoidance; trajectory prediction; multimodal maritime data

1. Introduction

Maritime operations are fundamental to global sectors such as international trade, energy logistics, fisheries, and environmental monitoring. These operations increasingly take place in complex and dynamic maritime environments, where autonomous systems are being explored to enhance situational awareness, navigational safety, and operational efficiency. The development of autonomous maritime platforms has been supported by advances in sensor technologies and data availability, enabling continuous monitoring and decision support across diverse operational scenarios.

Maritime sensing modalities include the Automatic Identification System (AIS), marine radar, electro-optical/infrared (EO/IR) cameras, and satellite imagery. Each modality provides complementary information, but also exhibits inherent limitations: AIS may be subject to coverage gaps and spoofing, EO/IR sensors can be affected by low-light or adverse weather, and Synthetic Aperture Radar (SAR) requires careful interpretation for reliable target detection. Reliance on a single sensor modality may therefore limit perception accuracy and autonomy in dynamic maritime domains.

Multimodal data fusion has emerged as a key approach to address these challenges. By integrating heterogeneous sources, fusion techniques can improve vessel detection, tracking, obstacle recognition, and trajectory prediction, while enhancing robustness to sensor failures or environmental conditions. Fusion strategies span early (raw data), intermediate (feature-level), and late (decision-level) integration, implemented using classical statistical methods or modern deep learning frameworks, including convolutional, recurrent, and transformer-based architectures.

This survey provides a structured review of multimodal perception and fusion for maritime autonomy. Its contributions include: (1) an overview of maritime sensing modalities and their operational characteristics; (2) analysis of challenges in multimodal integration, such as spatiotemporal misalignment and data heterogeneity; (3) a systematic review of fusion methodologies across different integration stages; (4) a summary of publicly available maritime datasets and evaluation practices; and (5) identification of open research questions and future directions, including self-supervised learning, cross-modal alignment, and real-time deployment. The review aims to support the design and evaluation of scalable, reliable, and generalizable multimodal perception systems for autonomous maritime platforms.

1.1. Related Literature Reviews

While existing literature has examined various aspects of maritime perception and autonomous operations, comprehensive reviews specifically focusing on multimodal perception remain limited. Some reviews explore single-sensor techniques. For example, the AIS-based review [1] provides a detailed analysis of ship tracking, traffic forecasting, and maritime monitoring using AIS data. Similarly, reviews on radar, electro-optical/infrared imagery, or satellite-based maritime observations highlight the applications and limitations of specific sensors, but rarely consider multimodal fusion [2–4]. More comprehensive reviews in the field of artificial intelligence and multimodal learning [5–7] cover alignment, collaborative learning, and fusion strategies; however, they are domain-agnostic and do not consider maritime-specific limitations such as asynchronous sensor streams, inclement weather, ship motion, or limited onboard computing resources. Broader maritime AI research [8,9] focuses on fleet optimization, autonomous navigation, or risk management. These observations highlight a gap in the literature: a lack of analytical research on multimodal fusion strategies specifically for maritime autonomy. This review addresses this gap by examining techniques for fusing data from AIS, radar, EO/IR cameras, and SAR sensors, focusing on applications such as ship detection, multi-target tracking, navigation, and anomaly detection. Key challenges, including spatiotemporal misalignment, sensor drift, and environmental variations, are discussed in the hope of providing insights for robust and scalable multimodal perception for future autonomous maritime systems.

1.2. Article's Organization

The rest of this article is structured as follows: Section 2 provides an overview of the primary sensing modalities used in maritime perception systems, discussing their acquisition platforms, data characteristics, and integration challenges. Multimodal fusion strategies are presented in Section 3, including both traditional (early, mid, late fusion) and deep learning–based approaches, with a focus on cross-modal representation and decision-level reasoning. Recent applications of these techniques to maritime tasks—such as object detection, tracking, behavior understanding, and anomaly detection—are reviewed in Section 4. Section 5 introduces representative multimodal maritime datasets, detailing sensor combinations, spatial-temporal coverage, and accessibility. A discussion of current challenges and future research directions is provided in Section 6, particularly concerning system robustness, scalability, and deployment under real-world conditions. The survey is concluded in Section 7.

2. Maritime Sensors: Types and Characteristics

Accurate perception of the maritime environment is essential for situational awareness and autonomous decision-making in unmanned surface vessels (USVs) and other autonomous maritime platforms. Reliable operation in complex and dynamic sea conditions requires the integration of multiple sensing technologies, each providing complementary information. Understanding the capabilities, limitations, and operational characteristics of these sensors is critical for designing effective multimodal perception and fusion strategies.

Key maritime sensors include radar, EO/IR cameras, LiDAR, sonar, AIS, and satellite-based observations. Radar ensures robust detection in adverse weather conditions such as fog or heavy rain, whereas EO/IR cameras offer high-resolution visual imaging but are sensitive to low-light and occlusions. AIS provides positional and navigational information for vessel tracking, though it depends on signal reception and voluntary transmission. LiDAR and sonar deliver precise range measurements and underwater detection, while satellites enable wide-area surveillance and environmental monitoring.

This chapter presents a comprehensive overview of these sensors, highlighting their operational roles, advantages, limitations, and applicability in autonomous maritime systems. Table 1 summarizes key characteristics and detection ranges across modalities. Subsequent sections examine each sensor type in detail, emphasizing how multimodal fusion enhances perception, supports vessel detection, tracking, obstacle avoidance, and trajectory prediction, and mitigates challenges arising from dynamic and uncertain maritime environments.

Table 1. A comparison of characteristics of maritime sensors.

Sensor	Data Type	Advantages	Limitations
AIS	Tabular -ID, -speed -position, -course	1) Global coverage 2) Standardized format 3) Supports trajectory prediction	1) Low update rate 2) Voluntary transmission 3) Vulnerable to spoofing
Radar	Time-Series Radar plots	1) Real-time detection 2) Works in adverse weather 3) Suitable for tracking	1) Low spatial resolution 2) Sea clutter 3)Limited for small targets
SAR	Radar imagery -grayscal -intensity	1) All-weather 2) Penetrates clouds/rain 3) Detects surface movement	1) Complex interpretation 2) High processing requirements 3) Limited resolution for small objects
EO	Optical Images -RGB -NIR	1) High-resolution imaging 2) Suitable for detection and recognition 3) Intuitive visual information	1) Sensitive to illumination 2) Poor performance at night/fog 3) Thermal images have limited structural detail
IR	Thermal Images -MWIR -LWIR	1) Operates in low-light/night 2) Robust in adverse weather 3) Useful for anomaly detection	1) Low resolution 2) Limited structural details 3) Sensitive to background temperature variations
Lidar	3D Point Cloud -RGB -NIR	1)High-precision 3D spatial data 2) Useful for obstacle detection 3) Provides range and shape info	1) Sensitive to weather 2) Limited range 3) Needs reflective surfaces
Sonar	Acoustic Images Time-series	1) Detects submerged objects 2) Effective for underwater target detection 3) Provides depth/shape info	1) Low resolution 2) Noise-prone signals 3) Limited to underwater targets

2.1. *Automatic Identification System*

The Automatic Identification System is a fundamental maritime communication system that enables real-time vessel tracking and situational awareness [10]. Vessels equipped with AIS transponders periodically broadcast standardized navigational messages that include the Maritime Mobile Service Identity (MMSI), geographic position, speed over ground (SOG), course over ground (COG), heading, vessel type, and navigational status. These broadcasts are received by coastal base stations, satellites, or nearby vessels, facilitating real-time maritime situational awareness and traffic coordination [11].

AIS data has become a fundamental resource for maritime monitoring due to its structured format, global coverage, and continuous availability. It is widely used for vessel tracking, behavior analysis, anomaly detection, and trajectory forecasting. In recent years, data-driven approaches—particularly deep learning models—have leveraged historical AIS trajectories for tasks such as route prediction, intent recognition, and risk assessment [1].

Despite its widespread utility, AIS presents several inherent limitations. These include susceptibility to spoofing, message loss or delay, reduced update frequency in satellite-based AIS (S-AIS), and the inability to track non-cooperative targets (e.g., vessels operating with AIS turned off or transmitting falsified information). Moreover, AIS provides only positional and kinematic data, lacking environmental or visual context critical for close-range perception and obstacle avoidance [12]. As a result, AIS is increasingly integrated with complementary sensing modalities—such as marine radar, EO imagery, and environmental datasets—to enhance robustness, situational awareness, and prediction accuracy in complex maritime environments [13–15].

2.2. Radar

Marine radar systems are widely deployed in maritime navigation and surveillance due to their ability to operate reliably under various environmental conditions [16]. By transmitting microwave pulses and receiving echoes from surrounding objects, radar provides real-time range, bearing, and relative motion information, making it particularly effective in low-visibility scenarios such as fog, rain, and nighttime operations [17,18]. Unlike optical sensors, radar does not rely on ambient light and can detect both cooperative and non-cooperative targets, including those not transmitting AIS signals. Radar outputs are typically represented as grayscale polar images with high temporal resolution but limited semantic information, which increases the difficulty of object classification and interpretation.

Despite its advantages, radar data remains underexploited in learning-based maritime perception. Most deep learning models are optimized for RGB imagery and perform poorly on radar data due to its sparse texture, low signal-to-noise ratio, and unique spatial encoding [19]. Furthermore, the limited availability of annotated, high-quality radar datasets from real maritime scenarios has hindered model training and benchmarking [20].

Recent research efforts have explored radar-specific representation learning, sensor fusion techniques combining radar with EO or AIS data [21,22], and the use of synthetic radar datasets generated through simulation [23]. Emerging deep learning architectures designed for radar characteristics—such as polar-to-Cartesian transformation and Doppler-aware processing—demonstrate potential in improving detection and tracking performance [24,25]. With increasing demand for robust all-weather maritime autonomy, radar-based perception is expected to become a critical complement to vision-based systems, particularly for close-range monitoring, obstacle avoidance, and navigation in challenging environments.

2.3. Electro-Optical Sensor

Electro-optical sensors—including visible-light cameras and satellite-based optical imaging systems—are widely used in maritime perception due to their ability to capture high-resolution, semantically rich visual data of the sea surface. These sensors provide detailed spatial context that supports tasks such as ship detection, classification, tracking, and visual verification, both in onboard systems and shore-based monitoring centers [26–28].

Compared to radar systems, EO sensors offer significantly higher spatial resolution and visual interpretability. Their compatibility with modern computer vision techniques, especially convolutional neural networks (CNNs), has led to their widespread adoption in maritime object detection and scene understanding [3]. Typical EO-based object detection workflows often involve preprocessing steps such as horizon line estimation, background subtraction, and motion-based foreground segmentation [28]. However, these traditional techniques are challenged by dynamic ocean environments, such as moving waves, changing illumination, and coastal clutter [29].

Recent advancements in deep learning have improved the robustness of EO-based perception, enabling end-to-end learning-based detection and classification under diverse sea conditions. Nevertheless, EO sensors face intrinsic limitations. Their performance deteriorates in adverse environments, such as nighttime, heavy fog, or rain, due to their reliance on ambient lighting and line-of-sight visibility [30–32]. In addition, visual quality may vary with sensor viewing angles, background complexity, and occlusion, all of which can negatively impact detection accuracy [33,34]. To overcome these challenges, EO sensors are often integrated with complementary modalities such as radar, infrared (IR), and AIS to enhance perception robustness under varying operational conditions [35–38].

2.4. Infrared (IR) Sensor

Infrared (IR) and thermal imaging sensors capture emitted or reflected infrared radiation, enabling target detection under low-light or obscured conditions. Unlike electro-optical (EO) sensors, IR sensors do not depend on ambient illumination, and thus are effective in nighttime, fog, smoke, and haze [39]. Based on spectral range, IR sensors are commonly categorized into near-infrared (NIR, 0.75–1.4 µm), shortwave infrared (SWIR, 1.4–3 μm), midwave infrared (MWIR, 3–5 μm), and longwave infrared (LWIR, 8–14 μm) [40,41]. LWIR sensors, in particular, are widely used for passive thermal imaging, detecting heat signatures from vessels, people, and other thermal sources.

IR systems are typically deployed on maritime platforms such as unmanned aerial vehicles (UAVs), patrol vessels, and coastal surveillance installations [42]. In practical applications, IR sensors are often fused with EO or radar data to enhance robustness across variable weather and lighting conditions [39]. Despite their advantages, IR sensors present several limitations. Thermal imagery generally has lower spatial resolution than EO imagery and is sensitive to background thermal noise, sea surface temperature gradients, and atmospheric absorption. These factors can degrade detection accuracy, particularly in cluttered or high-reflectance coastal environments [43].

Furthermore, the development of IR-based perception systems is hindered by the scarcity of publicly available, labeled IR datasets. Existing datasets are often limited in scale or scenario diversity, restricting the training and benchmarking of deep learning models. To address this, recent research has explored multi-band IR fusion, cross-modal learning strategies, and domain adaptation techniques to improve generalization and feature alignment with EO or radar modalities [44,45].

2.5. Satellite-Based Remote Sensing System

Satellite-based sensors provide wide-area maritime observation through various imaging modalities, including synthetic aperture radar (SAR), multispectral, hyperspectral, and optical sensors [46–48]. These sensors are deployed on Earth observation platforms such as Sentinel-1/2 (ESA), Landsat (NASA/USGS), and commercial constellations (e.g., PlanetScope, Maxar). Among them, SAR enables all-weather, day-and-night imaging of ocean surfaces, while optical and multispectral sensors support high-resolution detection of vessels, oil spills, sea ice, and other surface anomalies [49].

Satellite imagery plays a critical role in maritime domain awareness, particularly for large-scale monitoring, historical trajectory validation, and surveillance over remote or high-risk regions [50–52]. When fused with AIS data, satellite observations can support the detection of non-cooperative targets (e.g., dark vessels) and the validation of reported navigation paths. However, satellite-based sensing faces several operational constraints [2]. These include limited temporal resolution (e.g., revisit intervals from several hours to days), data latency, and weather-dependent quality degradation—particularly due to cloud cover in optical imagery. In addition, the spatial resolution of freely available satellite data is often insufficient for small-object detection or real-time situational awareness [53].

To address these limitations, recent studies have investigated multi-modal fusion strategies combining satellite data with AIS and radar inputs. Deep learning methods—such as attention-based fusion, anomaly detection networks, and pretraining with remote sensing imagery—have been applied to improve ship detection, trajectory forecasting, and event classification in maritime scenarios [54,55].

2.6. Sonar

Sonar sensors are critical for underwater monitoring in maritime domains. They use sound waves to detect objects, measure distances, and map the sea floor [56]. This makes them ideal for detecting submerged obstacles and underwater vessels, particularly because light is rapidly absorbed in water, limiting the effectiveness of optical cameras and LiDARs for underwater mapping. In contrast, sonar systems can achieve longer sensing ranges (up to several kilometers) and do not depend on lighting conditions. Additionally, navigation sonar systems can generate 3D surface models of the seafloor or underwater terrain, which are crucial for safe navigation and maritime mapping.

The data captured by sonar devices is typically represented as point clouds or depth profiles, providing critical insights into the maritime environment. This data is especially useful for activities like underwater exploration, port monitoring, and search-and-rescue operations. In maritime surveillance and navigation, Sonar data is often fused with other sensor modalities, such as radar, AIS, or EO sensors, to improve situational awareness and enhance decision-making in dynamic, cluttered maritime

doi:10.20944/preprints202510.0977.v1

environments [57,58]. For instance, when combined with radar data, sonar can help identify and avoid submerged objects that may be missed by radar alone.

However, sonar's primary limitation lies in its inability to detect objects above or on the water's surface, making it unsuitable for tracking surface vessels or aerial targets. Additionally, sonar performance is sensitive to environmental factors, such as water clarity, salinity, and temperature, which can affect the accuracy of depth measurements and object detection. Recent developments in multi-sensor fusion and machine learning algorithms are pushing the boundaries of sonar applications, enabling more accurate underwater mapping, anomaly detection, and real-time monitoring in maritime environments [59,60].

2.7. Static and Non-Sensor Data for Autonomous Maritime Systems

In addition to real-time sensor modalities such as AIS, radar, and EO/IR, autonomous maritime systems can benefit from the integration of static or non-sensor datasets that provide critical environmental and contextual priors. Representative examples include electronic navigational charts (ENCs), nautical charts, Geographic Information System (GIS)-based layers, bathymetric maps, maritime boundaries, and port infrastructure databases. These datasets are generally issued by hydrographic authorities or maritime organizations in standardized formats (e.g., S-57, S-100) and offer structured geospatial information such as seabed topography, shipping lanes, restricted areas, and hazard locations.

Although they lack temporal dynamics, these non-sensor data sources play a complementary role by anchoring sensor observations in a reliable geospatial frame of reference. When fused with dynamic sensor inputs, they enable enhanced tasks such as route planning, contextual behavior analysis, anomaly detection, grounding risk assessment, and compliance verification with maritime regulations. For instance, deviations from recommended shipping lanes can be more easily identified when AIS trajectories are overlaid with ENC data, while radar or EO detections can be cross-validated against known navigational hazards.

Nevertheless, the utility of static datasets depends on their accuracy, georeferencing quality, and update frequency. Outdated or misaligned charts can introduce systematic errors, potentially degrading autonomous navigation performance. Thus, effective maritime autonomy requires not only the fusion of multimodal dynamic sensors but also the incorporation of curated static geospatial data as a knowledge base for robust perception and decision-making.

3. Multimodal Fusion in Maritime Autonomy

Maritime perception systems rely on diverse sensor modalities to capture the dynamic and often adverse conditions at sea. Data from Automatic Identification System (AIS) messages, marine radar, electro-optical (EO) and infrared (IR) imagery, synthetic aperture radar (SAR), and nautical charts each provide complementary but partial information about the maritime environment. Multimodal learning integrates these heterogeneous sources, enabling more robust and context-aware perception, particularly under occlusion, noise, or missing data.

Multimodal deep learning (MMDL) methods aim to jointly model multiple data sources using deep neural networks, automatically extracting hierarchical features and learning cross-modal relationships in an end-to-end manner. Compared to traditional sensor fusion approaches such as Kalman filtering or rule-based decision fusion, MMDL can handle complex perception tasks like vessel detection, trajectory prediction, anomaly recognition, and maritime scene understanding more effectively.

MMDL methods are commonly classified according to their fusion strategies, which determine how information from multiple modalities is combined:

- **Early Fusion (Data-level)**: Raw data from different modalities are combined before feature extraction.
- **Middle Fusion (Feature-level)**: Features are first extracted independently from each modality, then fused.

- **Late Fusion (Decision-level)**: Each modality is processed separately, and the final decisions are combined.

Each fusion strategy exhibits distinct strengths and limitations. Early fusion can exploit cross-modal correlations effectively but may be sensitive to misaligned or noisy data. Intermediate fusion balances modality-specific processing with cross-modal interactions, while late fusion offers flexibility and robustness at the potential cost of losing fine-grained cross-modal relationships. Adapting these strategies to maritime scenarios often requires handling asynchronous sensor data, intermittent measurements, and challenging environmental conditions such as poor visibility or adverse weather.

This section focuses on the role of fusion strategies in maritime MMDL and the corresponding network architectures, providing a structured overview of how different approaches have been applied to this domain.

3.1. Fusion-Level-Based Methods

MMDL technology fuses data from different sensors to learn more representative features from multi-level representations, effectively reducing data uncertainty and enhancing the performance of downstream tasks such as ship detection, trajectory prediction, and anomaly recognition. In maritime environments, object detection becomes particularly challenging due to the varying lighting, weather conditions, and the dynamic nature of the sea, which can cause false detections. To address these challenges, MMDL provides a robust approach by combining complementary sensor data, such as Electro-Optical (EO) imagery and Infrared (IR) images, to improve detection accuracy.

Based on the hierarchical and temporal characteristics of the fusion, multimodal fusion methods are generally classified into three main types: early fusion, intermediate fusion, and late fusion. This classification approach is commonly found in previous multimodal machine learning review papers. Early fusion involves combining raw data from multiple sensors before any information extraction is performed, while intermediate fusion merges features extracted from each modality's raw data, allowing for more nuanced learning of complex relationships. Late fusion aggregates the outputs of independent detectors for final decision-making. In maritime applications, such as maritime vessel detection tasks, reference [61] implemented three CNN architectures for RGB+IR fusion: early fusion (stacking raw images), feature fusion (merging CNN feature maps), and late fusion (combining separate detector outputs).

3.1.1. Early Fusion

Data-level fusion, or early fusion, combines raw sensor data or pre-processed outputs from different modalities at the input stage. The main advantage of early fusion is that it preserves low-level correlations between modalities, reducing information loss during multimodal processing. This approach maximizes the complementary nature of sensor data to enhance system robustness and accuracy. Which is abundantly adopted in multi-source image fusion for image enhancement, especially in the application of remote sensing imaging by fusing infrared images and RGB images. Early fusion is effective when data from different modalities are well aligned in space and time or can be combined without extensive processing. For example, visual and infrared data can often be aligned at the pixel level, making them suitable for fusion. Similarly, radar and AIS (Automatic Identification System) data can be fused early when they are temporally or spatially synchronized. In maritime applications, early fusion is commonly used to combine data from different sensors, such as electro-optical cameras, infrared cameras, and radar, for tasks like ship detection, trajectory prediction, and anomaly detection. For example, combining electro-optical (EO) and infrared (IR) images improves ship recognition accuracy, especially in challenging conditions such as low light, haze, or night.

Recent maritime studies have demonstrated the effectiveness of data-level fusion in improving perception tasks under adverse conditions. For example, [61] proposed a CNN-based early fusion network that concatenates RGB and IR images into a four-channel tensor, significantly enhancing ship detection under occlusion and dim lighting conditions. [62,63] utilized both intrinsic camera parameters and extrinsic LiDAR-camera calibration for data integration, developing an image-based segmentation method for bollard detection and autonomous mooring applications. In a similar vein, [64] researched coastal mapping by fusing LiDAR data with multi-spectral or hyperspectral images, further enhancing environmental perception. These studies collectively highlight the effectiveness of raw-data fusion in improving object detection, pose estimation, and overall perception accuracy for complex maritime operations.

Early fusion can exploit low-level inter-modal correlations, reduce computational redundancy, and enable end-to-end learning. However,In maritime scenarios, sensor data often comes from disparate sources (such as SAR, optical, AIS, and environmental data), with significant inconsistencies in resolution, timescale, and sampling frequency. Early attempts to directly stitch or overlay data layers can easily lead to feature mismatches (for example, AIS images are captured at minute intervals, while SAR images may be captured only once every few days), introducing significant noise and bias. Furthermore, early fusion requires high data quality. If a modality is missing or interfered with (for example, if AIS is turned off or optical imagery is affected by weather), overall system performance can be significantly degraded, resulting in insufficient robustness.

3.1.2. Middle Fusion

Middle fusion (also known as feature-level fusion) is considered one of the most effective strategies among multimodal fusion approaches [65]. By operating at the feature level, this method enables deeper integration across modalities and captures both semantic abstractions and spatial details from heterogeneous data sources. In the maritime domain, where combining complementary information from different sensors is crucial, feature-level fusion has been widely adopted to enhance perception and decision-making.

Several representative studies illustrate its effectiveness. For instance, Achiri et al.[66] fused ship position, heading, and size features extracted separately from SAR images and AIS data, integrating them through an arithmetic mean function to enable more robust vessel identification. Liu et al. [67] combined optical and SAR data, first generating port slice images through data-level fusion and then employing feature-level fusion with saliency analysis, region growing, and joint shape-based classification to improve ship detection and recognition. Luo et al. [68] designed a multimodal deep learning trajectory prediction framework (MDL-TP) that integrates AIS-based behavioral features with environmental variables such as wind, visibility, and temperature, using modality-specific neural networks and feature-level fusion to jointly capture spatiotemporal vessel dynamics and environmental influences. These approaches demonstrate the potential of intermediate fusion to improve accuracy and robustness in complex maritime applications, even under noisy, incomplete, or asynchronous observations.

However, despite its advantages, the application of intermediate fusion in the maritime domain still faces several challenges. While mid-term fusion can partially alleviate the data layer disparity issue, it still presents challenges in feature alignment and semantic differences in maritime scenarios. For example, SAR imagery extracts geometric and scattering features, while AIS provides behavioral trajectories and navigation status. These features have significantly different semantic levels, making efficient feature complementation difficult. Furthermore, mid-term fusion often requires complex network architectures and high computational overhead, making it unsuitable for real-time monitoring and large-scale maritime applications. In addition, in the case of missing or low-quality modalities, mid-term fusion may still experience "modal bias" (the dominance of a certain modality leads to information imbalance).

3.1.3. Late Fusion

Decision-level fusion is a technique that improves accuracy in various applications by integrating the outputs from multiple models. This is useful in settings with asynchronous or unreliable modalities. Decision-level fusion has emerged as an effective strategy to enhance accuracy in maritime applications by integrating the outputs of multiple modality-specific models. In complex maritime surveillance scenarios, such as ship detection and classification, this technique can mitigate modality-specific limitations by aggregating predictions from independently trained networks.

For instance, in ship detection tasks, researchers have employed model ensembles that integrate outputs from SAR-based object detectors and EO-based classifiers through weighted voting or confidence-based stacking, significantly improving performance in congested coastal zones and under variable weather conditions. Such approaches leverage the robustness of SAR in low-visibility environments and the fine-grained detail of EO imagery under clear skies. In this context, several decision-level fusion methods have been explored in recent years. In [61], a CNN-based late fusion approach was proposed, where bounding box proposals from EO and IR detectors were combined at the decision level, demonstrating superior vessel detection performance compared with single-modality CNNs. Similarly,[69] investigated the integration of space-borne SAR and AIS for ship surveillance, highlighting improved association strategies and multi-source fusion techniques, and further introducing decision-level methods such as averaging, Dempster–Shafer evidence theory, and fuzzy reasoning. More recently, Chen et al.[70] proposed a decision-level fusion framework for ship detection using optical and SAR images, where Faster R-CNN independently detected targets from each modality, U-Net provided land–sea segmentation, and the final decision-level fusion combined detection outputs with segmentation results to enhance target discrimination.

In maritime applications, late-stage fusion often manifests itself as "independent decision-making for each modality followed by weighting." While simple to implement, it overlooks the underlying complementary information between modalities. For example, SAR can detect "stealth" vessels (without AIS), while AIS provides behavioral characteristics. If voting or weighting is performed solely at the decision-making level, the synergy between the two at the feature level may be missed. Furthermore, late-stage fusion is prone to bias when faced with data imbalance (e.g., AIS information far outweighs SAR imagery), making it difficult to meet the requirements of complex maritime tasks such as anomaly detection and behavior prediction.

3.2. Deep Learning-Based Fusion Techniques

Although fusion strategies are often categorized into early (data-level), intermediate (feature-level), and late (decision-level) integration, this taxonomy mainly reflects the stage at which multimodal information is combined within the processing pipeline. In contrast, deep learning–based approaches represent a classification from the perspective of implementation methods. Neural architectures such as convolutional, recurrent, and transformer-based models can be employed across different fusion stages: for instance, CNNs can perform early fusion by directly combining multi-channel inputs, attention-based models can realize feature-level fusion by aligning latent representations, and ensemble networks can enable decision-level fusion by aggregating multiple modality-specific outputs. Hence, the two perspectives are complementary—fusion levels define where integration occurs, while deep learning techniques define how it is implemented.

3.3. Challenges in Multimodal Data Integration

Integrating multiple maritime sensing modalities, such as EO, SAR, AIS, and radar, can significantly enhance situational awareness and perception accuracy. However, effective multimodal fusion faces several technical challenges that stem from the inherent heterogeneity and operational constraints of maritime data.

Key technical challenges include:

- **Limited multimodal datasets:** High-quality, temporally synchronized datasets combining EO, SAR, AIS, and radar remain scarce or proprietary. The lack of publicly available training and evaluation resources hampers the development and benchmarking of supervised fusion models.
- **Asynchronous and unbalanced modalities:** Maritime sensors often operate at different sampling rates, and data streams may be missing or partially corrupted (e.g., cloud-covered EO imagery,

spoofed or lost AIS signals). Fusion algorithms must be robust to such imbalances and capable of handling intermittent or incomplete inputs.

- **Temporal and spatial misalignment:** Differences in frame rates, latencies, and spatial resolutions across sensors introduce alignment challenges. Accurate temporal synchronization and geospatial registration are essential for reliable cross-modal feature integration [7,71].
- **Heterogeneous data formats and representations:** Modalities differ in structure (e.g., structured AIS messages vs. raster EO imagery), dimensionality, and semantic content. Learning joint representations that effectively capture complementary information is non-trivial [5,72].
- **Noise, incompleteness, and uncertainty:** Sensor observations are often affected by environmental conditions and operational limitations, such as radar clutter, EO/IR occlusions, or AIS signal loss. Fusion frameworks must be capable of handling noisy and missing data while maintaining predictive reliability [73,74].
- **Computational constraints for real-time operation:** Onboard maritime platforms typically have limited processing resources. Efficient architectures are required to fuse high-volume data streams in real time without compromising accuracy [75].

Addressing these technical challenges necessitates adaptive fusion architectures capable of handling asynchronous inputs, heterogeneous modalities, and cross-domain uncertainties. Recent solutions include attention-based fusion networks, cross-modal alignment strategies, and uncertainty-aware learning frameworks [76–79].

Table 1 summarizes key characteristics of common maritime data types, providing a reference for selecting appropriate fusion strategies. A comprehensive overview of publicly available multimodal maritime datasets, including sensor combinations, coverage regions, and accessibility, is presented in Chapter 5.

4. Applications and Modalities in Maritime Autonomy

Multimodal fusion has become a cornerstone of diverse maritime applications, each shaped by distinct sensing environments, operational demands, and data availability. Recent research highlights how different sensor modalities are selected and combined depending on task requirements: for instance, AIS data are frequently employed in trajectory prediction due to their rich spatiotemporal information; EO/IR imagery and radar signals are central to ship detection and classification; while search and rescue operations often integrate multiple sources—including satellite imagery and environmental sensors—to strengthen situational awareness. This section categorizes the major application domains of maritime autonomy and examines the typical sensing modalities, fusion strategies, and datasets used in each, providing insight into how multimodal approaches are tailored to specific maritime challenges.

To ensure a comprehensive and representative overview of these application domains, we systematically surveyed recent literature using Google Scholar and related databases. A targeted set of keywords was employed, including Autonomous ship, Maritime autonomy, Unmanned Surface Vehicle (USV), Unmanned Surface Vessel, Unmanned Maritime Vehicle (UMV), Maritime Autonomous Surface Ship (MASS), and Autonomous Surface Craft. This keyword-driven search strategy enabled us to capture a broad spectrum of studies covering different facets of maritime autonomy, thereby grounding the applications summarized in this section in a robust body of prior research.

4.1. Vessel Detection and Recognition

Accurate vessel detection, continuous tracking, and reliable ship-type identification are fundamental to modern maritime surveillance, supporting navigational safety, traffic management, and behavioral analytics. Commonly employed sensor modalities include radar (broad-range, all-weather detection), EO/IR imagery (visual confirmation and classification), and AIS, which provides spatiotemporal kinematic information—such as position, speed, heading—and vessel identity [80].

doi:10.20944/preprints202510.0977.v1

While AIS is invaluable for tracking and identification, its limitations—including signal loss, spoofing, noise, and incomplete coverage—can degrade performance in complex maritime environments. To address these challenges, recent research emphasizes multimodal fusion strategies, integrating AIS with radar and EO/IR data to enhance detection range, tracking robustness, and classification accuracy. For example, video imagery can complement AIS by providing vessel appearance and motion context [81], while radar ensures persistent tracking when AIS transmissions are absent or compromised [82].

Non-cooperative observing systems, such as optical and infrared cameras or radar platforms deployed from shore, vessels, aircraft, or satellites, further augment maritime surveillance. Among these, satellite-based sensors are particularly valuable due to their global coverage, regular revisit cycles, and high data acquisition rates. Synthetic Aperture Radar (SAR) enables ship detection under all weather and illumination conditions, whereas optical imagery provides higher spatial detail and visual confirmation, making it complementary to SAR for vessel characterization [4].

Building on these modalities, research on multimodal fusion has advanced along several directions. Deep neural networks with CNN or attention-based architectures have been employed to fuse EO/IR and radar features, improving discrimination of small or occluded vessels [20]. Integration of AIS with X-band SAR detections has been explored to enhance identification of small vessels and "dark ships," as well as to assess the impact of supplementary monitoring data on detection accuracy and false alarm reduction [15]. Furthermore, AIS fusion with satellite remote-sensing imagery, using point-set matching and fuzzy comprehensive decision methods, has been shown to significantly improve maritime target positioning, reducing errors by over 70% and providing precise vessel locations for operational monitoring [54].

Recent advances further extend multimodal fusion to underwater monitoring, where Automatic Identification System signals are combined with sonar data to improve target association and identification. Such approaches address challenges related to environmental noise, incomplete coverage, and temporal or spatial mismatches, demonstrating the potential of heterogeneous sensor fusion to enhance vessel detection and tracking in increasingly complex maritime environments [58].

4.2. Obstacle Detection and Classification

Obstacle detection and classification are critical components of maritime situational awareness, focusing on the identification of non-vessel objects such as floating debris, buoys, icebergs, and reefs. Accurate detection of these obstacles is essential for safe navigation and collision avoidance in both inland waterways and open seas. A variety of sensor modalities and computational methods have been explored for obstacle detection and recognition. Electro-optical and infrared cameras are widely used in conjunction with deep learning-based algorithms to detect and classify floating or surface objects. These systems leverage visual and infrared information to identify obstacles under varying illumination and weather conditions, providing fine-grained recognition of transient or small-scale hazards.

Radar and sonar systems complement optical sensors by enabling the detection of fixed or semi-fixed obstacles, such as buoys, piers, or submerged structures. Radar offers all-weather, long-range coverage, while sonar provides underwater perception, which is critical for submerged hazards that cannot be captured by optical sensors. For more detailed spatial and shape characterization, LiDAR and multibeam sonar systems have been applied to detect and classify both surface and underwater obstacles. LiDAR generates high-resolution three-dimensional point clouds for surface mapping, whereas multibeam sonar produces bathymetric and volumetric representations of underwater objects, facilitating precise localization and classification.

Recent research emphasizes sensor fusion strategies, which combine electro-optical, infrared, radar, sonar, and LiDAR data to exploit complementary strengths, enhance detection robustness, and improve classification accuracy in complex maritime environments. Such multimodal approaches are particularly valuable in scenarios with limited visibility, high traffic density, or heterogeneous obstacle types. For instance, a probabilistic fusion framework integrating LiDAR, radar, and cameras has been proposed to improve obstacle detection and classification under challenging conditions, such as glare or objects that interfere with LiDAR returns [83]. Similarly, autonomous surface vehicles have adopted image-based multisensor fusion approaches, where cameras provide primary semantic information and LiDAR point clouds supply spatial context. In these methods, tracking-assisted image detection leverages historical information to compensate for missed detections, while confidence-association-based fusion strategies are used to determine final targets [84].

Extending multimodal fusion to underwater environments, one study employs a dual-stream interactive network with attention-based feature fusion to combine optical and sonar images for robust underwater object detection [57]. This approach effectively improves detection accuracy compared with methods that rely on a single sensor and enhances visualization and spatial understanding of underwater environments, providing critical support for unmanned vehicle operations.

4.3. Scene Understanding

Scene understanding is a critical component of autonomous maritime and underwater navigation, encompassing the perception, reconstruction, and interpretation of the surrounding environment to support safe and efficient operations. In the context of underwater vehicles, accurate three-dimensional mapping and path planning are essential for autonomous navigation, obstacle avoidance, and mission execution.

Recent studies have focused on enhancing scene understanding through multi-sensor fusion. One line of research emphasizes multi-sonar fusion for three-dimensional reconstruction and environmental modeling. Forward-looking sonar, which is robust in turbid environments, can be complemented by profiling sonar to produce fused point clouds that mitigate sensor-specific ambiguities and noise, thereby improving mapping fidelity and supporting adaptive scan-path planning for autonomous underwater vehicles [60]. Simulation and experimental results indicate that such multi-sonar approaches yield higher-fidelity reconstructions than single-sensor solutions.

Beyond sonar-based reconstruction, other work explores perceptual enhancement through cross-modal fusion. For instance, a dual-stream interactive network with attention-based feature fusion has been proposed to combine optical and sonar images for robust underwater object detection, significantly improving detection accuracy compared with unimodal methods [59]. Similarly, in maritime surveillance scenarios, infrared and visible imagery from shipborne electro-optical pods can be fused to overcome challenges of illumination and weather, producing images that preserve critical targets while retaining the natural appearance of visible imagery [36]. These methods demonstrate that perceptual fusion across different modalities—whether optical, sonar, or infrared—can enhance both machine interpretation and human decision-making in complex maritime environments.

Together, advances in sonar-based reconstruction and multimodal perceptual fusion underscore that scene understanding in maritime contexts requires both accurate spatial models and enhanced multimodal perception. Integrating geometric and semantic information from heterogeneous sensors offers a promising path toward resilient, semantically rich environmental representations that support navigation and decision-making in complex marine environments.

4.4. Anomaly Detection and Behavior Recognition

Anomaly detection is a key component of modern maritime surveillance systems, enabling automated identification of unusual vessel behaviors or environmental conditions. Accurate detection of anomalies enhances situational awareness and supports the overall safety of maritime operations [85].

Maritime anomalies can be broadly categorized into position, environmental, and motion anomalies. Position anomalies arise when vessels deviate from expected or designated routes, such as cargo ships leaving prescribed lanes or ferries following atypical trajectories. Environmental anomalies are deviations influenced by external factors, including seasonal variations, day-of-week effects, or vessel-type-specific operational patterns. Motion anomalies encompass unusual behaviors in speed, heading, or maneuvering, such as navigating in the opposite direction, traveling at unusually high or low speeds, abrupt turns, or instantaneous stops. Collectively, these categories illustrate the diversity of anomalous vessel behaviors and underscore the need for robust detection methods capable of capturing deviations across multiple dimensions [86].

Detecting maritime anomalies is particularly challenging due to the heterogeneous and dynamic nature of sensor data. Sources such as Automatic Identification System (AIS) tracks, radar returns, electro-optical/infrared imagery, and sonar measurements are often noisy, incomplete, or temporally misaligned, necessitating sophisticated data processing and fusion strategies [87]. To address these challenges, a variety of multimodal fusion approaches have been proposed to leverage the complementary strengths of different sensor modalities.

Weighted track fusion algorithms that integrate AIS and X-band radar data have been developed to enable accurate ship monitoring even in complex maritime traffic [13]. In these approaches, AIS provides primary positional information, while radar ensures continuous tracking when AIS signals are missing or unreliable. By combining local information entropy with weighted fusion, such methods can effectively identify abnormal vessel behaviors, including deviations in speed, course, position, spacing, and timing.

Beyond AIS and radar, multi-source frameworks have been explored for small vessel detection and anomaly recognition. For example, SVIADF combines synthetic aperture radar, AIS tracks, and wide-area optical imagery to detect and classify small or low-visibility vessels [88]. In the detection stage, YOLO-based object proposals are further refined using constant false alarm rate analysis, which adaptively adjusts detection thresholds according to local background noise, thereby reducing false positives and improving recall.

Recent work has also demonstrated the potential of integrating GPS trajectory data with visual information for anomaly detection in Unmanned Surface Vehicles [89]. The Dual-Branch and Dual-View Multimodal Learning framework employs a visual dual-branch module to extract holistic and localized semantics and a multimodal semantic fusion module to align and enhance cross-modal representations. Experiments on real-world USV datasets indicate that this approach significantly outperforms unimodal baselines, emphasizing the importance of multimodal fusion in capturing spatial-temporal interdependencies and improving anomaly detection in complex maritime environments.

4.5. Trajectory Prediction

Trajectory prediction of maritime vessels constitutes a fundamental task in intelligent maritime traffic management, underpinning applications such as collision avoidance, traffic flow regulation, and navigational behavior analysis. The advent of the Automatic Identification System (AIS) has facilitated large-scale collection of time-series kinematic data, thereby enabling the development of data-driven predictive models. Nevertheless, models trained solely on AIS data often face challenges in highly dynamic maritime environments, where factors such as wind, ocean currents, and multi-vessel interactions introduce substantial uncertainty [90].

Despite their ubiquity, raw AIS data streams suffer from several inherent limitations, including incomplete coverage, transmission latency, measurement noise, and heterogeneous data standards. These deficiencies may compromise the accuracy and robustness of trajectory forecasting. To address these issues, recent research has shifted toward multimodal integration strategies that combine AIS with auxiliary information sources such as remote sensing imagery, nautical charts, and environmental observations. By providing enriched contextual information, multimodal fusion frameworks enhance intent inference and improve the reliability of predictive systems.

A key line of work focuses on incorporating exogenous environmental influences into trajectory prediction. Multimodal frameworks that jointly analyze AIS kinematics and environmental factors (e.g., wind, visibility, temperature) have been proposed. For instance, the Multimodal Deep Learning Trajectory Prediction (MDL-TP) framework integrates meteorological and oceanographic data with AIS streams [91], while the Adaptive Multimodal Prediction (AMP) model employs parallel feature extraction networks and a gated fusion mechanism to enhance robustness under diverse operating conditions [92]. Both approaches demonstrate that leveraging environmental priors significantly reduces predictive uncertainty and improves reliability in practical deployments.

Beyond environmental variables, multimodal research has also explored the fusion of dynamic sensing modalities such as radar with static geospatial priors like nautical charts [93]. A representative example is the Trajectory Prediction Network, which formulates vessel trajectory forecasting as a multi-task sequence-to-sequence problem by integrating AIS/GPS trajectories with radar imagery and water–land segmentation derived from navigational charts. To support this line of investigation, the Inland Shipping Dataset (ISD) was released, offering aligned AIS, radar, and chart data for benchmarking. Experimental evaluations on ISD reveal that multimodal fusion substantially outperforms single-modality baselines, thereby highlighting the effectiveness of integrating vessel kinematics, perception data, and static priors for accurate trajectory forecasting.

4.6. Collision Avoidance & Risk Assessment

Collision avoidance and risk assessment are critical high-level decision-making tasks in intelligent maritime navigation, directly influencing vessel safety and operational efficiency. While conceptually distinct from obstacle detection and trajectory prediction, all three tasks often rely on similar sensor modalities such as AIS, radar, optical/infrared cameras, and environmental measurements. Obstacle detection operates at the perception level, focusing on the real-time identification and localization of surrounding objects ("what" is in the environment). Trajectory prediction addresses the forecasting layer, integrating temporal sequences of kinematic and environmental data to anticipate future vessel positions and intent ("where" vessels are likely to move). Building upon these inputs, collision avoidance and risk assessment function at the decision-making layer, combining situational awareness and predicted trajectories to evaluate encounter risks and generate safe navigational actions ("how" to act). This hierarchical distinction underscores the complementary roles of multimodal inputs across perception, forecasting, and decision-making in intelligent maritime navigation systems.

The dynamic nature of maritime traffic and environmental conditions introduces substantial uncertainty, arising from vessel maneuvering, weather disturbances, and interactions among heterogeneous traffic participants. Such uncertainty can limit the effectiveness of approaches relying on single-source or simplistic information, highlighting the need for integrated multimodal strategies. At the perception and forecasting levels, obstacle detection and trajectory prediction exploit sensor-driven multimodal fusion, combining AIS, radar, and optical/infrared data with environmental measurements. Obstacle detection focuses on accurately identifying and localizing surrounding vessels and obstacles in real time, while trajectory prediction leverages temporal sequences of kinematic and environmental data to forecast future positions and intent. These fused inputs provide the situational awareness and predicted motion information necessary for informed decision-making at higher levels.

Building on these inputs, collision avoidance and risk assessment employ information-driven multimodal fusion [94–96], integrating high-level data such as AIS-derived traffic flow, weather and environmental conditions, historical navigation accidents, and vessel parameters. Multi-modal deep fusion techniques have been shown to enhance the accuracy and robustness of risk assessment under complex maritime conditions, enabling more effective decision-making for ship operators [97,98]. Extending this approach, several studies have combined additional data sources—including ship traffic flow, weather, historical accidents, and vessel parameters—to develop ship navigation risk early warning models. By incorporating multi-modal learning with transfer learning, these models generalize across different sea areas, maintaining high predictive accuracy and supporting comprehensive maritime safety management [99].

Moreover, near-miss incidents have been incorporated into multi-modal deep fusion frameworks to improve early risk detection. For example, Wu et al. [100] proposed a near-miss risk identification method that integrates AIS, weather information, historical navigation accidents, and vessel parameters. This approach constructs a near-miss probability model using kernel density estimation and employs machine learning algorithms to analyze potential navigation risks from high-dimensional, multi-source data. The framework not only effectively identifies navigation risks but also predicts potentially hazardous situations earlier, capturing operational and environmental factors that, while not immediately dangerous, indicate elevated risk levels.

4.7. Search and Rescue (SAR) Operations

Maritime search and rescue (MSAR) operations demand rapid, reliable, and weather-resilient detection of distressed individuals, life-saving equipment, or debris under challenging visibility and environmental conditions. To address these challenges, multimodal perception and learning have become key enablers, integrating data from electro-optical and infrared imagery, radar, and distress beacons. The fusion of these complementary sources provides spatial and semantic diversity, which is crucial for robust detection and localization in complex maritime environments.

SAR missions differ from general maritime environmental perception or obstacle detection tasks in sensor configuration and operational focus. While both rely on radar, EO/IR cameras, and LiDAR, SAR prioritizes the rapid detection of small, fragile targets, such as individuals or life rafts, often emphasizing thermal imaging fused with high-resolution EO sensors and distress beacons. LiDAR provides complementary geometric cues, reducing false positives. In contrast, environmental perception and obstacle detection focus on global situational awareness and collision avoidance, typically relying on radar and AIS for large targets and EO/IR or LiDAR for close-range identification. This distinction highlights that while SAR and general perception share sensors, their priorities and fusion strategies differ.

Conceptually, MSAR can be divided into two complementary stages: maritime search and maritime rescue. The search phase is particularly critical, as it focuses on the detection and localization of distressed individuals, life-saving equipment, or debris [101]. Given challenges such as small target size, weak signals, and adverse environmental conditions, this phase naturally relies on the fusion of multiple sensing modalities. Thermal signatures enable weak-target detection, LiDAR provides geometric cues, radar offers long-range coverage, and EPIRB signals support detection of cooperative targets. Collectively, these modalities enhance the robustness of perception under a variety of environmental conditions.

In contrast, the rescue phase emphasizes intervention and execution, including safe approach, payload delivery, and retrieval of persons or objects. While it depends on the results of the search phase, rescue operations also incorporate real-time environmental perception, such as obstacle detection and navigation, to ensure operational safety and efficiency. Consequently, the search stage can be regarded primarily as a multi-sensor fusion perception task, whereas the rescue stage focuses mainly on planning and control, with sensory inputs serving to support safe execution.

Research on unmanned maritime MSAR systems remains in its early stages, primarily involving four major platforms: unmanned aerial vehicles (UAVs), unmanned surface vessels (USVs), unmanned underwater vehicles (UUVs), and their heterogeneous collaborations [101]. Key research directions for these platforms include search path planning, path-following control, search communications, and visual perception. While UAVs have proven useful in various SAR applications, their effectiveness in maritime environments is constrained by limited flight range, vulnerability to adverse weather, and inability to perform direct intervention. In contrast, unmanned or autonomous surface vessels (USVs/ASVs) offer notable advantages: they are more resilient to environmental disturbances, can carry substantial quantities of critical relief supplies, and are capable of deploying flotation aids or inflatable rafts for immediate use [102].

In addition to platform design considerations, autonomous surface vessels have been developed specifically for deepwater MSAR applications. For example, a USV prototype equipped with GPS and underwater sensors was successfully tested for searching victims, black boxes, debris, or other evidence on the surface and underwater. The vessel supports both autonomous navigation and manual operation via a ground station, which monitors telemetry data and provides control commands within a range of approximately 100 meters. This study highlights how autonomous platforms can significantly reduce crew risk during MSAR operations [103].

While the above work emphasizes autonomous platform capabilities and basic multi-sensor integration, other studies focus on multi-modal target detection for human-in-water scenarios. Recent research has proposed early-fusion detection pipelines combining LiDAR point clouds and thermal

doi:10.20944/preprints202510.0977.v1

imagery. In this approach, LiDAR provides geometric cues to mitigate thermal failures caused by haze or background noise, while thermal cameras compensate for LiDAR's sparse point density, enabling long-range human classification [104]. Such LiDAR–thermal fusion demonstrates how multimodal integration can substantially enhance robustness under adverse conditions and reduce false alarms, highlighting a promising direction for resilient SAR perception systems.

4.8. Illegal Activities Detection

Illegal maritime activities, including human trafficking, smuggling, and other non-cooperative vessel operations, pose significant threats to coastal security and international law enforcement. Detecting such vessels is particularly challenging because they often attempt to evade surveillance by disabling transponders, mimicking legitimate traffic patterns, or exploiting AIS vulnerabilities. Consequently, effective monitoring requires the fusion of heterogeneous maritime data, such as EO and SAR imagery, AIS signals, radar surveillance, and contextual socio-economic information.

A representative example is the detection of so-called "dark ships," vessels that deliberately disable AIS transponders to evade monitoring. To address this challenge, recent studies have explored multi-modal fusion frameworks that combine optical and SAR satellite imagery with AIS signals. By leveraging oriented bounding box–based object detection and advanced feature association algorithms, these approaches improve the accuracy of vessel identification under complex maritime conditions and enable effective day–night, all-weather surveillance [105].

Another research direction emphasizes trajectory-based detection of smuggling vessels. For instance, [106] introduced a framework that fuses high-speed radar trajectories with meteorological data, employing a parallel temporal convolutional network (TCN) for motion and weather feature extraction, followed by an LSTM-based decision fusion module for trajectory discrimination. By explicitly accounting for external factors such as poor visibility and nighttime conditions, this method achieves robust performance in distinguishing covert smuggling activities from normal maritime traffic.

Beyond trajectory analysis, multimodal pipelines integrating remote sensing imagery and AIS have also been developed to enhance illegal activity detection. One study proposed a system that combines SAR imagery, high-resolution optical data, and AIS signals, using an enhanced YOLOv10s detection model optimized for small vessels in cluttered maritime scenes [107]. The pipeline further incorporates AIS cross-matching algorithms, explainable AI tools (Eigen-CAM), and even carbon emission assessments during training to balance accuracy with sustainability. In addition, the release of the HS3-S2 dataset, featuring multimodal satellite images with varying resolutions, provides a valuable benchmark for improving robustness against noisy maritime backgrounds. Together, these developments illustrate the potential of multimodal fusion frameworks to deliver reliable monitoring solutions in AIS-poor or adversarial scenarios.

Table 2. Task analysis based on sensor fusion.

Senario	Specific Task	Ref	Tensor Types
Maritime Perception	Vessel Detection & Tracking	[14,81] [15,54,55] [82,108] [58]	AIS + EO AIS + SAR AIS + Radar AIS + Sonar
	Obstacle Detection & Recognition	[57] [84] [83]	EO + Sonar LiDAR + EO LiDAR + EO + Radar
	Scene Understanding	[36] [60] [59] [59]	EO + IR Multi-Sonar EO + Sonar EO + IR + LiDAR
Behavior Understanding	Anomaly Detection	[13] [88] [89]	AIS + Radar AIS + SAR EO + GPS
	Trajectory Prediction	[91,92] [93] [109]	AIS + Environment AIS + ENC + Radar AIS + Satellite images
	Collision Avoidance & Risk Assessment	[] [] []	
Decision & Mission Support	Search and Rescue Support	[104] [103] [110]	LiDAR + IR GPS + Sonar LiDAR + EO
	Illegal Activities Detection	[105,107] [106]	EO + SAR + AIS Radar + Meteorological data
	Autonomous Navigation	[111] [112] [113]	EO + IR + LiDAR + Radar + GPS EO + IR + LiDAR + Radar + ENC EO + Sonar + Radar + GNSS + ENC

4.9. Autonomous Navigation

Autonomous navigation is a core capability of Maritime Autonomous Surface Ships (MASS), enabling vessels to make safe and efficient decisions based on perceived environmental information. While scene understanding establishes situational awareness, autonomous navigation focuses on the decision-making and control processes that transform perception into action. These processes encompass three key tasks—path planning, collision avoidance, and trajectory tracking—where multimodal sensor fusion is essential to ensure robustness in complex or uncertain maritime conditions.

Path planning constitutes the initial step, where safe and efficient routes are generated at both global and local scales. At the global level, the integration of AIS and remote sensing data supports route optimization by accounting for traffic density, weather, and restricted zones. At the local level, onboard sensors such as radar, sonar, and cameras provide real-time updates on dynamic obstacles, which can be processed through search-based, optimization-based, or reinforcement learning methods. Multimodal fusion mitigates single-sensor weaknesses, for instance compensating for degraded optical imagery in foggy environments.

Collision avoidance requires rapid responses to encounters with nearby vessels or floating objects. The joint use of AIS, radar, and optical imagery enables both target recognition and short-term trajectory prediction, guided by international navigation rules (COLREGs). Recent approaches, including model predictive control and deep reinforcement learning, leverage multimodal fusion to balance safety, compliance, and efficiency in dynamic maritime scenarios.

Trajectory tracking and control translate planned routes into precise maneuvers under changing ocean conditions. The fusion of GNSS/INS with visual odometry improves positioning accuracy, while environmental sensors such as wind meters and current profilers provide feedback for adaptive control. These multimodal strategies ensure stable navigation performance even in the presence of currents, winds, or sensor noise.

Representative contributions illustrate the diverse applications of multimodal fusion. The Pohang Canal Dataset integrates multiple perception and navigation sensors (cameras, LiDAR, radar, GPS, and AHRS) along restricted waterways, offering a valuable benchmark for testing path planning and collision avoidance strategies. More recently, a cross-attention transformer–based approach demonstrated the potential of deep multimodal fusion, combining RGB and infrared imagery with sparse LiDAR, radar, and electronic charts to construct a bird's-eye view representation of vessel surroundings, enabling robust navigation even in adverse weather. Earlier studies, such as Wright and Baldauf's work, combined visual landmark detection with radar and sonar-based terrain tracking, emphasizing redundancy and certification for safety in autonomous navigation.

Together, these studies highlight how multimodal integration advances path planning, collision avoidance, and trajectory control, while also underscoring the need for redundancy and standardized benchmarks to ensure safe and reliable operation of future autonomous ships.

Summary and Outlook

The breadth of multimodal applications across maritime domains—from pollution monitoring and search and rescue to port logistics and tactical coordination—highlights the transformative potential of cross-sensor fusion and intelligent data integration. These emerging use cases demand systems that are not only capable of combining diverse modalities such as AIS, EO/SAR imagery, radar, audio, and textual data, but that can also reason over this information in real time, under uncertainty, and often with limited supervision.

A recurring theme across all application areas is the challenge of modality heterogeneity, where data formats, temporal resolutions, and sensor characteristics differ widely. For example, integrating AIS logs with satellite imagery or VHF distress signals requires sophisticated fusion architectures capable of aligning spatial, temporal, and semantic representations. Similarly, data sparsity and label scarcity—especially in rare-event scenarios such as illegal smuggling or extreme weather—limit the applicability of fully supervised deep learning models and call for the development of self-supervised, transfer learning, and domain adaptation techniques.

Another major bottleneck is the real-time operational requirement found in safety-critical applications like fleet coordination and SAR missions. This places strong constraints on inference latency, communication bandwidth, and system robustness, necessitating lightweight yet expressive fusion models that can operate reliably under degraded sensing conditions.

Moreover, explainability and trustworthiness are increasingly important, particularly in military, regulatory, and legal contexts. Decision-support systems in maritime operations must provide interpretable outputs that align with human expectations and can be audited or verified post-deployment.

While academic research has produced promising multimodal architectures tailored to individual maritime tasks, a key future direction is bridging the gap toward generalizable and deployable systems. This includes the design of shared benchmarks, the creation of large-scale and diverse multimodal datasets, and the development of unified frameworks that can accommodate new modalities or mission profiles with minimal retraining.

In conclusion, multimodal learning presents a compelling paradigm for advancing maritime intelligence, safety, and sustainability. Realizing its full potential will require concerted efforts in model design, system integration, and interdisciplinary collaboration across oceanography, computer vision, signal processing, and marine operations.

5. Public Datasets

Multimodal learning in maritime domains demands not only advanced algorithms but also high-quality datasets that reflect the diversity, complexity, and operational challenges of real-world scenarios. Building upon the multimodal data types introduced in Chapter 2, this chapter surveys publicly available datasets and benchmarks that support research in maritime perception, tracking, rescue, and decision-making.

Despite recent progress, public datasets for multimodal maritime applications remain limited in quantity, scope, and consistency. Existing datasets often suffer from a lack of standardized annotation formats, limited temporal or spatial resolution, and modality imbalances (e.g., missing synchronized labels across EO, radar, and AIS). Furthermore, task-specific benchmarks for multimodal fusion in maritime contexts—such as cross-modal retrieval, anomaly detection, or sensor alignment—are still underdeveloped. This section categorizes key datasets by task and modality, highlighting both their potential and limitations.

5.1. Vessel Detection

Ship detection datasets are mainly divided into two categories: Satellite remote sensing + radar fusion: SSDD, AIR-SARShip1.0 (AIR-SARShip2.0 has been released but not included in the statistics), SAR-Ship1.0; Satellite remote sensing + thermal infrared fusion: TISD.

SSDD is the first SAR ship detection dataset. It was constructed by collecting publicly available SAR images, cropping them to approximately 500×500 pixel regions, and manually annotating ship locations. However, its annotation accuracy was low and standards were inconsistent. SSDD+, released in 2021, introduced rotated bounding box annotation, provided both PASCAL VOC and COCO annotation formats, and established stricter testing specifications. It is suitable for small object detection and object orientation estimation, but the data size is still limited. Common benchmark models include Faster R-CNN, SSD, RetinaNet, and the YOLO series.

SAR-Ship-Dataset, based on Gaofen-3 and Sentinel-1 imagery, is the first large-scale SAR ship detection dataset covering complex environments. Common benchmark models include Faster R-CNN, SSD, RetinaNet, and the YOLO series.

AIR-SARShip1.0, built on Gaofen-3 images, uses the PASCAL VOC format and vertical bounding box annotation, covering a wider range of ship types and scenarios. Common benchmarks include Faster R-CNN, Cascade R-CNN, SSD, RetinaNet, and the YOLO series.

TISD is constructed from three-band infrared images collected by the SDGSAT-1 TIS sensor. It provides vertical bounding box annotations and can handle detection scenarios in complex weather conditions, light reflections, and low-light environments. However, due to the limited interpretability of CNN methods and the fuzzy boundaries of infrared targets, they are prone to false positives and missed detections. Common benchmarks include YOLOv5s, Faster R-CNN, SSD, and RetinaNet.

Overall, the aforementioned detection datasets commonly suffer from the following issues: 1. The number of offshore images is insufficient, limiting the model's generalization ability in complex offshore environments; 2. Ships are difficult to distinguish from land or are occluded; 3. Ship size definitions still rely on manual annotation; 4. The image resolution is relatively low, making fine classification difficult, resulting in ships being classified as a single category.

5.2. Vessl classification

Ship classification datasets are primarily divided into three categories: 1. Satellite remote sensing + visible light fusion: HRSC2016, FGSD, and ShipRSImageNet; 2. Satellite remote sensing + radar fusion: OpenSARShip; 3. Visible light + infrared fusion: VAIS and SEAGULL.

HRSC2016 is the first ship recognition dataset built based on high-resolution, unobstructed Google Earth imagery. It initially classifies ships into 25 categories and provides port location annotations using bounding boxes, rotated bounding boxes, and pixel-level classification. However, its coverage of ship categories and scenes is limited, and its image diversity is insufficient. Common benchmarks include R^{2}CNN, RRPN, RoI Transformer, Gliding Vertex, and RetinaNet-R.

FGSD is constructed from high-resolution Google Earth satellite imagery, covering ports in multiple countries. It provides 43 ship categories and multi-level labels, annotated using horizontal and rotated bounding boxes, and uses a ship's "V-shaped structure" to indicate orientation. While its coverage is broader than HRSC2016, the data is still primarily sourced from Google Earth. Common benchmarks include Faster R-CNN and R^{2}CNN.

ShipRSImageNet collects optical remote sensing images from various sensors, platforms, locations, and seasons worldwide. Using horizontal bounding boxes, rotated bounding boxes, and polygon annotations, it defines a four-level classification system and 50 ship categories, along with weather information. Its classification accuracy surpasses the aforementioned datasets. However, its limitation is that optical remote sensing cannot operate at night. Common benchmarks include Faster R-CNN+FPN, SSD, and Cascade Mask R-CNN+FPN.

OpenSARShip2.0 is built based on Sentinel-1 SAR imagery and AIS information, covering 87 scenes. It uses SNAP 3.0 and OCLT for semi-automatic annotation, expanding the scale of SAR ship data and enabling spatiotemporal correlation between AIS and SAR features. However, due to low resolution and imbalanced class distribution, its classification performance and generalization capabilities are limited. Common benchmarks include ResNet, the VGG series with attention mechanism, and multi-scale dual-attention CNN.

VAIS, a dataset capturing image sequences at six ports over a nine-day period, classifies six types of ships and is tested using CNNs and Gnostic Fields. It is a representative dataset for early implementations of EO and IR cross-modal fusion. However, it is small in scale, suffers from uneven matching of images, and has a limited number of categories and scenes.

SEAGULL, built on maritime surveillance video sequences, fuses visible, near-infrared, infrared, and hyperspectral data. This richness and robustness make it particularly suitable for detection and tracking tasks. However, the sample distribution is uneven at night, in adverse weather conditions, and for small targets at long distances, and spectral alignment and synchronization are also challenging. Common benchmarks include Faster R-CNN, YOLOv2, background subtraction, optical flow, and Kalman filter tracking methods.

5.3. Obstacle detection

Obstacle detection datasets are mainly divided into three categories: 1. Visible light + infrared + lidar + radar fusion: PoLaRIS; 2. Visible light + infrared fusion: Singapore Maritime Dataset; 3. Visible light + lidar fusion: SeePerSea.

PoLaRIS is built based on five video sequences from the Pohang canal dataset. It provides multimodal data from RGB, infrared, lidar, and radar, annotating 3D object bounding boxes, depth, and dynamic tracking information. It supports object annotations as small as 10×10 pixels, making it suitable for early hazard detection. However, its scenes are relatively simple and do not fully reflect perspective changes and motion blur.

The Singapore Maritime Dataset, constructed from footage captured by a Canon 70D camera in Singapore waters over 11 months, covers a variety of lighting and weather conditions. It uses manual annotations, resulting in high resolution and diverse backgrounds. However, it suffers from relatively simple scenes, high bounding box noise, and class imbalance.

SeePerSea, the first 3D multimodal water obstacle detection dataset, was collected by Catabot and manned vessels across three regions over a period of nearly four years. The dataset integrates RGB, LiDAR point clouds, and navigation data, covering multiple locations, weather conditions, and backgrounds. However, it only defines three target categories (ships, buoys, and other hazards) and lacks infrared and radar modalities, making it less robust in low-light conditions and inclement weather.

5.4. Auto Navigation

Autonomous navigation datasets are mainly divided into three categories: 1. Visible light + infrared + radar fusion: Pohang Canal dataset; 2. Visible light + lidar + radar fusion: MOANA; 3. Visible light + infrared + lidar + radar fusion: Maritime Sensor Fusion Benchmark.

The Pohang Canal dataset was collected by a small vessel in the Pohang Canal, South Korea. It covers the complex and narrow waterway and integrates multimodal information (including GPS, attitude, acceleration, etc.), making it suitable for target detection and tracking research. However, its limited geographic coverage and the lack of some GPS data make dynamic navigation tasks such as long-term trajectory prediction challenging.

MOANA, collected in Ulsan, South Korea and Singapore, integrates dual-frequency radar, LiDAR, stereo cameras, and GNSS, covering both structured and unstructured scenes. W-band radar enhances long- and short-range detection capabilities. However, its modal imbalance and limited geographic coverage make cross-modal alignment a prominent issue.

The Maritime Sensor Fusion Benchmark Dataset, collected by the milliAmpere platform, integrates five electro-optical cameras, five infrared cameras, radar, and lidar, making it suitable for multimodal detection and tracking research. While this dataset offers high fidelity, it is limited in size, lacks coverage of rare scenarios, and maintains multi-sensor time synchronization, which remains challenging.

5.5. Other Tasks

Some tasks only find a single public dataset: 1. Scene understanding (visible light + radar fusion): WaterScenes; 2. Traffic control (visible light + AIS fusion): FVessel; 3. Search and rescue (visible light + infrared fusion): SeaDroneSee.

WaterScenes is the first 4D radar-camera fusion dataset, collected over six months, covering a variety of temporal and environmental conditions, and providing annotations for seven categories of instances. However, it has a single viewpoint, lacks multi-camera information, and suffers from uneven data distribution.

FVessel, constructed from video and AIS data, covers the Wuhan section of the Yangtze River. It proposes an asynchronous matching method for AIS and video trajectories, focusing on addressing occlusion and trajectory fusion issues. However, its scale is limited, it lacks extreme weather scenarios, and its modal coverage is limited.

SeaDroneSee, based on drone data collected at different altitudes and viewpoints, annotates objects such as swimmers. It is large-scale, more realistic, and supports multiple lighting and background conditions. However, it has a single modality, uneven categories, and lacks spatiotemporal continuity.

Table 3. Summary of data .

Application	Main Modalities	Representative Datasets	Common Pixels	Ref
Vessel Detection	Satellite sensing+Radar	SSDD	500×500	[114]
		SSDD+	500×500	[115]
		SAR-Ship-Dataset	crop:256×256	[116]
		AIR-SARShip1.0	raw:3000×3000 crop:500×500	[117]
	Satellite sensing+IR	TISD	768×768	[118]
Vessl classification	Satellite sensing+EO	HRSC2016	raw:from300×300to1500×900	[119]
		FGSD	930×930	[120]
		ShipRSImageNet	raw:930×930	[121]
	Radar+AIS	OpenSARShip2.0	-	[122]
	EO+IR	VAIS	EO:145833 IR:8544	[123]
		SEAGULL	EO:1920×1080 IR:384×288	[124]
Obstacle detection	EO+IR+Lidar+Radar	PoLaRIS	EO:2048×1080(PNG) 2464×2048(JPG) IR:640×512	[125]
	EO+IR	Singapore Maritime Dataset	EO:1080×1920	[126]
	EO+Lidar	SeePerSea	EO:640×480	[127]
Scene understanding	EO+Radar	WaterScenes	EO:1920×1080(raw) 640×640(crop)	[128]
Traffic Monitoring	AIS + EO	FVessel	2560×1440	[129]
Search and rescue	EO+IR	SeaDronesSee	EO:from3840×2160to5456×3632	[130]
Auto navigation	EO+IR+Radar	Pohang canal dataset	EO:2048×1080(PNG) 2464×2048(JPG) IR:640×512	[111]
	EO+Lidar+Radar	MOANA	-	[131]
	EO+IR+Lidar+Radar	Maritime Sensor Fusion Benchmark	EO: 1224x1020 IR: 640x512	[132]

6. Future and Challenge

MMDL is increasingly applied in the maritime domain, yet many fusion paradigms and architectures developed in other fields remain underexplored. For example, cross-modal large language models (LLMs) and vision-language transformers (e.g., CLIP, BLIP) have demonstrated strong generalization in open-domain tasks, but their adaptation to maritime scenarios—such as combining EO imagery, radar, AIS, and textual metadata (e.g., ship logs, weather reports)—remains limited. Similarly, recent advances in foundation models and prompt-tuned multimodal encoders hold promise for anomaly detection and event prediction, though challenges such as data scarcity and domain shift hinder direct application.

Beyond these, methods like multimodal causal inference, spatiotemporal graph fusion, and dynamic modality gating—successful in medical imaging and autonomous driving—have not been systematically investigated for maritime perception. This highlights the need for cross-domain knowledge transfer to address unique maritime challenges.

Despite recent progress, several obstacles remain for deploying robust deep learning–based multimodal systems in real-world maritime environments. These challenges arise from the heterogeneity of sensor modalities, the dynamic and uncertain nature of the ocean, and limitations in dataset availability and onboard computational capacity. In addition, operational requirements such as real-time processing, reliable navigation, and safety compliance impose strict constraints on system design. To provide a structured overview, we categorize the remaining challenges into interconnected technical dimensions.

6.1. Data Synchronization and Temporal Alignment

A fundamental challenge in multimodal perception for autonomous maritime systems is the temporal alignment of heterogeneous sensor data. Maritime sensors—including AIS, radar, EO, IR, sonar, and satellite observations—operate at different sampling rates and may exhibit variable temporal accuracy, leading to mismatches that complicate real-time fusion and decision-making. This problem is further exacerbated in maritime environments by sensor drift, communication latency, and irregular signal transmission, particularly in remote or high-traffic areas where AIS or satellite updates may be delayed [1,4].

Effective temporal alignment is crucial when combining historical and real-time data for tasks such as vessel detection, trajectory prediction, and anomaly recognition. Advanced approaches leveraging deep learning, including recurrent neural networks (RNNs) and transformer-based architectures, have shown promise in predicting and compensating for time lags in other domains, such as autonomous driving and UAV perception [71,133]. Self-supervised learning techniques have also been proposed to learn temporal correspondences without extensive labeled data, reducing the impact of asynchronous streams and improving robustness in real-world operations [134].

In the maritime context, adapting these AI-based temporal alignment methods to account for environmental factors, variable sensor update rates, and heterogeneous data types is essential for building reliable, real-time multimodal fusion frameworks that underpin autonomous navigation and situational awareness.

6.2. Limited Computational and Communication Resources

Autonomous maritime platforms frequently operate in remote or offshore environments, where both computational power and communication bandwidth are inherently constrained. These platforms must process and fuse large volumes of heterogeneous sensor data in real time to support navigation, situational awareness, and decision-making. The imbalance between the data volume generated by these multimodal sensors and the onboard processing capabilities poses a critical challenge for timely and reliable fusion of sensor information.Consequently, the development of lightweight and computationally efficient fusion architectures is essential to ensure that autonomous maritime systems can operate effectively under limited onboard resources [135].

Communication bandwidth constitutes another major limitation. High-resolution EO/IR imagery, radar returns, and satellite-derived datasets produce large data streams that can exceed the transmission capacity of maritime networks, particularly in areas far from shore-based infrastructure [136]. This constraint can introduce delays, reduce the timeliness of situational awareness, and hinder coordinated operations across platforms.

To address these limitations, recent research has explored the integration of edge computing and onboard data pre-processing. By performing computation directly on the vessel or unmanned surface/underwater platforms, edge computing reduces the reliance on high-bandwidth links and enables low-latency decision-making [137,138]. In parallel, data compression techniques, including both lossless and lossy methods, have been applied to reduce the volume of sensor data transmitted without significantly degrading information quality [139].

Combining edge computing with efficient compression algorithms is particularly beneficial for multimodal fusion systems, where synchronizing and processing heterogeneous data streams is computationally intensive. These strategies collectively support real-time perception, trajectory prediction, and anomaly detection in maritime environments, even under stringent resource constraints.

6.3. Security and Privacy Concerns

The increasing reliance on digital technologies in maritime operations has introduced significant security and privacy challenges. Sensitive information, such as vessel locations, cargo details, and operational patterns, must be protected to prevent unauthorized access, spoofing, or manipulation [140,141]. Cyberattacks targeting navigation, communication, or data processing systems can disrupt maritime operations and pose safety risks.

Maritime platforms often operate in remote or offshore regions with intermittent connectivity, complicating secure data transmission. Conventional encryption and authentication protocols, while effective in terrestrial networks, may be insufficient under such conditions [142]. Blockchain and distributed ledger technologies have been proposed to enhance data integrity and trustworthiness, enabling secure sharing of AIS and sensor data across heterogeneous maritime systems [143,144].

Legacy maritime systems further exacerbate security risks, as many were not designed for internet connectivity and real-time data exchange. Ensuring compliance with international regulations, such as the International Maritime Organization's (IMO) guidelines on maritime cybersecurity [145,146], while maintaining operational efficiency, remains a critical challenge.

Future research should focus on adaptive, intelligent security protocols tailored for maritime multimodal fusion systems. Such solutions must dynamically detect, assess, and respond to cyber threats in real time, ensuring both data integrity and operational safety. Integration of lightweight encryption, anomaly detection using AI, and resilient communication frameworks will be key for secure and efficient autonomous maritime operations [147,148].

6.4. AI Trustworthiness and Explainability in Maritime Applications

As AI and deep learning models increasingly support decision-making in maritime operations, ensuring their trustworthiness, reliability, and explainability has become critical. Maritime tasks such as autonomous navigation, collision avoidance, and emergency response are safety-critical, and operators must be able to understand and oversee AI-driven decisions [149,150]. The black-box nature of many deep learning models can undermine confidence, particularly in high-stakes scenarios where human intervention remains essential.

Multimodal fusion in maritime systems—integrating AIS, radar, EO/IR imagery, and other sensors—amplifies the complexity of interpretability. Operators need transparency in how fused sensor data informs AI outputs, especially under dynamic environmental conditions, such as variable sea states, adverse weather, or high vessel traffic [151]. Explainable AI (XAI) methods, including attention-based visualization, feature attribution, and uncertainty quantification, can provide insights into model reasoning and highlight which sensor inputs most influence predictions [152,153].

Human-AI collaboration is essential: AI should augment operator decision-making rather than replace it. Confidence measures and interpretable outputs allow operators to assess prediction reliability and make informed decisions, particularly when the system faces ambiguous or conflicting sensor data [154,155]. Developing visualization tools and interface designs that clearly present contributions from multiple sensor modalities is a key step toward fostering trust in AI-enabled maritime systems.

In summary, establishing trustworthy and explainable AI is fundamental for safe and effective deployment of autonomous maritime platforms. Future research should focus on integrating XAI techniques into multimodal fusion frameworks, enabling operators to understand, evaluate, and control AI-driven decisions in real-time maritime environments.

7. Conclusions

Multimodal data fusion plays a central role in advancing maritime autonomy by integrating heterogeneous sensors such as AIS, radar, EO/IR imagery, sonar, LiDAR, and satellite observations. These techniques enhance vessel detection, tracking, obstacle recognition, and trajectory prediction under complex and dynamic maritime conditions. This survey systematically categorized fusion strategies into early, intermediate, and late integration, reviewing classical deep learning–based approaches. Key applications, including maritime situational awareness, anomaly detection, search and rescue, and domain monitoring, were analyzed alongside available datasets and benchmarking platforms. Despite progress, challenges remain in temporal-spatial alignment, data sparsity, environmental robustness, and standardized evaluation. Future research directions include self-supervised cross-modal alignment, real-time edge-friendly fusion frameworks, simulation-to-real generalization, and domain-adapted foundation models. Addressing these challenges is essential for developing scalable, reliable, and robust multimodal systems for autonomous maritime operations.

Author Contributions: Conceptualization, C.W.; investigation, C.W.; resources, C.W.; writing—original draft preparation, C.W.,Y.X.; writing—review and editing, D.Z.; funding acquisition, C.W. All authors have read and agreed to the published version of the manuscript.

Funding: ‘This research was funded by Dalian Maritime University, grant number 3132025232.

Data Availability Statement: All datasets included in this study are public datasets and can be obtained by contacting the links in the relevant papers or directly searching on the web.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

EO	Electro-Optical Sensor
IR	Infrared Sensor
AIS	Automatic Identification System
SAR	Synthetic Aperture Radar
MMDL	Multimodal deep learning

References

1. Svanberg, M.; Santén, V.; Hörteborn, A.; Holm, H.; Finnsgård, C. AIS in maritime research. *Marine Policy* **2019**, *106*, 103520. https://doi.org/https://doi.org/10.1016/j.marpol.2019.103520.
2. Kanjir, U.; Greidanus, H.; Oštir, K. Vessel detection and classification from spaceborne optical images: A literature survey. *Remote Sensing of Environment* **2018**, *207*, 1–26. https://doi.org/https://doi.org/10.1016/j.rse.2017.12.033.
3. Lyu, H.; Shao, Z.; Cheng, T.; Yin, Y.; Gao, X. Sea-Surface Object Detection Based on Electro-Optical Sensors: A Review. *IEEE Intelligent Transportation Systems Magazine* **2023**, *15*, 190–216. https://doi.org/10.1109/MITS.2022.3198334.
4. Zhang, Z.; Zhang, L.; Wu, J.; Guo, W. Optical and synthetic aperture radar image fusion for ship detection and recognition: Current state, challenges, and future prospects. *IEEE Geoscience and Remote Sensing Magazine* **2024**.
5. Sleeman IV, W.C.; Kapoor, R.; Ghosh, P. Multimodal classification: Current landscape, taxonomy and future directions. *ACM Computing Surveys* **2022**, *55*, 1–31.
6. Jabeen, S.; Li, X.; Amin, M.S.; Bourahla, O.; Li, S.; Jabbar, A. A review on methods and applications in multimodal deep learning. *ACM Transactions on Multimedia Computing, Communications and Applications* **2023**, *19*, 1–41.
7. Tang, Q.; Liang, J.; Zhu, F. A comparative review on multi-modal sensors fusion based on deep learning. *Signal Processing* **2023**, *213*, 109165.
8. Portillo Juan, N.; Negro Valdecantos, V.; Troch, P. Advancing artificial intelligence in ocean and maritime engineering: Trends, progress, and future directions. *Ocean Engineering* **2025**, *339*, 122077. https://doi.org/https://doi.org/10.1016/j.oceaneng.2025.122077.
9. Artificial Intelligence in Maritime Transportation: A Comprehensive Review of Safety and Risk Management Applications. *Applied Sciences* **2024**, *14*. https://doi.org/10.3390/app14188420.
10. Robards, M.; Silber, G.; Adams, J.; Arroyo, J.; Lorenzini, D.; Schwehr, K.; Amos, J. Conservation science and policy applications of the marine vessel Automatic Identification System (AIS)—a review. *Bulletin of Marine Science* **2016**, *92*, 75–103. https://doi.org/10.5343/bms.2015.1034.

11. Tu, E.; Zhang, G.; Rachmawati, L.; Rajabally, E.; Huang, G.B. Exploiting AIS data for intelligent maritime navigation: A comprehensive survey from data to methodology. *IEEE Transactions on Intelligent Transportation Systems* **2017**, *19*, 1559–1582.
12. Huang, W.; Feng, H.; Xu, H.; Liu, X.; He, J.; Gan, L.; Wang, X.; Wang, S. Surface Vessels Detection and Tracking Method and Datasets with Multi-Source Data Fusion in Real-World Complex Scenarios. *Sensors* **2025**, *25*. https://doi.org/10.3390/s25072179.
13. Ming, W.C.; Yanan, L.; Lanxi, M.; Jiuhu, C.; Zhong, L.; Sunxin, S.; Yuanchao, Z.; Qianying, C.; Yugui, C.; Xiaoxue, D.; et al. Intelligent marine area supervision based on AIS and radar fusion. *Ocean Engineering* **2023**, *285*, 115373.
14. Huang, Z.; Hu, Q.; Lu, L.; Mei, Q.; Yang, C. Online Estimation of Ship Dimensions by Combining Images with AIS Reports. *Journal of Marine Science and Engineering* **2023**, *11*, 1700.
15. Lee, Y.K.; Jung, H.C.; Kim, K.; Jang, Y.; Ryu, J.H.; Kim, S.W. Assessment of maritime vessel detection and tracking using integrated SAR Imagery and AIS/V-Pass Data. *Ocean Science Journal* **2024**, *59*, 27.
16. Huang, W.; Liu, X.; Gill, E.W. Ocean wind and wave measurements using X-band marine radar: A comprehensive review. *Remote sensing* **2017**, *9*, 1261.
17. Cui, C.; Ma, Y.; Lu, J.; Wang, Z. Radar enlightens the dark: Enhancing low-visibility perception for automated vehicles with camera-radar fusion. In Proceedings of the 2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC). IEEE, 2023, pp. 2726–2733.
18. Ersü, C.; Petlenkov, E.; Janson, K. A Systematic Review of Cutting-Edge Radar Technologies: Applications for Unmanned Ground Vehicles (UGVs). *Sensors* **2024**, *24*, 7807.
19. Srivastav, A.; Mandal, S. Radars for autonomous driving: A review of deep learning methods and challenges. *IEEE Access* **2023**, *11*, 97147–97168.
20. Liu, X.; Li, Y.; Wu, Y.; Wang, Z.; He, W.; Li, Z. A Hybrid Method for Inland Ship Recognition Using Marine Radar and Closed-Circuit Television. *Journal of Marine Science and Engineering* **2021**, *9*.
21. Kim, H.; Kim, D.; Lee, S.M. Marine Object Segmentation and Tracking by Learning Marine Radar Images for Autonomous Surface Vehicles. *IEEE Sensors Journal* **2023**, *23*, 10062–10070. https://doi.org/10.1109/JSEN.2023.3259471.
22. Sun, S.; Lyu, H.; Dong, C. AIS aided marine radar target tracking in a detection occluded environment. *Ocean Engineering* **2023**, *288*, 116133.
23. Ninos, A.; Hasch, J.; Alvarez, M.E.P.; Zwick, T. Synthetic radar dataset generator for macro-gesture recognition. *IEEE Access* **2021**, *9*, 76576–76584.
24. Chen, X.; Huang, W. Spatial–Temporal Convolutional Gated Recurrent Unit Network for Significant Wave Height Estimation From Shipborne Marine Radar Data. *IEEE Transactions on Geoscience and Remote Sensing* **2022**, *60*, 1–11. https://doi.org/10.1109/TGRS.2021.3074075.
25. Horstmann, J.; Bödewadt, J.; Carrasco, R.; Cysewski, M.; Seemann, J.; Streβer, M. A coherent on receive X-band marine radar for ocean observations. *Sensors* **2021**, *21*, 7828.
26. Wang, L.; Fan, S.; Liu, Y.; Li, Y.; Fei, C.; Liu, J.; Liu, B.; Dong, Y.; Liu, Z.; Zhao, X. A review of methods for ship detection with electro-optical images in marine environments. *Journal of Marine Science and Engineering* **2021**, *9*, 1408.
27. Giompapa, S.; Croci, R.; Di Stefano, R.; Farina, A.; Gini, F.; Graziano, A.; Lapierre, F. Naval target classification by fusion of IR and EO sensors. In Proceedings of the Electro-Optical and Infrared Systems: Technology and Applications IV. SPIE, 2007, Vol. 6737, pp. 277–288.
28. Prasad, D.K.; Rajan, D.; Rachmawati, L.; Rajabally, E.; Quek, C. Video processing from electro-optical sensors for object detection and tracking in a maritime environment: A survey. *IEEE Transactions on Intelligent Transportation Systems* **2017**, *18*, 1993–2016.
29. Ballan, L.; Melo, J.G.; van den Broek, S.P.; Baan, J.; Heslinga, F.G.; Huizinga, W.; Dijk, J.; Dilo, A. EO and radar fusion for fine-grained target classification with a strong few-shot learning baseline. In Proceedings of the Signal Processing, Sensor/Information Fusion, and Target Recognition XXXIII. SPIE, 2024, Vol. 13057, pp. 206–218.
30. Ma, Z.; Wen, J.; Liang, X. Video image clarity algorithm research of USV visual system under the sea fog. In Proceedings of the International Conference in Swarm Intelligence. Springer, 2013, pp. 436–444.
31. Prasad, D.K.; Prasath, C.K.; Rajan, D.; Rachmawati, L.; Rajabally, E.; Quek, C. Challenges in Video Based Object Detection in Maritime Scenario Using Computer Vision. *International Journal of Computer and Information Engineering* **2017**, *11*, 31 – 36.

32. Chaar, M.M.; Raiyn, J.; Weidl, G. Improving the Perception of Objects Under Daylight Foggy Conditions in the Surrounding Environment. *Vehicles* **2024**, *6*, 2154–2169.
33. Zheng, Y.; Chen, Z.; Lv, D.; Li, Z.; Lan, Z.; Zhao, S. Air-to-air visual detection of micro-uavs: An experimental evaluation of deep learning. *IEEE Robotics and automation letters* **2021**, *6*, 1020–1027.
34. Ruan, J.; Cui, H.; Huang, Y.; Li, T.; Wu, C.; Zhang, K. A review of occluded objects detection in real complex scenarios for autonomous driving. *Green energy and intelligent transportation* **2023**, *2*, 100092.
35. Bijelic, M.; Gruber, T.; Mannan, F.; Kraus, F.; Ritter, W.; Dietmayer, K.; Heide, F. Seeing through fog without seeing fog: Deep multimodal sensor fusion in unseen adverse weather. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 11682–11692.
36. Liu, Y.; Dong, L.; Xu, W. Infrared and visible image fusion for shipborne electro-optical pod in maritime environment. *Infrared Physics & Technology* **2023**, *128*, 104526.
37. Narayanan, R.M.; Wood, N.S.; Lewis, B.P. Assessment of various multimodal fusion approaches using synthetic aperture radar (SAR) and electro-optical (EO) imagery for vehicle classification via neural networks. *Sensors* **2023**, *23*, 2207.
38. Nagaraju, B.; Rajesh, R. Integrated Electro Optic Infra Red(EO/IR) Simulation for Airborne Maritime Surveillance. In Proceedings of the 2024 IEEE Space, Aerospace and Defence Conference (SPACE), 2024, pp. 501–504. https://doi.org/10.1109/SPACE63117.2024.10667798.
39. Leonard, C.L.; DeWeert, M.J.; Gradie, J.; Iokepa, J.; Stalder, C.L. Performance of an EO/IR sensor system in marine search and rescue. In Proceedings of the Airborne Intelligence, Surveillance, Reconnaissance (ISR) Systems and Applications II. SPIE, 2005, Vol. 5787, pp. 122–133.
40. Cocking, J.; Narayanaswamy, B.E.; Waluda, C.M.; Williamson, B.J. Aerial detection of beached marine plastic using a novel, hyperspectral short-wave infrared (SWIR) camera. *ICES Journal of Marine Science* **2022**, *79*, 648–660, [https://academic.oup.com/icesjms/article-pdf/79/3/648/43513594/fsac006.pdf]. https://doi.org/10.1093/icesjms/fsac006.
41. Nirgudkar, S.; Robinette, P. Beyond Visible Light: Usage of Long Wave Infrared for Object Detection in Maritime Environment. In Proceedings of the 2021 20th International Conference on Advanced Robotics (ICAR), 2021, pp. 1093–1100. https://doi.org/10.1109/ICAR53236.2021.9659477.
42. Goddijn-Murphy, L.; Williamson, B.J.; McIlvenny, J.; Corradi, P. Using a UAV thermal infrared camera for monitoring floating marine plastic litter. *Remote Sensing* **2022**, *14*, 3179.
43. Bustos, N.; Mashhadi, M.; Lai-Yuen, S.K.; Sarkar, S.; Das, T.K. A systematic literature review on object detection using near infrared and thermal images. *Neurocomputing* **2023**, *560*, 126804. https://doi.org/https://doi.org/10.1016/j.neucom.2023.126804.
44. Hu, C. Remote detection of marine debris using satellite observations in the visible and near infrared spectral range: Challenges and potentials. *Remote Sensing of Environment* **2021**, *259*, 112414.
45. Nagaraju, B.; Rajesh, R. Integrated Electro Optic Infra Red(EO/IR) Simulation for Airborne Maritime Surveillance. In Proceedings of the 2024 IEEE Space, Aerospace and Defence Conference (SPACE), 2024, pp. 501–504. https://doi.org/10.1109/SPACE63117.2024.10667798.
46. Panda, S.S.; Rao, M.N.; Thenkabail, P.S.; Misra, D.; Fitzgerald, J.P. Remote sensing systems—Platforms and sensors: Aerial, satellite, UAV, optical, radar, and LiDAR. In *Remote Sensing Handbook, Volume I*; CRC Press, 2016; pp. 3–86.
47. Soldi, G.; Gaglione, D.; Forti, N.; Di Simone, A.; Daffinà, F.C.; Bottini, G.; Quattrociocchi, D.; Millefiori, L.M.; Braca, P.; Carniel, S.; et al. Space-based global maritime surveillance. Part I: Satellite technologies. *IEEE Aerospace and Electronic Systems Magazine* **2021**, *36*, 8–28.
48. Kavzoglu, T.; Tso, B.; Mather, P.M. *Classification methods for remotely sensed data*; CRC press, 2024.
49. Zhao, T.; Wang, Y.; Li, Z.; Gao, Y.; Chen, C.; Feng, H.; Zhao, Z. Ship detection with deep learning in optical remote-sensing images: A survey of challenges and advances. *Remote Sensing* **2024**, *16*, 1145.
50. Shaban, A. Use of satellite images to identify marine pollution along the Lebanese coast. *Environmental Forensics* **2008**, *9*, 205–214.
51. Sasaki, K.; Sekine, T.; Burtz, L.J.; Emery, W.J. Coastal marine debris detection and density mapping with very high resolution satellite imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **2022**, *15*, 6391–6401.
52. Mattyus, G. Near real-time automatic marine vessel detection on optical satellite images. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* **2013**, *40*, 233–237.
53. Zucchetta, M.; Madricardo, F.; Ghezzo, M.; Petrizzo, A.; Picciulin, M. Satellite-based monitoring of small boat for environmental studies: a systematic review. *Journal of Marine Science and Engineering* **2025**, *13*, 390.

54. Wang, X.; Song, X.; Zhao, Y. Identification and Positioning of Abnormal Maritime Targets Based on AIS and Remote-Sensing Image Fusion. *Sensors* **2024**, *24*. https://doi.org/10.3390/s24082443.
55. Graziano, M.D.; Renga, A.; Moccia, A. Integration of Automatic Identification System (AIS) Data and Single-Channel Synthetic Aperture Radar (SAR) Images by SAR-Based Ship Velocity Estimation for Maritime Situational Awareness. *Remote Sensing* **2019**, *11*. https://doi.org/10.3390/rs11192196.
56. Aubard, M.; Madureira, A.; Teixeira, L.; Pinto, J. Sonar-Based Deep Learning in Underwater Robotics: Overview, Robustness, and Challenges. *IEEE Journal of Oceanic Engineering* **2025**.
57. Yu, F.; Xiao, F.; Li, C.; Cheng, E.; Yuan, F. AO-UOD: A Novel Paradigm for Underwater Object Detection Using Acousto–Optic Fusion. *IEEE Journal of Oceanic Engineering* **2025**.
58. Zhao, W.; Xiong, X.; Cheng, X.; Zhang, X.; Wang, D. The Feature-level Fusion of AIS and Sonar Information based on LSTM-Transformer Joint Model. In Proceedings of the Journal of Physics: Conference Series. IOP Publishing, 2025, Vol. 3007, p. 012053.
59. Kim, H.G.; Seo, J.; Kim, S.M. Underwater optical-sonar image fusion systems. *Sensors* **2022**, *22*, 8445.
60. Rho, S.; Joe, H.; Sung, M.; Kim, J.; Kim, S.; Yu, S.C. Multi-Sonar Fusion-Based Precision Underwater 3D Reconstruction for Optimal Scan Path Planning of AUV. *IEEE Access* **2025**.
61. Farahnakian, F.; Heikkonen, J. Deep learning based multi-modal fusion architectures for maritime vessel detection. *Remote Sensing* **2020**, *12*, 2509.
62. Subedi, D.; Jha, A.; Tyapin, I.; Hovland, G. Camera-lidar data fusion for autonomous mooring operation. In Proceedings of the 2020 15th IEEE Conference on Industrial Electronics and Applications (ICIEA). IEEE, 2020, pp. 1176–1181.
63. Garczyńska-Cyprysiak, I.; Kazimierski, W.; Włodarczyk-Sielicka, M. Neural Approach to Coordinate Transformation for LiDAR–Camera Data Fusion in Coastal Observation. *Sensors* **2024**, *24*, 6766.
64. Wang, J.; Wang, L.; Feng, S.; Peng, B.; Huang, L.; Fatholahi, S.N.; Tang, L.; Li, J. An overview of shoreline mapping by using airborne LiDAR. *Remote Sensing* **2023**, *15*, 253.
65. Li, S.; Tang, H. Multimodal alignment and fusion: A survey. *arXiv preprint arXiv:2411.17040* **2024**.
66. Achiri, L.; Guida, R.; Iervolino, P. SAR and AIS fusion for maritime surveillance. In Proceedings of the 2018 IEEE 4th International Forum on Research and Technology for Society and Industry (RTSI). IEEE, 2018, pp. 1–4.
67. Liu, J.; Chen, H.; Wang, Y. Multi-source remote sensing image fusion for ship target detection and recognition. *Remote Sensing* **2021**, *13*, 4852.
68. Luo, J.; Xiao, Y.; Li, Y.; Xiao, Y.; Yao, W. Multimodal deep learning framework for vessel trajectory prediction. *Ocean Engineering* **2025**, *336*, 121766. https://doi.org/https://doi.org/10.1016/j.oceaneng.2025.121766.
69. Zhao, Z.; Ji, K.; Xing, X.; Zou, H.; Zhou, S. Ship surveillance by integration of space-borne SAR and AIS–further research. *The Journal of Navigation* **2014**, *67*, 295–309.
70. Chen, J.; Xu, X.; Zhang, J.; Xu, G.; Zhu, Y.; Liang, B.; Yang, D. Ship target detection algorithm based on decision-level fusion of visible and SAR images. *IEEE Journal on Miniaturization for Air and Space Systems* **2023**, *4*, 242–249.
71. Huang, K.; Shi, B.; Li, X.; Li, X.; Huang, S.; Li, Y. Multi-modal sensor fusion for auto driving perception: A survey. *arXiv preprint arXiv:2202.02703* **2022**.
72. Baltrusaitis, T.; Ahuja, C.; Morency, L.P. Multimodal Machine Learning: A Survey and Taxonomy. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *41*, 423–443. https://doi.org/10.1109/TPAMI.2018.2798607.
73. Deng, C.; Deng, Z.; Lu, S.; He, M.; Miao, J.; Peng, Y. Fault diagnosis method for imbalanced data based on multi-signal fusion and improved deep convolution generative adversarial network. *Sensors* **2023**, *23*, 2542.
74. Zhang, Q.; Wei, Y.; Han, Z.; Fu, H.; Peng, X.; Deng, C.; Hu, Q.; Xu, C.; Wen, J.; Hu, D.; et al. Multimodal fusion on low-quality data: A comprehensive survey. *arXiv preprint arXiv:2404.18947* **2024**.
75. Wang, Y.; Huang, W.; Sun, F.; Xu, T.; Rong, Y.; Huang, J. Deep multimodal fusion by channel exchanging. *Advances in neural information processing systems* **2020**, *33*, 4835–4845.
76. Sun, Y.; Fu, Z.; Sun, C.; Hu, Y.; Zhang, S. Deep multimodal fusion network for semantic segmentation using remote sensing image and LiDAR data. *IEEE Transactions on Geoscience and Remote Sensing* **2021**, *60*, 1–18.
77. Chen, S.; Ma, X.; Zhang, H.; Li, H.; Sun, B.; Wang, Z. Real-Time Depth Completion With Multimodal Feature Alignment. *IEEE Transactions on Neural Networks and Learning Systems* **2025**.
78. Shao, Z.; Wang, H.; Cai, Y.; Chen, L.; Li, Y. UA-Fusion: Uncertainty-Aware Multimodal Data Fusion Framework for 3D Object Detection of Autonomous Vehicles. *IEEE Transactions on Instrumentation and Measurement* **2025**.

79. Chen, L.; Wang, J.; Mortlock, T.; Khargonekar, P.; Al Faruque, M.A. Hyperdimensional uncertainty quantification for multimodal uncertainty fusion in autonomous vehicles perception. In Proceedings of the Proceedings of the Computer Vision and Pattern Recognition Conference, 2025, pp. 22306–22316.
80. Kanjir, U.; Greidanus, H.; Oštir, K. Vessel detection and classification from spaceborne optical images: A literature survey. *Remote sensing of environment* **2018**, *207*, 1–26.
81. Guo, Y.; Liu, R.W.; Qu, J.; Lu, Y.; Zhu, F.; Lv, Y. Asynchronous trajectory matching-based multimodal maritime data fusion for vessel traffic surveillance in inland waterways. *IEEE Transactions on Intelligent Transportation Systems* **2023**, *24*, 12779–12792.
82. Liu, W.; Liu, Y.; Gunawan, B.A.; Bucknall, R. Practical moving target detection in maritime environments using fuzzy multi-sensor data fusion. *International Journal of Fuzzy Systems* **2021**, *23*, 1860–1878.
83. Stanislas, L.; Dunbabin, M. Multimodal sensor fusion for robust obstacle detection and classification in the maritime RobotX challenge. *IEEE Journal of Oceanic Engineering* **2018**, *44*, 343–351.
84. Zhang, Q.; Shan, Y.; Zhang, Z.; Lin, H.; Zhang, Y.; Huang, K. Multisensor fusion-based maritime ship object detection method for autonomous surface vehicles. *Journal of Field Robotics* **2024**, *41*, 493–510.
85. Stach, T.; Kinkel, Y.; Constapel, M.; Burmeister, H.C. Maritime anomaly detection for vessel traffic services: A survey. *Journal of Marine Science and Engineering* **2023**, *11*, 1174.
86. Riveiro, M.; Pallotta, G.; Vespe, M. Maritime anomaly detection: A review. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* **2018**, *8*, e1266.
87. Wang, Y.; Liu, J.; Liu, R.W.; Liu, Y.; Yuan, Z. Data-driven methods for detection of abnormal ship behavior: Progress and trends. *Ocean Engineering* **2023**, *271*, 113673. https://doi.org/https://doi.org/10.1016/j.oceaneng.2023.113673.
88. Chen, L.; Hu, Z.; Chen, J.; Sun, Y. SVIADF: Small Vessel Identification and Anomaly Detection Based on Wide-Area Remote Sensing Imagery and AIS Data Fusion. *Remote Sensing* **2025**, *17*, 868.
89. Tao, M.; Jiang, Z.; Shen, X.; Liu, Y.; Fang, X.; Chen, Z.; Zhao, G.; Li, X.; Ma, J.; Yu, P.S. Anomaly detection in unmanned surface vehicles via multimodal learning. *Ocean Engineering* **2025**, *340*, 122204.
90. Zhang, X.; Fu, X.; Xiao, Z.; Xu, H.; Qin, Z. Vessel trajectory prediction in maritime transportation: Current approaches and beyond. *IEEE Transactions on Intelligent Transportation Systems* **2022**, *23*, 19980–19998.
91. Luo, J.; Xiao, Y.; Li, Y.; Xiao, Y.; Yao, W. Multimodal deep learning framework for vessel trajectory prediction. *Ocean Engineering* **2025**, *336*, 121766.
92. Xiao, Y.; Hu, Y.; Liu, J.; Xiao, Y.; Liu, Q. An Adaptive Multimodal Data Vessel Trajectory Prediction Model Based on a Satellite Automatic Identification System and Environmental Data. *Journal of Marine Science and Engineering* **2024**, *12*, 513.
93. Dijt, P.; Mettes, P. Trajectory prediction network for future anticipation of ships. In Proceedings of the Proceedings of the 2020 international conference on multimedia retrieval, 2020, pp. 73–81.
94. Yang, X.; Zhi, J.; Zhang, W.; Xu, S.; Meng, X. A novel data-driven prediction framework for ship navigation accidents in the Arctic region. *Journal of Marine Science and Engineering* **2023**, *11*, 2300.
95. Lan, H.; Ma, X.; Qiao, W.; Deng, W. Determining the critical risk factors for predicting the severity of ship collision accidents using a data-driven approach. *Reliability Engineering & System Safety* **2023**, *230*, 108934.
96. Dong, H.; Zhen, R.; Gu, Q.; Lin, Z.; Chen, J.; Yan, K.; Chen, B. A novel collaborative collision avoidance decision method for multi-ship encounters in complex waterways. *Ocean Engineering* **2024**, *313*, 119512.
97. Ma, Y.; Liu, Q.; Yang, L. Machine learning-based multimodal fusion recognition of passenger ship seafarers' workload: A case study of a real navigation experiment. *Ocean Engineering* **2024**, *300*, 117346.
98. Mauro, F.; Vassalos, D. Time to capsize for damaged passenger ships in adverse weather conditions. A Multi-modal analysis. *Ocean Engineering* **2024**, *299*, 117409.
99. Wu, Z.; Wang, S.; Xu, H.; Shi, F.; Li, Q.; Li, L.; Qian, F. Research on ship safety risk early warning model integrating transfer learning and multi-modal learning. *Applied Ocean Research* **2024**, *150*, 104139.
100. Wu, Z.; Wang, S.; Li, L.; Wang, Y.; Wang, X. Ship near-miss risk identification method based on multi-modal deep fusion model under complex weather and sea conditions. *Ocean Engineering* **2025**, *329*, 121119.
101. Li, J.; Zhang, G.; Jiang, C.; Zhang, W. A survey of maritime unmanned search system: Theory, applications and future directions. *Ocean Engineering* **2023**, *285*, 115359.
102. Matos, A.; Silva, E.; Cruz, N.; Alves, J.C.; Almeida, D.; Pinto, M.; Martins, A.; Almeida, J.; Machado, D. Development of an Unmanned Capsule for large-scale maritime search and rescue. In Proceedings of the 2013 OCEANS-San Diego. IEEE, 2013, pp. 1–8.
103. Mansor, H.; Norhisam, M.H.; Abidin, Z.Z.; Gunawan, T.S. Autonomous surface vessel for search and rescue operation. *Bulletin of Electrical Engineering and Informatics* **2021**, *10*, 1701–1708.

104. Ponzini, F.; Van Hamme, D.; Martelli, M. Human detection in marine disaster search and rescue scenario: a multi-modal early fusion approach. *Ocean Engineering* **2025**, *340*, 122341.
105. Li, F.; Yu, K.; Yuan, C.; Tian, Y.; Yang, G.; Yin, K.; Li, Y. Dark Ship Detection via Optical and SAR Collaboration: An Improved Multi-Feature Association Method Between Remote Sensing Images and AIS Data. *Remote Sensing* **2025**, *17*, 2201.
106. Hu, Z.; Sun, Y.; Zhao, Y.; Wu, W.; Gu, Y.; Chen, K. Msif-Sstr: A Ship Smuggling Trajectory Recognition Method Based on Multi-Source Information Fusion. *Available at SSRN 5294246*.
107. Galdelli, A.; Narang, G.; Pietrini, R.; Zazzarini, M.; Fiorani, A.; Tassetti, A.N. Multimodal AI-enhanced ship detection for mapping fishing vessels and informing on suspicious activities. *Pattern Recognition Letters* **2025**, *191*, 15–22.
108. Lei, J.; Sun, Y.; Wu, Y.; Zheng, F.; He, W.; Liu, X. Association of AIS and radar data in intelligent navigation in inland waterways based on trajectory characteristics. *Journal of Marine Science and Engineering* **2024**, *12*, 890.
109. Duca, A.L.; Bacciu, C.; Marchetti, A. A K-nearest neighbor classifier for ship route prediction. In Proceedings of the OCEANS 2017 - Aberdeen, 2017, pp. 1–6. https://doi.org/10.1109/OCEANSE.2017.8084635.
110. Khaled, D.; Aly, H.; Khaled, M.; Mahmoud, N.; Shabaan, S.; Abdellatif, A. Development of a sustainable unmanned surface vehicle (USV) for search and rescue operations. In Proceedings of the the international undergraduate research conference. The Military Technical College, 2021, Vol. 5, pp. 462–468.
111. Chung, D.; Kim, J.; Lee, C.; Kim, J. Pohang canal dataset: A multimodal maritime dataset for autonomous navigation in restricted waters. *The International Journal of Robotics Research* **2023**, *42*, 1104–1114.
112. Dagdilelis, D.; Grigoriadis, P.; Galeazzi, R. Multimodal and Multiview Deep Fusion for Autonomous Marine Navigation. *arXiv preprint arXiv:2505.01615* **2025**.
113. Wright, R.G. Intelligent autonomous ship navigation using multi-sensor modalities. *TransNav: International Journal on Marine Navigation and Safety of Sea Transportation* **2019**, *13*.
114. Li, J.; Qu, C.; Shao, J. Ship detection in SAR images based on an improved faster R-CNN. In Proceedings of the 2017 SAR in Big Data Era: Models, Methods and Applications (BIGSARDATA). IEEE, 2017, pp. 1–6.
115. Zhang, T.; Zhang, X.; Li, J.; Xu, X.; Wang, B.; Zhan, X.; Xu, Y.; Ke, X.; Zeng, T.; Su, H.; et al. SAR ship detection dataset (SSDD): Official release and comprehensive data analysis. *Remote Sensing* **2021**, *13*, 3690.
116. Wang, Y.; Wang, C.; Zhang, H.; Dong, Y.; Wei, S. A SAR dataset of ship detection for deep learning under complex backgrounds. *remote sensing* **2019**, *11*, 765.
117. Xian, S.; Zhirui, W.; Yuanrui, S.; Wenhui, D.; Yue, Z.; Kun, F. AIR-SARShip-1.0: High-resolution SAR ship detection dataset. *Journal of Radars* **2019**, *8*, 852–863.
118. Li, L.; Yu, J.; Chen, F. TISD: a three bands thermal infrared dataset for all day ship detection in spaceborne imagery. *Remote Sensing* **2022**, *14*, 5297.
119. Liu, Z.; Yuan, L.; Weng, L.; Yang, Y. A high resolution optical satellite image dataset for ship recognition and some new baselines. In Proceedings of the International conference on pattern recognition applications and methods. SciTePress, 2017, Vol. 2, pp. 324–331.
120. Chen, K.; Wu, M.; Liu, J.; Zhang, C. FGSD: A dataset for fine-grained ship detection in high resolution satellite images. *arXiv preprint arXiv:2003.06832* **2020**.
121. Zhang, Z.; Zhang, L.; Wang, Y.; Feng, P.; He, R. ShipRSImageNet: A large-scale fine-grained dataset for ship detection in high-resolution optical remote sensing images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **2021**, *14*, 8458–8472.
122. Li, B.; Liu, B.; Huang, L.; Guo, W.; Zhang, Z.; Yu, W. OpenSARShip 2.0: A large-volume dataset for deeper interpretation of ship targets in Sentinel-1 imagery. In Proceedings of the 2017 SAR in Big Data Era: Models, Methods and Applications (BIGSARDATA). IEEE, 2017, pp. 1–5.
123. Zhang, M.M.; Choi, J.; Daniilidis, K.; Wolf, M.T.; Kanan, C. VAIS: A dataset for recognizing maritime imagery in the visible and infrared spectrums. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition workshops, 2015, pp. 10–16.
124. Ribeiro, R.; Cruz, G.; Matos, J.; Bernardino, A. A data set for airborne maritime surveillance environments. *IEEE Transactions on Circuits and Systems for Video Technology* **2017**, *29*, 2720–2732.
125. Choi, J.; Cho, D.; Lee, G.; Kim, H.; Yang, G.; Kim, J.; Cho, Y. Polaris dataset: A maritime object detection and tracking dataset in pohang canal. In Proceedings of the 2025 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2025, pp. 13626–13632.
126. Prasad, D.K.; Rajan, D.; Rachmawati, L.; Rajabally, E.; Quek, C. Video processing from electro-optical sensors for object detection and tracking in a maritime environment: A survey. *IEEE Transactions on Intelligent Transportation Systems* **2017**, *18*, 1993–2016.

127. Jeong, M.; Chadda, A.; Ren, Z.; Zhao, L.; Liu, H.; Zhang, A.; Jiang, Y.; Achong, S.; Lensgraf, S.; Roznere, M.; et al. SeePerSea: Multi-modal Perception Dataset of In-water Objects for Autonomous Surface Vehicles. *IEEE Transactions on Field Robotics* **2025**.
128. Yao, S.; Guan, R.; Wu, Z.; Ni, Y.; Huang, Z.; Liu, R.W.; Yue, Y.; Ding, W.; Lim, E.G.; Seo, H.; et al. Waterscenes: A multi-task 4d radar-camera fusion dataset and benchmarks for autonomous driving on water surfaces. *IEEE Transactions on Intelligent Transportation Systems* **2024**, *25*, 16584–16598.
129. Guo, Y.; Liu, R.W.; Qu, J.; Lu, Y.; Zhu, F.; Lv, Y. Asynchronous trajectory matching-based multimodal maritime data fusion for vessel traffic surveillance in inland waterways. *IEEE Transactions on Intelligent Transportation Systems* **2023**, *24*, 12779–12792.
130. Varga, L.A.; Kiefer, B.; Messmer, M.; Zell, A. Seadronessee: A maritime benchmark for detecting humans in open water. In Proceedings of the Proceedings of the IEEE/CVF winter conference on applications of computer vision, 2022, pp. 2260–2270.
131. Jang, H.; Yang, W.; Kim, H.; Lee, D.; Kim, Y.; Park, J.; Jeon, M.; Koh, J.; Kang, Y.; Jung, M.; et al. MOANA: Multi-radar dataset for maritime odometry and autonomous navigation application. *The International Journal of Robotics Research* **2025**, p. 02783649251354897.
132. Helgesen, Ø.K.; Vasstein, K.; Brekke, E.F.; Stahl, A. Heterogeneous multi-sensor tracking for an autonomous surface vehicle in a littoral environment. *Ocean Engineering* **2022**, *252*, 111168.
133. Lu, H.; Zhang, Y.; Zhang, C.; Niu, Y.; Wang, Z.; Zhang, H. A multi-sensor fusion approach for maritime autonomous surface ships berthing navigation perception. *Ocean Engineering* **2025**, *316*, 119965.
134. Mouawad, I.; Brasch, N.; Manhardt, F.; Tombari, F.; Odone, F. Time-to-label: Temporal consistency for self-supervised monocular 3D object detection. *IEEE Robotics and Automation Letters* **2022**, *7*, 8988–8995.
135. Luo, C.; Wu, F.; Xiong, R.; Xu, G.; Liu, W. Edge Computing-Enabled Lightweight Deep Neural Network for Real-Time Video Surveillance in Maritime Cyber-Physical Systems. In Proceedings of the 2024 9th International Conference on Intelligent Computing and Signal Processing (ICSP). IEEE, 2024, pp. 1342–1350.
136. Alqurashi, F.S.; Trichili, A.; Saeed, N.; Ooi, B.S.; Alouini, M.S. Maritime communications: A survey on enabling technologies, opportunities, and challenges. *IEEE Internet of Things Journal* **2022**, *10*, 3525–3547.
137. Chen, H.; Wen, Y.; Huang, Y.; Xiao, C.; Sui, Z. Edge computing enabling internet of ships: A survey on architectures, emerging applications, and challenges. *IEEE Internet of Things Journal* **2024**.
138. Liu, R.W.; Guo, Y.; Nie, J.; Hu, Q.; Xiong, Z.; Yu, H.; Guizani, M. Intelligent edge-enabled efficient multi-source data fusion for autonomous surface vehicles in maritime internet of things. *IEEE Transactions on Green Communications and Networking* **2022**, *6*, 1574–1587.
139. Jurdana, I.; Lopac, N.; Wakabayashi, N.; Liu, H. Shipboard data compression method for sustainable real-time maritime communication in remote voyage monitoring of autonomous ships. *Sustainability* **2021**, *13*, 8264.
140. Akpan, F.; Bendiab, G.; Shiaeles, S.; Karamperidis, S.; Michaloliakos, M. Cybersecurity challenges in the maritime sector. *Network* **2022**, *2*, 123–138.
141. Bothur, D.; Zheng, G.; Valli, C. A critical analysis of security vulnerabilities and countermeasures in a smart ship system **2017**.
142. Goudosis, A.; Katsikas, S.K. Secure ais with identity-based authentication and encryption. *TransNav: International Journal on Marine Navigation and Safety of Sea Transportation* **2020**, *14*, 287–298.
143. Gai, K.; Tang, H.; Li, G.; Xie, T.; Wang, S.; Zhu, L.; Choo, K.K.R. Blockchain-based privacy-preserving positioning data sharing for IoT-enabled maritime transportation systems. *IEEE Transactions on Intelligent Transportation Systems* **2022**, *24*, 2344–2358.
144. Wang, S.; Zhen, L.; Xiao, L.; Attard, M. Data-driven intelligent port management based on blockchain. *Asia-Pacific Journal of Operational Research* **2021**, *38*, 2040017.
145. Tabish, N.; Chaur-Luh, T. Maritime autonomous surface ships: A review of cybersecurity challenges, countermeasures, and future perspectives. *IEEe Access* **2024**, *12*, 17114–17136.
146. Melnyk, O.; Drozdov, O.; Kuznichenko, S. Cybersecurity in Maritime Transport: An International Perspective on Regulatory Frameworks and Countermeasures. *Lex Portus* **2025**, *11*, 7.
147. Lielbārde, S.; Brilingaitė, A.; Bukauskas, L.; Roponena, E.; Citskovska, E.; Pirta, R. Maritime Cyber Resilience: Bridging Cybersecurity and Regulatory Frameworks. In Proceedings of the European Interdisciplinary Cybersecurity Conference. Springer, 2025, pp. 217–228.
148. Polikarovskykh, O.; Malaksiano, M.; Piterska, V.; Daus, Y.; Tkachenko, M. Measures to Counter Cyber Attacks on Maritime Transportation. In *Maritime Systems, Transport and Logistics I: Safety and Efficiency of Operation*; Springer, 2025; pp. 197–212.

doi:10.20944/preprints202510.0977.v1

149. Glomsrud, J.A.; Ødegårdstuen, A.; Clair, A.L.S.; Smogeli, Ø. Trustworthy versus explainable AI in autonomous vessels. In Proceedings of the Proceedings of the International Seminar on Safety and Security of Autonomous Vessels (ISSAV) and European STAMP Workshop and Conference (ESWC), 2019, Vol. 37.
150. Veitch, E.; Alsos, O.A. Human-centered explainable artificial intelligence for marine autonomous surface vehicles. *Journal of Marine Science and Engineering* **2021**, *9*, 1227.
151. Prasad, D.K.; Prasath, C.K.; Rajan, D.; Rachmawati, L.; Rajabally, E.; Quek, C. Maritime situational awareness using adaptive multi-sensor management under hazy conditions. *arXiv preprint arXiv:1702.00754* **2017**.
152. Brown, K.E.; Talbert, D.A. Using explainable ai to measure feature contribution to uncertainty. In Proceedings of the The international FLAIRS conference proceedings, 2022, Vol. 35.
153. Nazir, M.A.; Evangelista, E.; Bukhari, S.M.S.; Sharma, R. A survey of feature attribution techniques in explainable AI: taxonomy, analysis and comparison. *Annals of Mathematics and Computer Science* **2025**, *28*, 115–126.
154. Madsen, A.N. Decision Transparency during Autonomous Collision Avoidance: On Human-AI Compatibility and Autonomous Ships **2024**.
155. Taner, T.; Cengiz, Z.S.; Ünal, H.T.; Mendi, A.F.; Nacar, M.A. AR-Based Hybrid Human-AI Decision Support System for Maritime Navigation. In Proceedings of the 2025 3rd Cognitive Models and Artificial Intelligence Conference (AICCONF). IEEE, 2025, pp. 1–6.

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.