Faculty of International Maritime Studies
Kasetsart University

Maritime Technology and Research

https://so04.tci-thaijo.org/index.php/MTR

Maritime Technology and Research

Review Article

Maritime vision datasets for autonomous navigation: A comparative analysis

Nico Jungbauer[1,*], **Hai Huang**[2] **and Helmut Mayer**[2]

[1]TKMS ATLAS ELEKTRONIK GmbH, Sebaldsbrücker Heerstraße 235, Bremen 28309, Germany
[2]Institute for Applied Computer Science, Bundeswehr University Munich, Werner-Heisenberg-Weg 39, Neubiberg 85579, Germany

[*]Corresponding author's e-mail address: nico.jungbauer@tkmsgroup.com

Article information
Received: February 21, 2025
Revision: May 15, 2025
Accepted: June 1, 2025

Keywords
Maritime domain;
Datasets;
Surface vessels;
Computer vision;
Deep learning;
Object detection

Abstract

Artificial intelligence is becoming an increasingly essential component in many areas, with notable advancements being made in the field of maritime computer vision. The employed deep learning models require substantial quantities of high-quality training data that are specifically tailored to the tasks for which they are being applied in the maritime domain. Training autonomous navigation systems for unmanned surface vehicles has been significantly enhanced by using extensive visual datasets captured through high-quality cameras, enabling these systems to learn from diverse environmental scenarios and improve the decision-making accuracy. However, the identification of suitable publicly accessible maritime vision datasets is challenging, and there is currently no broad overview of datasets that have been specifically designed for computer vision tasks related to unmanned surface vehicles in the maritime domain. This survey addresses the identified research gap by providing a comprehensive and systematic overview of open-source vision datasets containing ships, taking into account the specific task, the surrounding environment, and additional available data, such as infrared images or time series information. It is our aim to assist new researchers in the field of maritime computer vision to gain a rapid overview and facilitate initial access to this domain, enabling them to identify the most suitable dataset for their particular task.

1. Introduction

In the maritime industry, the application of artificial intelligence (AI) to autonomous systems promises significant advancements, potentially resulting in more robust and capable self-operating systems (Abdelsalam & Elnabawi, 2024). One task for these systems is fully autonomous shipping, which will require unmanned surface vehicles (USVs), according to the International Maritime Organization (IMO) (International Maritime Organisation, 2024). A key challenge in reaching the highest IMO automation level is autonomous navigation, which may rely on computer vision-based object detection for collision avoidance, as visual images provide valuable and complementary information compared to, e.g., LiDAR data (Wang, 2021). Maritime environments present unique challenges for object detection due to their dynamic and often unpredictable nature. Applications such as maritime surveillance, safety monitoring (Gribbestad et al., 2021), illegal fishing detection, and environmental protection rely heavily on accurate and real-time identification of vessels and other maritime objects. Effective optical object detection in these environments requires robust methods that can handle different conditions, including different lighting (day, night), weather conditions (fog, rain, etc.), and complex backgrounds such as waves and reflections. The desire for both accuracy and low reaction time in obstacle detection systems has led to the use of sophisticated deep learning architectures such as convolutional neural networks (CNNs) and vision transformers (Dosovitskiy et al., 2020). However, for these advanced models to reach their full potential, high-quality training datasets with labeled images are essential, yet often neglected (Sambasivan et al., 2021). The availability of large-scale, generic computer vision datasets such as ImageNet (Deng et al., 2009), COCO (Lin et al., 2014), and PASCAL VOC (Everingham et al., 2015) has significantly accelerated deep learning research. However, domain-specific datasets are essential to maximize performance in order to obtain safe and reliable AI. Unfortunately, the availability of comprehensive and diverse open-source maritime image datasets is rather limited. Existing datasets vary widely in terms of size, image quality, image diversity, labeling accuracy, and the range of scenarios they cover. This diversity makes it challenging for researchers and practitioners to select the most suitable dataset for their specific application needs. To the best of the authors' knowledge, there is currently only one survey comparing maritime vision datasets since 2015 (Su et al., 2023). Although this survey includes many of the relevant publicly available datasets up to 2023, it also comprises datasets with aerial images and omits several important datasets, which are discussed in Section 2. In addition, several new datasets have been published since its release, making a new systematic literature survey highly valuable. Accordingly, a comprehensive overview of maritime vision datasets is presented, analyzing their characteristics, strengths, and limitations with a particular focus on their usability for autonomous navigation. In this context, navigation refers to the task of guiding a vessel from a starting point to a destination, with obstacle avoidance as a key subtask to ensure safe transit. The contributions of this paper can be summarized as follows:

- Providing a holistic overview of maritime vision datasets for USVs from 2015 to October 2024.
- Giving recommendations for datasets specific to USV navigation tasks, significantly simplifying training dataset selection.

The remainder of this survey is structured as follows. In Section 2, an in-depth analysis of the publicly available maritime vision datasets is performed. Section 3 compares and discusses the datasets presented in the previous section. Section 4 then concludes the work by highlighting current gaps in maritime vision datasets and outlining directions for future research.

2. Analysis of maritime vision datasets

2.1 Criteria for dataset selection

The main goal of this paper is to select criteria that are suitable for providing a structured overview of open-source maritime vision datasets relevant to autonomous navigation. Therefore, only annotated datasets comprising RGB (Red, Green, Blue) images are considered, excluding a dataset containing only infrared images (Nirgudkar et al., 2023). Furthermore, bird's-eye view images were not taken into account, as they cannot be used on-board for navigation. Due to the significant improvement in image quality over the years, datasets published before 2015, as well as purely synthetic datasets, are excluded. The reason for this is that images generated by, e.g., generative adversarial networks (GANs) or diffusion models usually do not match real images sufficiently (Bird & Lotfi, 2024). To ensure reasonable generalization, datasets with fewer than 1,000 images, fewer than 20 videos, or object-specific datasets containing less than 50 % images of water vehicles, e.g., images of sea buoys (Liu et al., 2020), swimmers (Khan et al., 2024) or floating waste (Cheng et al., 2021), are excluded. Only datasets with previously published papers are taken into account. The search is not narrowed by applying thresholds on image resolution, number of annotations if known, accuracy of annotations, or the percentage of real and synthetic data within a single dataset, as long as it contains real images. Annotations must be manually verified and contain at least one of the following: bounding box, pixel-wise segmentation (either semantic, instance, or panoptic), horizon estimation, or more than one class, with one class being ship. The ship requirement is especially important, as the failure to detect and avoid obstacles can lead to severe consequences when high monetary-value moving objects are overlooked, posing potential risks to human life.

2.2 Criteria for dataset analysis

The analysis of maritime vision datasets is essential for determining their suitability for various object detection and scene understanding tasks. Each dataset has distinct characteristics regarding size, image resolution, scene diversity, and the type and quality of annotations. To facilitate a comprehensive comparison, the following criteria will be used: dataset size, number and usefulness of classes, annotation quality, image quality and resolution, scene diversity, scenario coverage, and temporal and dynamic content. In autonomous shipping, the usefulness of classes refers to how relevant the dataset's classes are for real-world object detection and navigation, ensuring that models can effectively handle key objects and scenarios in maritime environments.

Dataset size and class diversity are used to assess the generalization capability to real-world scenarios. Accordingly, the total number of images and the diversity of object classes (e.g., various types of vessels, buoys, and marine wildlife) within the dataset are taken into account. Larger datasets with a wide range of object classes are generally more beneficial for training deep learning models. Nevertheless, the issue of class balance must also be taken into account in order to prevent the introduction of bias and to meet the requirements of the specific application tasks (Johnson & Khoshgoftaar, 2019).

Image quality is a crucial aspect for assessing the overall usability of images. This encompasses a range of factors, including resolution, sharpness, signal-to-noise ratio, and dynamic range, which can have a significant impact on the performance of computer vision models. Higher-resolution images are preferred as they provide more detailed features. Additionally, newer datasets are preferred since they often contain images captured by newer cameras with larger sensors, resulting in better signal-to-noise ratios. Newer cameras also usually offer higher dynamic range, improved noise suppression, and better optical stabilization, allowing for longer exposure times.

Scene diversity refers to the range of environmental conditions represented in the dataset, such as variations in weather (clear sky, fog, rain), lighting conditions (day, dawn, night), sea states (calm, disturbed, rough waters), and types of environment (river-like, lake-like). Diversity leads to a variety of image characteristics, including differing levels of reflection, under- or over-exposure, sun glare, low visibility, the presence of vegetation, and the possibility of stained lenses. Scene diversity is crucial for assessing model robustness in real-world conditions.

Scenario coverage provides a detailed assessment of how well the dataset represents different maritime scenarios, such as maneuvering, open sea, coastal, or port traffic. Particular attention is given to challenging scenarios, with occlusions or cluttered backgrounds, where objects occlude or overlap each other, complicating detection, classification, and tracking.

Annotation quality refers to the accuracy, consistency, and comprehensiveness of annotations, which may consist of bounding boxes, segmentation masks, and class labels. The quality of annotations is generally expected to be highest when one or more experts with domain-specific knowledge are involved, followed by external labeling companies, which are regarded as specialists in general annotation tasks. Finally, volunteers are assumed to produce the lowest annotation quality. Attention also needs to be given to the level of granularity (e.g., semantic segmentation vs. instance segmentation vs. panoptic segmentation) and the availability of additional data, such as infrared images, AIS data, location information, and temporal annotations. In addition, the source of the annotations, whether from experts or volunteers, will be noted. If annotations are performed by external companies specializing in computer vision annotation, they are considered expert annotations.

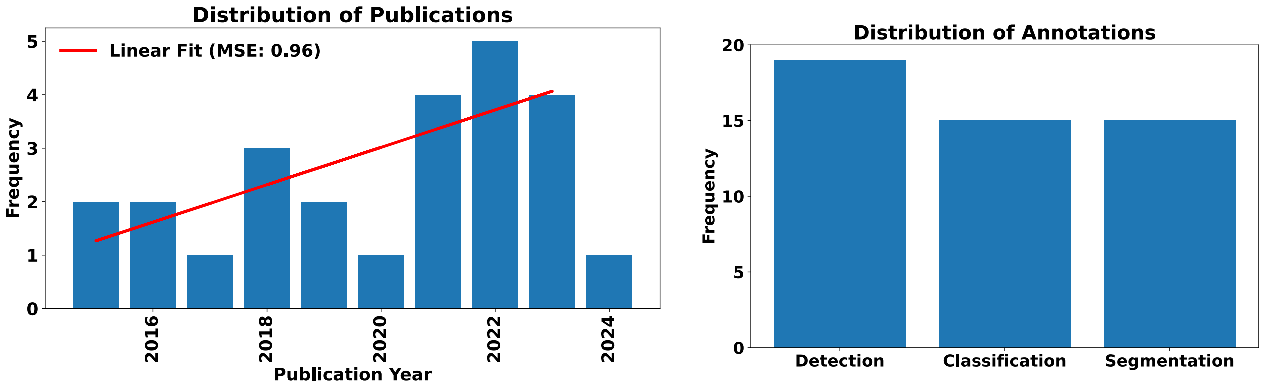
Temporal and dynamic content (if applicable) will be analyzed for datasets containing video sequences or temporal data. This assessment will consider the consistency of annotations across frames and the dataset's suitability for motion-based tasks, such as tracking and activity recognition.

The following assessment will highlight each dataset's strengths and limitations, providing recommendations for their most effective use in various maritime computer vision tasks such as vessel detection, sea horizon estimation, segmentation, and classification.

2.3 Overview of relevant datasets

Table 1 lists the 25 evaluated datasets that meet the criteria defined in Section 2.1, which form the foundation of this survey. **Figure 1(a)** illustrates the distribution of these datasets with respect to publication year. It can be observed that there is at least one publication associated with a dataset each year, with up to five datasets appearing per year. A linear fit was performed for the years 2015 - 2023, excluding the ongoing year 2024. The positive slope is 0.35 a^{-1} ±0.14 a^{-1}, suggesting an upward trend, though the high uncertainty prevents a definitive conclusion. Furthermore, **Figure 1(b)** shows that detection annotations (21 datasets) are the most common, followed by classification (16 datasets) and segmentation (11 datasets). All 25 datasets are categorized into six categories based on the annotations, which are also referred to as the “tasks”:

1. Detection (2 datasets)
2. Classification (1 datasets)
3. Segmentation (3 datasets)
4. Detection and Classification (11 datasets)
5. Detection and Segmentation (4 datasets)
6. Detection, Classification and Segmentation (4 datasets)



(a) (b)

Figure 1 (a) Distribution of publication years with a linear fit up to 2023, indicating a positive trend; (b) Types of annotations available in the datasets included in this survey. Annotations comprise bounding boxes for detection, class labels, and segmentation labels. Datasets containing multiple annotation types are counted multiple times.

Additionally, **Figure 2** displays the number of citations for each dataset and the publication year of the associated paper. Citation data were obtained from Google Scholar on October 2, 2024. This helps to identify which datasets are most frequently referenced by researchers. Some datasets are easier to find because they are explicitly mentioned in the publication title and offer straightforward download access, which may also affect citation counts. Another influencing factor is that the majority of papers not only present the dataset but also propose new maritime computer vision techniques that may attract researchers’ primary interest. The total number of citations tends to decrease (slope: −22.9 citations a^{-1} ± 6.3 citations a^{-1}) for more recent publications. This is expected, given that newer publications have had less time to accumulate citations. However, two datasets stand out, with citations surpassing other publications within a ±2-year range: the SMD and the newer SMD-plus dataset. These are the most cited datasets for periods 2015 - 2022 and 2020 - 2024, respectively, indicating broad interest in these datasets.

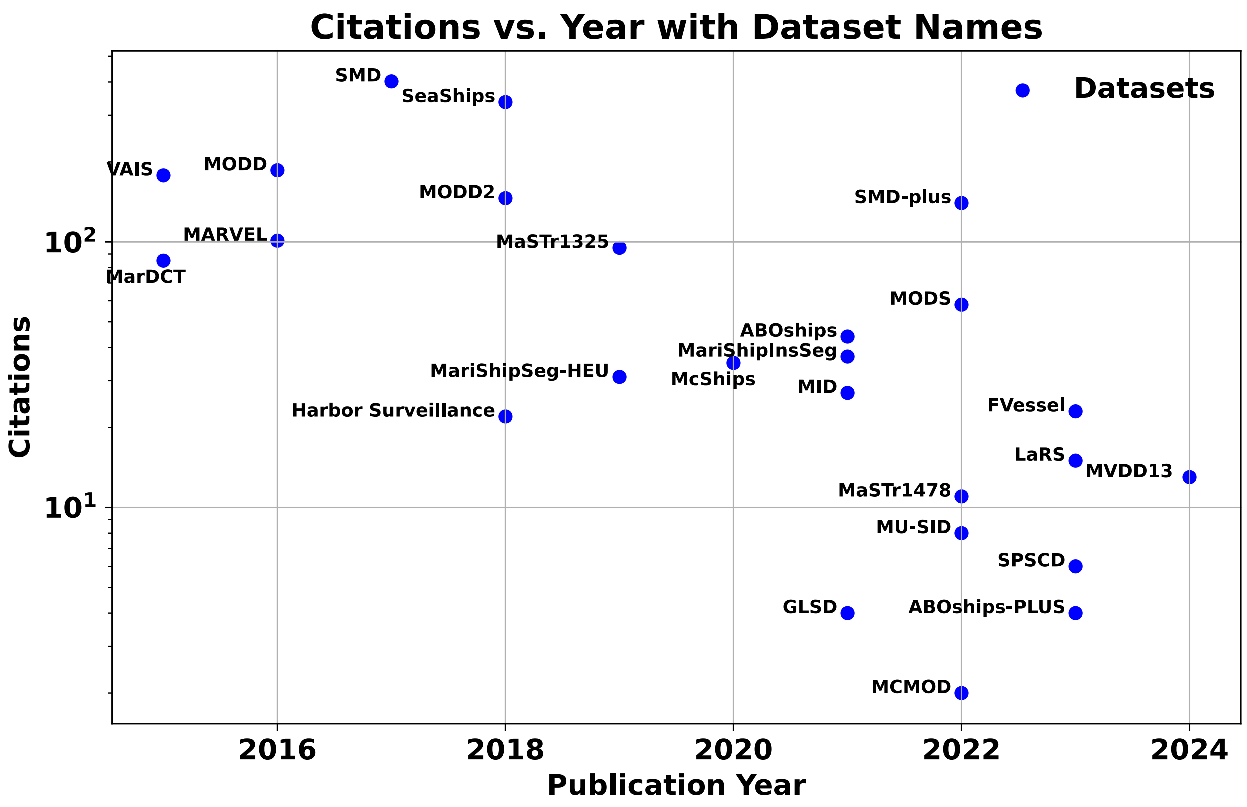


Figure 2 Year of publication and corresponding citation counts from Google Scholar.

Figure 3 shows the number of images contained in each dataset along with the corresponding class counts. The MARVEL dataset stands out, since it has nearly an order of magnitude more images and labels than any other dataset. On the other hand, MaSTr1325 and MaSTr1478 have relatively few images, largely due to their pixel-wise semantic annotations. GLSD also contains a large number of images and an above-median class count, though it is cited less frequently, potentially due to its method of creation, which includes simulated and painted images.

2.3.1 Detection

Bounding boxes around vessels are the most common annotation type in the selected datasets (**Figure 1**). Papers are listed in this section if they exclusively contain detection-related annotations, meaning they do not include any classes or segmentation masks, and thus do not fall into one of the other five categories specified above.

Only two datasets fall into this category. One is the Harbor Surveillance dataset (Zwemer et al., 2018), a novel maritime resource designed to enhance vessel detection in real-world surveillance scenarios. It contains 48,966 images at a resolution of 1,536×2,048 pixels, captured from ten different camera viewpoints over 73 days, spanning six months. The dataset was automatically generated by selecting frames from video recordings, followed by manual post-processing to ensure that each image contains at least one ship, with ships then manually annotated. In total, 70,513 ships are labeled in multiple orientations and under varying background conditions, covering different angles and scenes. While the dataset includes a wide range of ship types, it has some limitations in terms of the variety of viewpoints and scenes. In addition, many ships are only partially visible due to camera cut-offs, increasing the difficulty of vessel detection and making the dataset less relevant for real-world scenarios. The authors emphasize the need to expand scene diversity in future iterations to enhance its effectiveness for maritime object detection in surveillance applications.

Table 1 provides a comparison of different optical maritime vision datasets. It includes name, release year, image and video types, sensor information, resolution, tasks (D: detection, C: classification, S: segmentation), number of classes, and scene diversity. When datasets combine different types of cameras, other than visual cameras, the names and numbers refer to the annotated visual images only. The year column indicates the publication date of the corresponding paper. The classes column shows the number of classes from the object classification task. If object classification annotations are not present, it reflects the number of classes used in semantic segmentation, which may include "thing" categories (discrete objects) and "stuff" categories (continuous regions or materials without distinct boundaries). The instances column gives the number of annotated objects present in all images combined, except for the FVessel dataset, where the total number of ships in all videos is given. Scene diversity is categorized into three levels, indicating the geographic spread of data acquisition: datasets recorded in a single contiguous area (e.g., one lake or the vicinity of one city) are labeled with a minus sign (–), those captured across two to four distinct locations are marked as neutral (0), and datasets collected from more than four geographically separate areas are denoted with a plus sign (+).

Table 1 Comparison of different optical maritime vision datasets.

Name	Year	Images	Videos	Sensor	Resolution	Task	Classes	Instances	Scene Diversity
VAIS (Zhang et al., 2015)	2015	2,865	0	ISVI IC-C25	5,056×5,056	D, C	6 /15	N/A	+
MarDCT (Bloisi et al., 2015)	2015	6,743	0	multiple	240×800	D, C, S	24	N/A	-
MARVEL (Gundogdu et al., 2017)	2016	1,607,190	0	multiple	multiple	C	197	1,607,190	+
MODD (Kristan et al., 2015)	2016	4,454	12	AXIS 207W	480×640	D, S	1	N/A	-
SMD (Prasad et al., 2017)	2017	20,367	36	Canon 70D	1,080×1,920	D, C	10	192,980	0
Harbor Surveillance (Zwemer et al., 2018)	2018	48,966	0	N/A	1,536×2,048	D	1	70,513	-
MODD2 (Bovcon et al., 2018)	2018	11,675	28	Vrmagic VRmS-14/C-COB	1,278×958	D, S	1	N/A	-
SeaShips (Shao et al., 2018)	2018	31,455	0	multiple	1,080×1,920	D, C	6	40,077	0
MaSTr1325 (Bovcon et al., 2019)	2019	1,325	0	Vrmagic VRmS-14/C-COB	958×1,278	S	3	N/A	-
MariShipSeg-HEU (Zhang et al., 2020)	2020	3,560	0	multiple	N/A	D, C, S	N/A	3,560	+
Mcships (Zheng & Zhang, 2020)	2020	14,709	0	multiple	multiple	D, C	13	26,529	+
GLSD (Shao et al., 2024)	2021	152,576	0	multiple	multiple	D, C	13	212,357	+
ABOships (Iancu et al., 2021)	2021	9,880	0	N/A	720×1,280	D, C	9	41,967	-
MID (Liu et al., 2021)	2021	2,655	8	N/A	480×640	D, S	1	~9,800	0
MariShipInsSeg (Sun et al., 2022)	2022	4,001	0	multiple	N/A	D, S	1	8,413	+
MCMOD (Sun et al., 2023)	2022	16,166	0	N/A	1,080×1,920	D, C	10	98,590	0
MODS (Bovcon et al., 2021)	2022	24,090	0	multiple	N/A	D, C, S	3	145,334	0
MU-SID (Hashmani & Umair, 2022)	2022	2,673	0	Nikon D3400	1,080×1,920	S	2	2,673	+
SMD-plus (Kim et al., 2022)	2022	20,367	36	Canon 70D	1,080×1,920	D, C	7	177,698	0
MaSTr1478 (Žust & Kristan, 2022)	2022	1,478	0	Vrmagic VRmS-14/C-COB	958×1,278	S	3	N/A	-
ABOships-PLUS (Iancu et al., 2023)	2023	15,838	0	N/A	720×1,280	D, C	4	33,227	-
SPSCD (Petkovic et al., 2023)	2023	19,337	0	Dahua DH-TPC - PT8620A-T	1,080×1,920	D, C	12	27,849	-
FVessel (Guo et al., 2023)	2023	7,625	26	HIKVISION DS-2DC4423IW-D	N/A	D	1	107	-
LaRS (Žust et al., 2023)	2023	4,006	0	multiple	multiple	D, C, S	8	N/A	+
MVDD13 (Wang et al., 2024)	2024	35,474	0	multiple	multiple	D, C	13	N/A	+

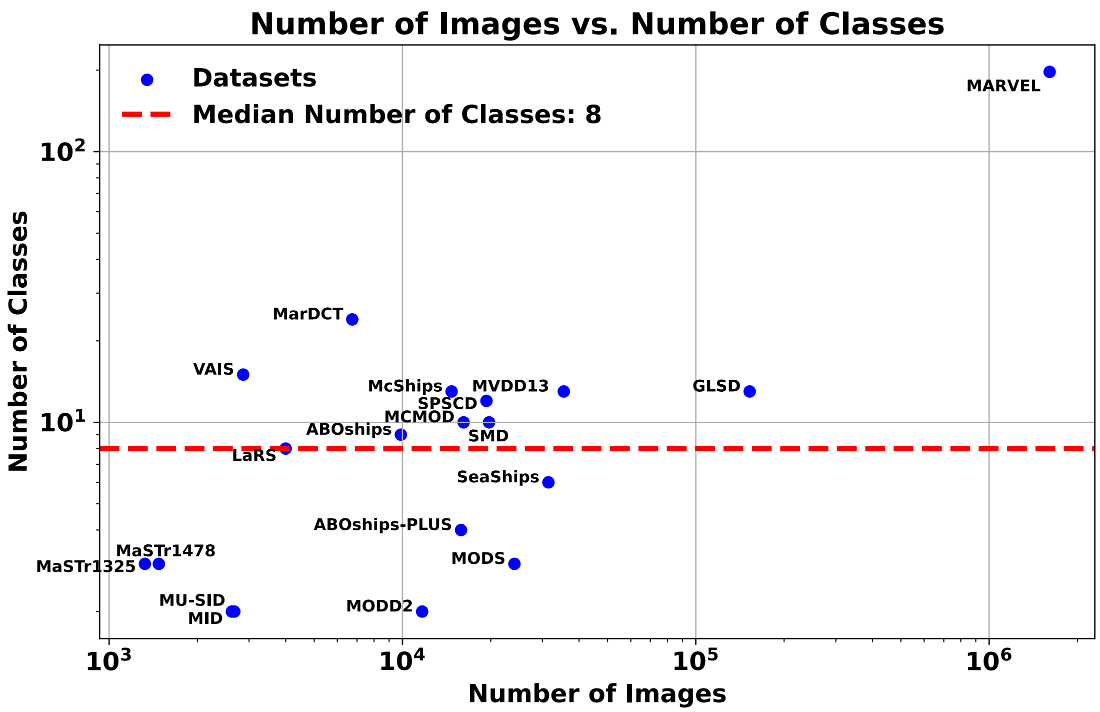


Figure 3 Number of images and classes for each dataset. For segmentation-only datasets, the number of semantic classes is given. Otherwise, the number of object classes is stated. Datasets with only one class (e.g., detection) or lacking specific information on the number of images or classes are excluded from this plot.

The FVessel dataset (Guo et al., 2023) is designed as a benchmark for vessel detection, tracking, and data fusion tasks. Two versions of the dataset are available: version one, comprising 26 videos, can be directly downloaded, while version two, with nine videos, is available to researchers on request. In this discussion, version one is focused on. It consists of 7,625 images (**Figure 4**) derived from 26 RGB videos, each paired with corresponding AIS data. This data was collected under a variety of weather conditions along the Wuhan Segment of the Yangtze River. Data acquisition was carried out using a HIKVISION DS-2DC4423IW-D dome camera and a Saiyang AIS9000-08 Class-B AIS receiver, capturing diverse locations such as bridge and riversides. The dataset reflects a range of maritime conditions, including sunny, cloudy, and low-light settings, providing a realistic set of challenges for maritime surveillance. With meticulous annotation, the dataset contains 107 vessels in total to ensure precise object tracking. These annotations, in the form of bounding boxes, cover all 26 videos, recorded at one-second intervals, incorporating various vessel types and occlusion scenarios. This multi-source dataset is a valuable resource for developing and testing advanced vessel tracking and data fusion algorithms.

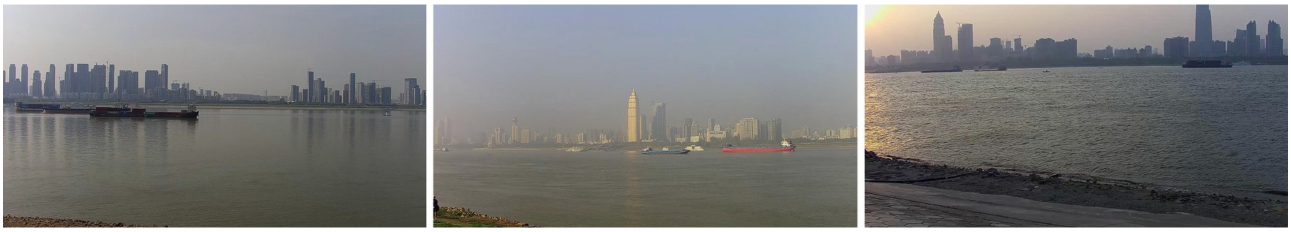
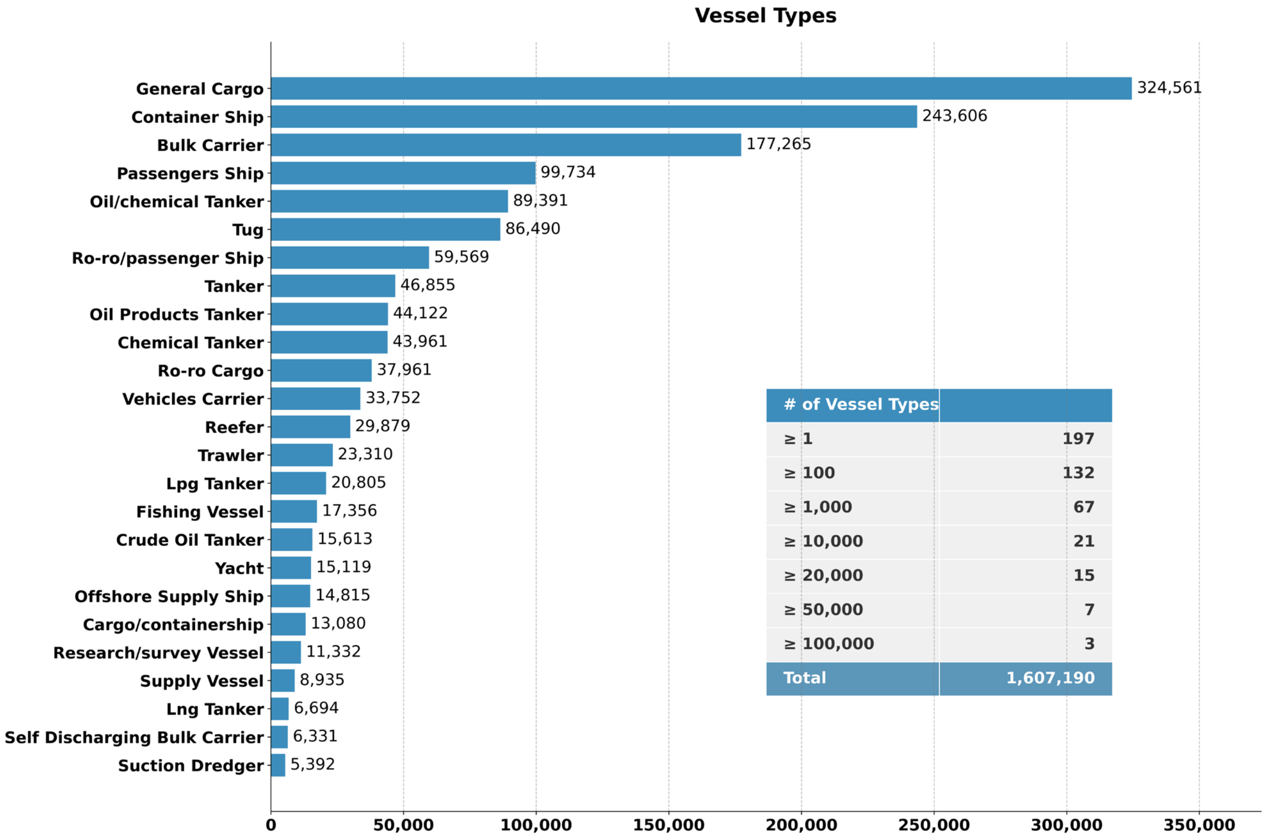


Figure 4 Example images from the FVessel dataset taken from Guo et al. (2023).

2.3.2 Classification

Classification implies that at least two different vessel types are present in the dataset. It is the second most common annotation type among the selected datasets (**Figure 1**), without bounding boxes or segmentation labels. Datasets are listed in this section if they do not fall into the detection section or any of the other four remaining categories specified above. Only one dataset falls into this category.

In particular, the MARVEL dataset (Gundogdu et al., 2017) is a large dataset of maritime vessel images consisting of 1,607,190 images of various resolutions taken by different cameras and uploaded to the Shipspotting website by hobby photographers. It is the largest publicly available maritime dataset and contains the most diverse set of vessel categories. The dataset includes 197 vessel types, with 103,701 unique IMO numbers and over 8,000 unique vessels. The number of unique vessels differs from the number of unique IMO numbers because ships identical in construction can carry different IMO numbers. To address classification challenges, the MARVEL dataset uses a semi-supervised clustering scheme to merge some vessel types, reducing the 197 categories into 26 superclasses, as shown in **Figure 5**. In total, 1,190,169 images are available for superclass classification. The dataset provides detailed annotations for most images, including attributes such as the year it was built, beam (the width of a ship at its widest point), draught (the distance from the bottom of the hull to the waterline), gross tonnage (an index based on a ship's internal volume), and summer deadweight tonnage (a measure of the ship's carrying capacity). The Maritime Mobile Service Identity (MMSI) is also included, which uniquely identifies ship stations. Superclasses, along with the vast number of images, makes the MARVEL dataset an essential resource for a variety of tasks, including vessel classification, verification, retrieval, and recognition. However, one drawback is the varying image resolution. The dataset's primary focus is on image classification, as the images typically feature vessels under ideal conditions such as close-up views, simple backgrounds, and clear weather.



# of Vessel Types	
≥ 1	197
≥ 100	132
≥ 1,000	67
≥ 10,000	21
≥ 20,000	15
≥ 50,000	7
≥ 100,000	3
Total	1,607,190

Figure 5 Instances of different vessel types in the MARVEL dataset (Gundogdu et al., 2017).

2.3.3 Segmentation

Segmentation is the least common annotation type among the selected datasets (**Figure 1**), meaning that semantic segmentation labels are present. Horizon annotation lines are subsumed under semantic segmentation, as they divide the image into three meaningful regions (below horizon, horizon, and above horizon). Datasets in this section do not comprise classification labels or bounding boxes. There are no instance or panoptic segmentation datasets in this section because, if the corresponding segmentation labels are present, bounding boxes can be drawn.

The MaSTr1325 dataset (Bovcon et al., 2019) includes 1,325 images with a resolution of 958×1,278 pixels, created for training deep learning models on semantic segmentation tasks for USVs. Captured over 24 months in the coastal waters of Koper, Slovenia, with a Vrmagic VRmS-14/C-COB CCD camera mounted 0.7 meters above water on a real USV, it provides a 132.1-degree field of view. Images were captured under various weather conditions and at different times of the day to enhance diversity. Each image is manually labeled at the pixel level with three semantic categories: sea, sky, and environment, and the annotations are reviewed by experts to ensure quality. This dataset is particularly valuable for developing USV obstacle detection techniques using semantic segmentation.

The MaSTr1478 dataset (Žust & Kristan, 2022) extends the MaSTr1325 dataset with 153 additional images (**Figure 6**) and preceding frames for enhanced temporal context in maritime obstacle detection. The additional images, captured by the same camera system, focus on challenging scenarios, such as reflections and sun glitter, which often confuse single-frame object detection models (Žust & Kristan, 2022). The authors suggest that, by incorporating temporal context, this dataset is crucial for advancing maritime perception systems for USVs.

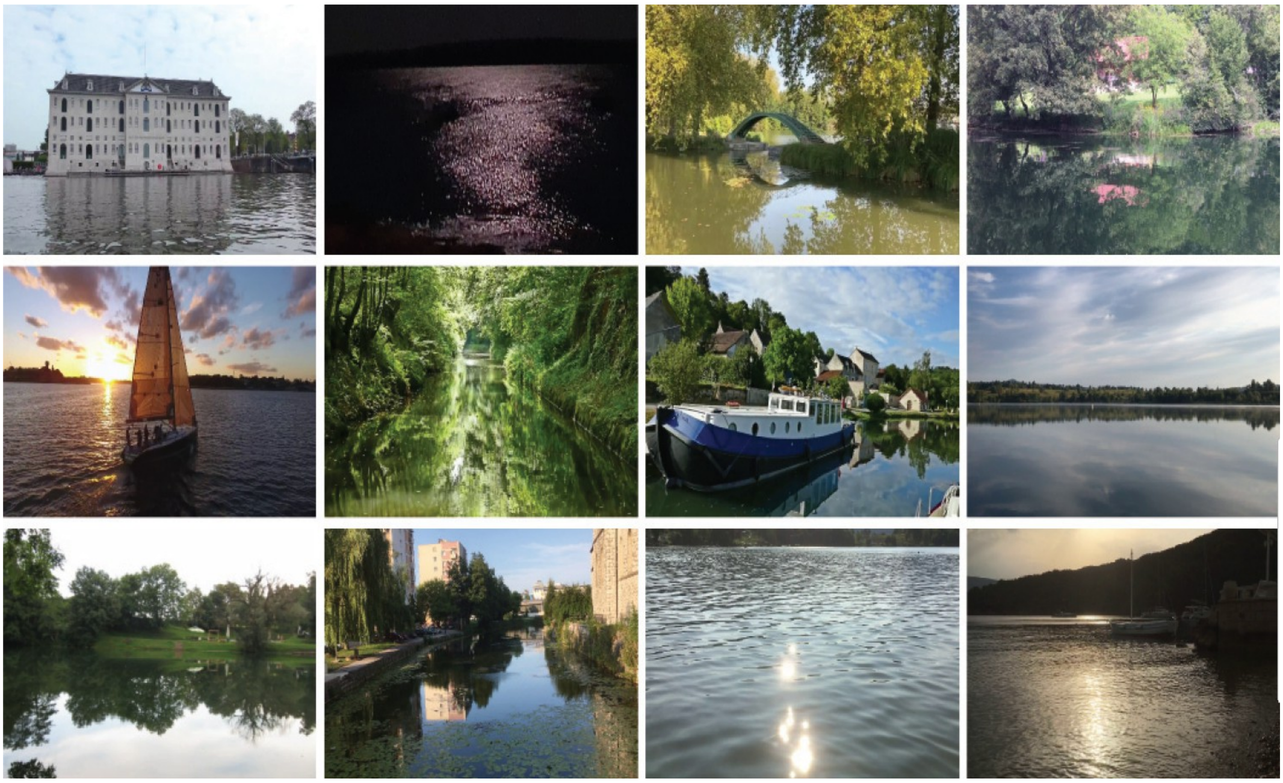


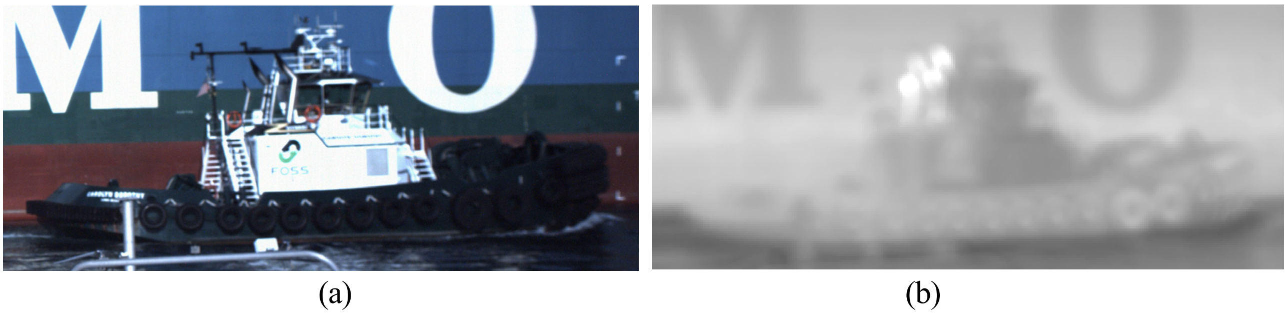
Figure 6 Example images in MaSTr1478 that were added to MaSTr1325. They show sun glitter, object reflections, and low-light conditions, and are taken from Žust & Kristan (2022).

The Manzoor-Umair Sea Image Dataset (MU-SID) (Hashmani & Umair, 2022) is designed for sea horizon line (SHL) detection and consists of 2,673 high-definition (1,080×1,920 pixels) RGB images, captured with a Nikon D3400. It includes diverse maritime scenes from five distinct locations in West Malaysia, recorded over 14 months (February 2020 to April 2021). The dataset incorporates 36 geographical, seasonal, and maritime conditions, spanning locations in the Straits of Malacca and the South China Sea, with varied sea states, environmental conditions, and times of day. Each image is annotated with SHL and supported by weather station data on wind and sea state conditions, making MU-SID unique by including weather information. This dataset features a wide range of SHL angles, from calm seas to rough waves, as well as challenging conditions such as glare, fog, and partial occlusion. As such, it is an essential resource for the further development of SHL detection methods in a maritime context.

2.3.4 Detection and classification

Detection and classification are addressed by eleven out of the 25 datasets. Papers are listed in this section if they provide some form of detection along with a classification label, without segmentation masks.

The VAIS dataset (Zhang et al., 2015) is a comprehensive maritime dataset containing 1,623 visible-spectrum (RGB) images and 1,242 infrared (IR) images (**Figure 7**). Among these, 1,088 paired images include both visible and infrared data, while an additional 154 infrared images capture nighttime scenes. The dataset focuses on ship type classification with bounding box annotations, and each ship instance is manually labeled. The RGB images are of low resolution, with the majority having a resolution of less than 224×224 pixels. These images were taken with an ISVI IC-C25 camera and the infrared images with a Sofradir-EC Atom 1,024 camera. Collected over nine days, with daily image capture sessions between 10 and 15 hours across six different piers, VAIS includes a range of environmental and lighting conditions. The dataset includes various ship types, such as 26 cargo ships and 9 barges, i.e., merchant vessels, 41 sailing ships with sails up and 24 with sails down, 11 ferries, and 4 tour boats classified as medium passenger vessels, as well as 8 fishing boats along with 14 other medium-sized ships. Additionally, there are 19 tugboats, as well as 36 small boats comprising 28 speedboats, 6 jetskis, 25 small pleasure boats, and 13 large pleasure boats. Bounding boxes with an area smaller than 200 pixels were discarded to maintain a reasonable size for object detection. Nighttime images, which are available only in the infrared spectrum, are provided as individual captures (singletons) that add diversity to the dataset by addressing the challenges of nighttime maritime scenes. The variety of ship types and the combination of visible and infrared data make the VAIS dataset an essential resource for ship detection and classification in both day and night scenarios.



(a) (b)

Figure 7 (a) VAIS example optical image and (b) corresponding infrared image (Zhang et al., 2015).

The SeaShips dataset (Shao et al., 2018) is a large-scale, precisely annotated dataset for ship detection, focusing on object detection rather than image segmentation. It contains 31,455 images (see **Figure 8**) with a resolution of 1,080×1,920 pixels, extracted from 10,080 video clips captured by a coastal surveillance system around Hengqin Island, Zhuhai City, China.

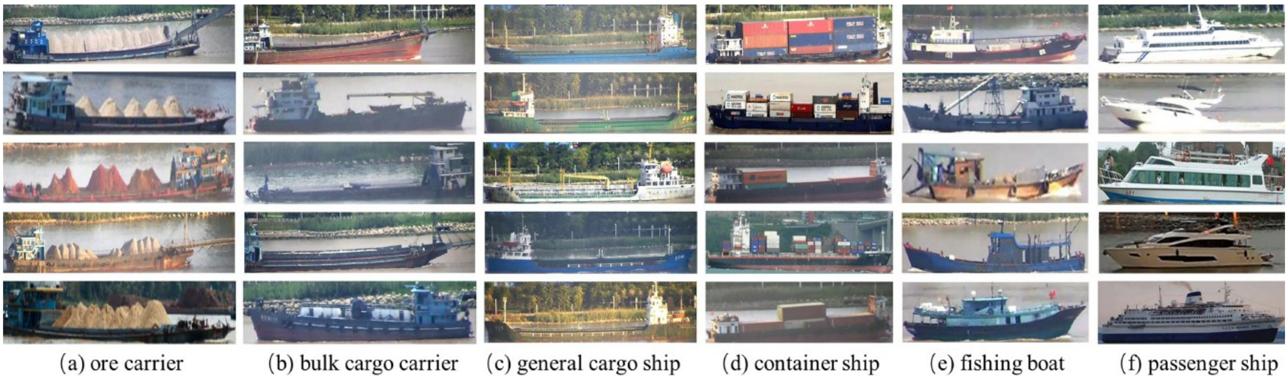


Figure 8 Cropped example images from the SeaShips dataset (Shao et al., 2018).

With 156 cameras deployed at 50 sites, the dataset covers 53 km² of coastal area. Images were collected over four months-January, April, August, and October-in 2017 and 2018, with daily capture times from 6:00 a.m. to 8:00 p.m. Images feature six common ship types: ore carrier, bulk cargo carrier, general cargo ship, container ship, fishing boat, and passenger ship. Ships are captured under various conditions, including different scales, hull parts, lighting, viewpoints, backgrounds, and occlusions.

Approximately 10 % of the ships are partially occluded by others, and ships with only parts of their hulls visible are also included. The dataset contains approximately 500 unique ships with manually annotated ship-type labels and high-precision bounding boxes. One frame was selected approximately every two seconds, and empty frames were discarded to ensure high-quality data. SeaShips is one of the best datasets for coastal ship classification tasks, although it does not include any images of warships. The Singapore Maritime Dataset (SMD) (Prasad et al., 2017) is a comprehensive dataset developed for maritime object detection and tracking in Singaporean waters. The dataset includes video recordings captured from both fixed on-shore platforms and moving vessels, providing a diverse range of maritime environments. It originally consisted of 81 videos: 40 from fixed on-shore platforms, 11 from on-board moving vessels, and 30 near-infrared on-shore videos (**Figure 9**. The dataset comprises 20,367 RGB images extracted from these videos, all captured at a resolution of 1,080×1,920 pixels using a Canon 70D camera. Recordings were taken between July 2015 and May 2016, at various times of the day, including 40 minutes before sunrise, at sunrise, midday, afternoon, evening, and up to two hours after sunset. The dataset also contains challenging weather conditions, such as haze and rain, providing varied environmental contexts. SMD has 10 object classes, 6 of which are different ship types. The annotations were performed by student volunteers, who manually labeled various objects in the video frames. The non-expert annotation indicates a potential need for further improvement in labeling accuracy.

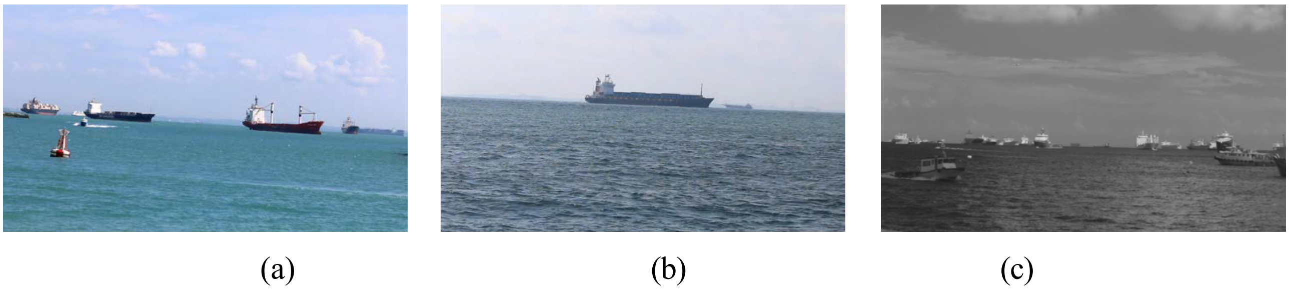


Figure 9 (a) SMD on-shore image; (b) on-board image and (c) near-infrared image. Images taken from Prasad et al. (2017).

To address the challenges observed when using the original SMD, the Singapore Maritime Dataset plus (SMD-plus) (Kim et al., 2022) was introduced. The original SMD contained several annotation issues, including inaccurately placed bounding boxes- some excessively loose, encompassing background areas, and others overly tight, capturing only parts of objects- as well as misclassified object labels (Kim et al., 2022). In addition, significant class imbalance limited the effectiveness of the dataset for training machine learning models. To resolve these issues, SMD-plus provides refined annotations for improved consistency and accuracy in bounding box placement and class labeling. The dataset also addresses class imbalance by discarding rare classes and merging similar ones to simplify classification. For instance, the "boat" and "speedboat" classes were combined, and the "ferry" class was merged into the "boat" category. Rare classes such as "person", "flying bird and plane", and "swimming person" were excluded to maintain a focus on more relevant maritime objects. Therefore, the SMD-plus dataset defines seven primary classes: Ferry, Buoy, Vessel/Ship, Boat, Kayak, Sailboat, and Other. These enhancements make SMD-plus a robust and better-balanced dataset, particularly suitable for developing and evaluating maritime object detection algorithms in complex environments.

The McShips dataset (Zheng & Zhang, 2020) is a large-scale collection of ship images intended for detection and fine-grained categorization. It consists of 14,709 annotated images captured by multiple cameras at various resolutions, with a minimum resolution of 500×500 pixels. This dataset also includes images taken under adverse conditions, such as fog, to enhance its diversity. Frames without ships were removed to maintain relevance, yielding a nearly balanced distribution of 7,953 warship images and 8,942 civilian ship images. Images were curated using a web crawler and search engines, and each was annotated by experts at least five times to ensure high accuracy. The dataset presents significant challenges due to high intra-class variability, which is caused by differences in viewpoint, weather conditions, illumination, scale changes, occlusion, and cluttered backgrounds. McShips includes 13 ship categories, divided into 6 warship classes (aircraft carrier, submarine, landing ship, auxiliary ship, destroyer, and missile boat) and 7 civilian ship classes (sailboat, speedboat, fishing boat, passenger ship, container ship, tugboat, and support ship).

The Global Large-Scale Ship Database (GLSD) (Shao et al., 2024) was created for ship detection and contains 152,576 images with resolutions ranging from 90×90 to 6,509×6,509 pixels. The GLSD combines images from a video monitoring system deployed in the Zhuhai Hengqin New Area, China, and images gathered via web crawling. The dataset encompasses more than 3,000 ports and 33 shipping routes across China, America, and Europe. The GLSD includes both iconic and non-iconic images, some taken from elevated or bird's-eye perspectives. Furthermore, some images contain visual disturbances such as watermarks, painted elements, and mosaic patterns, which the authors claim mimic real-world noise. The dataset consists of 13 different ship categories, namely sailing boats, fishing boats, passenger ships, warships, general cargo ships, container ships, bulk carriers, barges, ore carriers, speed boats, canoes, oil carriers, and tugs. Each image has been manually annotated by a professional labeling team, with bounding boxes marking ship instances and global attributes like location, weather, and lighting conditions. In total, there are over 212,357 annotated ship instances. The scale, composition and geographic diversity of the GLSD make it unique for training and testing ship detection algorithms for various computer vision tasks.

The Multi-Category Large-Scale Dataset for Maritime Object Detection (MCMOD) (Sun et al., 2023) was developed to enhance detection and classification capabilities for maritime vessels. This dataset comprises 16,166 labeled images at a resolution of 1,080×1,920 pixels, captured using three onshore video cameras in Hainan, China. MCMOD includes diverse scenes recorded 24 hours a day, encompassing various weather conditions such as fog, rain, and evening settings. Notably, MCMOD's high scene diversity is due to camera movements that include rotation and zooming, enabling to capture maritime traffic with variable illumination, object sizes, and perspectives. MCMOD contains 98,590 maritime objects across 10 categories, covering a broad range of vessel types comprising fishing vessels, speedboats, engineering ships, cargo ships, yachts, sailboats, buoys, rafts, and cruise ships.

The Split Port Ship Classification Dataset (SPSCD) (Petkovic et al., 2023) is a rather new dataset aimed at evaluating the performance of detection and classification algorithms for ships in passenger ports in the Mediterranean. It includes 19,337 high-resolution images captured by a Dahua DH-TPC-PT8620A-T camera at 1,080×1,920 pixel resolution. This dataset focuses on small- to medium-sized ships typically not monitored by systems such as AIS or VTS (Vessel Traffic System). The dataset includes images recorded under diverse weather conditions and throughout the day, providing a wide range of backgrounds and scene variations. It captures ships primarily from the port and starboard side as they enter or exit the Port of Split, Croatia. Ship categories reflect the specific maritime traffic profile of this region and include 12 classes, which consist of speed craft, fishing boats, and large ferries, excluding military and cargo-specific ships. The camera used for data collection was mounted 9 meters above sea level, recording video sequences from February 2020 to December 2022. The dataset contains 27,849 manually annotated ship instances, making it a valuable complement to other datasets.

The ABOships dataset (Iancu et al., 2021) is a maritime vessel detection dataset comprising 9,880 images with a resolution of 720×1,280 pixels, collected over 13 days along a single route from Turku to Ruissalo in Southwest Finland (**Figure 10**). The dataset is derived from 135 videos, recorded with a 65° field-of-view camera at 15 FPS, capturing various maritime vessels under a wide range of conditions, including background variation, atmospheric changes, occlusion, and scale variation. The dataset was extracted from 720p videos, covering urban areas along the Aura River, port scenes, and parts of the Finnish Archipelago under different lighting conditions. Each object was manually labeled as one of 11 categories, including seamarks, 9 vessel types (e.g., motorboats, sailboats, ferries, and cargo ships), and various floating objects, resulting in 41,967 annotated instances. To ensure labeling accuracy, an OpenCV-based tracker (Bradski, 2000) was employed to refine object labels across frames. The dataset's diversity, spanning various maritime environments, object sizes, and occlusions, enhances its value for machine vision applications in dynamic maritime surveillance.

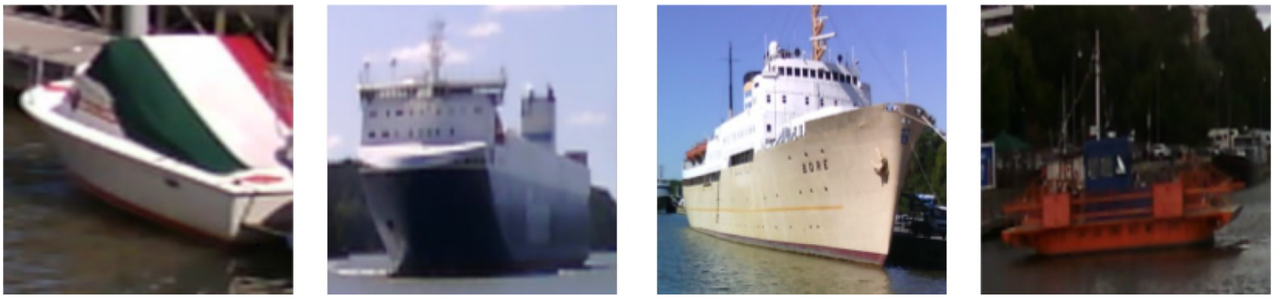


Figure 10 Example objects in the ABOships dataset taken from Iancu et al. (2021). From left to right: boat, cargo ship, cruise ship and ferry.

The ABOships-PLUS dataset (Iancu et al., 2023) is an enhanced version of the original ABOships dataset that addresses several limitations, such as class underrepresentation and the inclusion of small objects with areas below 16×16 pixels. In the original ABO-ships, only 1.3 % of the images contain objects from the class "Miscellaneous", and just 1.58 % of the images contain "Cargo ships", with only 200 and 161 annotations, respectively. ABOships-PLUS is expanded to 15,838 images and reclassifies objects into the four broader superclasses: Sailboat, Powerboat, Ship, and Stationary, eliminating smaller objects and increasing the total number of annotations to 33,227. ABOships-PLUS also accounts for frequent occlusions of maritime objects, boosting overall object detection performance by refining annotations and balancing object classes.

Finally, the Marine Vessel Detection Dataset 13 (MVDD13) (Wang et al., 2024) is a comprehensive dataset specifically designed for detection from USVs, featuring 35,474 annotated images. As the most recent dataset in this survey, MVDD13 reflects real-world challenges with images captured from multiple viewing angles (front, astern, and side views) and varying weather conditions, including cloudy, rainy, and foggy settings (**Figure 11**). The images were collected using a 360° panoramic camera mounted 2 meters above the water on a USV, supplemented with additional military vessel images from platforms such as ShipSpotting and MarineTraffic. The dataset includes 13 vessel categories, spanning military and civilian types such as cargo ships, cruise ships, tankers, sailboats, tugs, fishing boats, warships, and submarines. The MVDD13 dataset emphasizes real-world maritime environments, capturing scale variations, different lighting conditions, diverse viewpoints, and occlusions, enhancing its utility for USV visual perception systems. Annotation quality is ensured through a strict process in which images are labeled only if three experts agree on the classification. Occluded objects are annotated based on visible features, making the MVDD13 dataset a rich resource for advancing marine vessel detection with deep learning models.

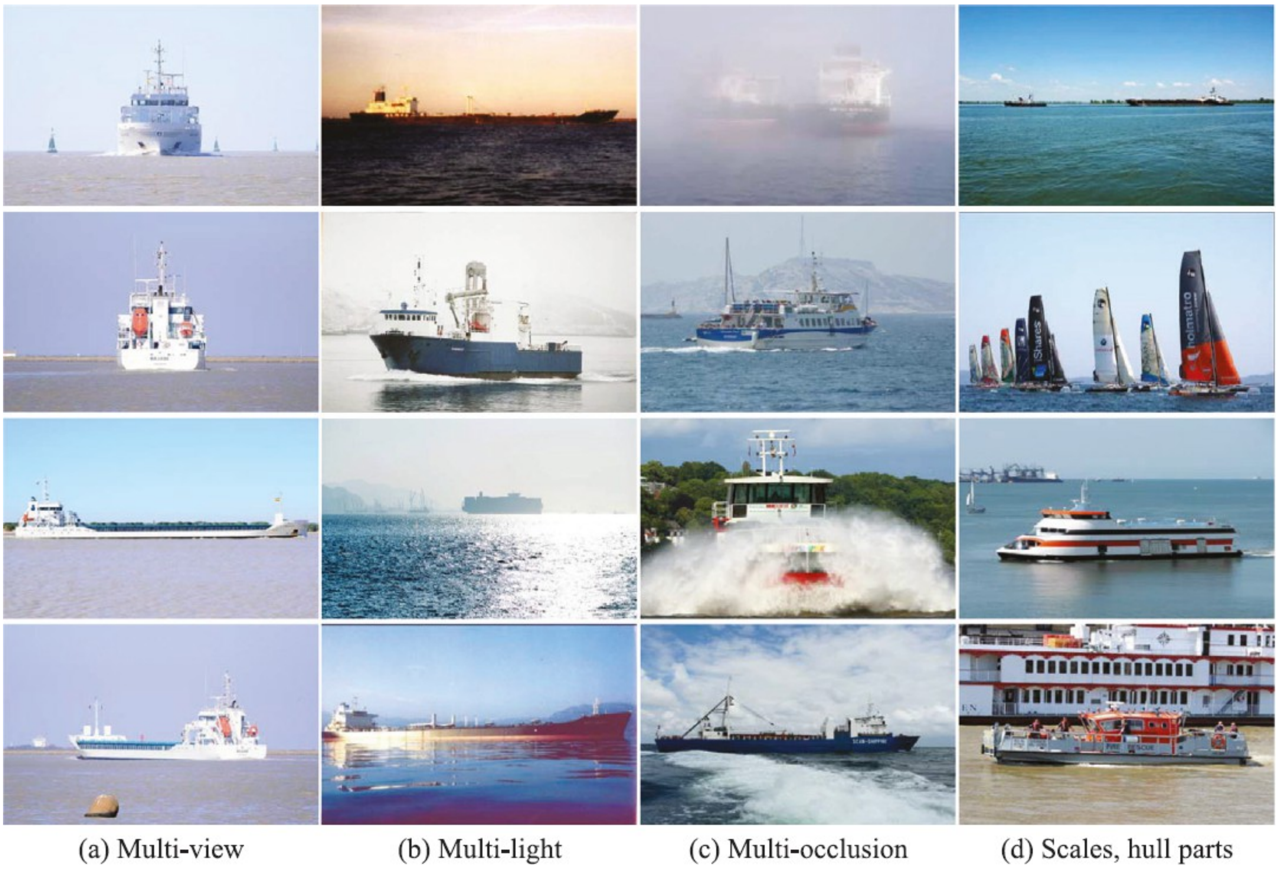


Figure 11 Example images from the MVDD13 dataset (Wang et al., 2024)

Detection and segmentation

Detection and segmentation annotations are provided by only three out of the 25 datasets. Papers are listed in this section if they provide any sort of detection with additional segmentation.

The Marine Obstacle Detection Dataset (MODD) (version 1) (Kristan et al., 2015) was primarily designed for testing obstacle detection algorithms for USVs. The dataset was obtained from video sequences (**Figure 12**) at a resolution of 480×640 pixels, recorded from multiple platforms, predominantly a 2.2-meter long USV equipped with an Axis 207 W camera. The camera was mounted approximately 0.7 meters above the water surface with a field of view of around 55°, aimed at the front of the vessel. It automatically adjusted to lighting conditions. Video sequences were collected in the Gulf of Trieste, near the port of Koper in Slovenia, over several months, capturing different weather conditions and times of day. To provide realistic operational environments, the videos include both normal navigation and simulated near-collision situations.

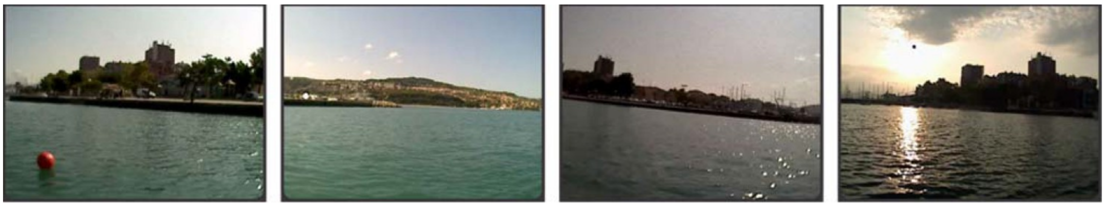


Figure 12 Example images obtained from videos of the MODD dataset (Kristan et al., 2015)

The MODD includes 10 videos recorded under normal conditions and 2 additional sequences marked as extreme conditions, where the USV is directly facing the sun. Each frame of the videos has been manually annotated, with the annotations verified by an expert, yielding 4,454 annotated images. The annotations include segmentation labels for water, sky, and horizon/shoreline, but the dataset does not provide detailed obstacle classification beyond these categories. The focus is on obstacle detection rather than classification, and the dataset is used for assessing the robustness of algorithms, with the best performance observed when using the YCrCb and Lab color spaces (Kristan et al., 2015).

The Marine Obstacle Detection Dataset 2 (MODD2) (Bovcon et al., 2018) is a maritime dataset containing 11,675 images and 28 videos, captured over 15 months in the Gulf of Koper, Slovenia, along with synchronized Inertial Measurement Unit (IMU) data. The dataset was designed to simulate realistic USV navigation scenarios, where obstacles pose potential threats (**Figure 13**). The dataset captures a variety of weather conditions and times of day to enhance visual diversity. In 25 of the 28 sequences, at least one obstacle is present, and dangerous situations, such as near-collisions, are simulated. Sun glitter and environmental reflections are present in many sequences, adding complexity to the data. A stereo camera system with a 132.1° field of view was mounted 0.7 meters above the water surface to capture the data. Obstacles are categorized into two classes: large obstacles, which intersect the water’s edge, and small obstacles, fully contained below the water-edge polygon. The annotations include manually labeled obstacles with bounding boxes and water-edge polygons. Annotations were produced in two stages, first by crowdsourcing and later refined by maritime computer vision experts to ensure high-quality labels. Additionally, the horizon was annotated in frames where it was clearly visible. MODD2 is valuable for testing and advancing obstacle detection for USV systems.

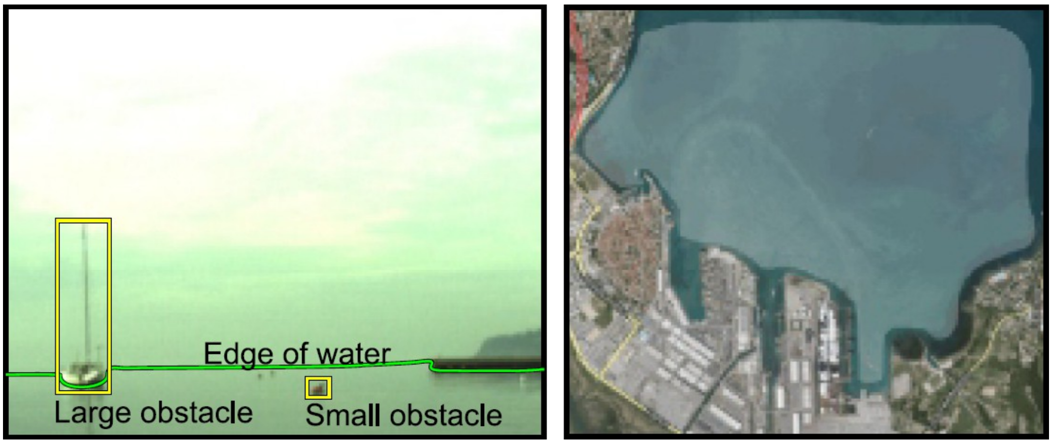


Figure 13 Left: Annotated example frame of the MODD2 dataset. Right: The highlighted area shows the coastal region of Koper, Slovenia, where the data were acquired (Bovcon et al., 2018).

The Maritime Imagery Dataset (MID) (Liu et al., 2021) is a high-definition dataset designed for USVs, focusing on the detection of marine obstacles under varied conditions. Captured in the coastal waters of Qingdao and Shanghai, China, over several months, MID contains a total of 2,655 labeled images, each with a resolution of 480×640 pixels. The images were recorded using a camera with a 50° field of view, mounted approximately 1 meter above the water surface. MID covers diverse maritime conditions encountered by coastal USVs, such as water reflections, visual blur from adverse weather, dim illumination, sun glitter, and horizon tilt due to waves and the motion of the USV. The dataset contains eight video sequences collected at different times of day and across multiple weather conditions, providing a comprehensive look at typical environmental challenges for USVs. Each sequence is annotated for obstacle detection, with dynamic obstacles labeled either as large (straddling the water's edge) or small (fully surrounded by water). The MID provides a challenging benchmark for developing and testing obstacle detection algorithms for USVs, with realistic visual distortions, scale variations, and occlusions that add to its utility for maritime perception research. The MariShipInsSeg dataset (Sun et al., 2022) is a collection of 4,001 manually annotated images from multiple cameras, specifically designed for ship instance segmentation in maritime environments. Approximately 60 % of the images were sourced from the Internet and 5 % from the SeaShips dataset, and the remainder consists of ship images from the COCO and VOC datasets. A key feature of the MariShipInsSeg dataset is the inclusion of ships under challenging and unusual maritime conditions, such as large waves, water reflections, dim lighting, and scenes with crowded ships, reflecting the complexities of the real world. MariShipInsSeg contains an average of 2.1 ship instances per image, ensuring a diverse range of ship types and scenarios for training and testing segmentation models. This dataset provides detailed instance-level annotations (**Figure 14**) for ships in various environments, making it valuable for advancing segmentation algorithms in maritime applications.

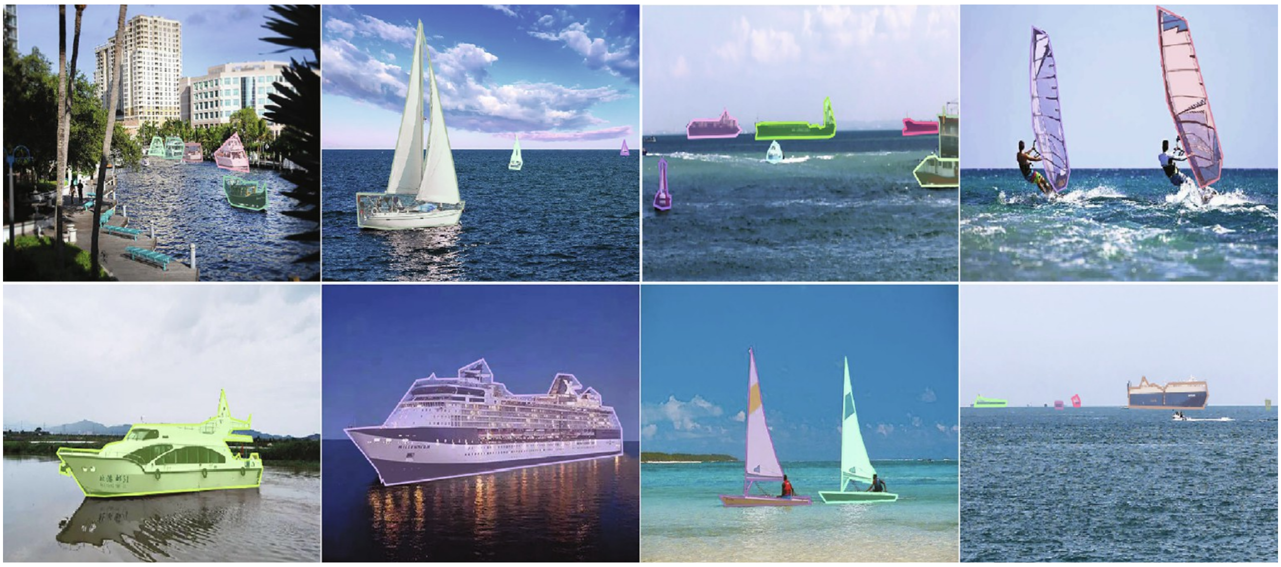


Figure 14 Example images with ground truth segmentations of the MariShipInsSeg dataset taken from Sun et al. (2022).

2.3.4 Detection, classification and segmentation

A dataset is only listed in this section if it provides detection, classification and segmentation labels. Only 4 out of the 25 datasets contain annotations for all three tasks. The Maritime Detection, Classification, and Tracking dataset (MarDCT) (Bloisi et al., 2015) was created to monitor boat traffic along the Grand Canal of Venice, Italy, covering a 6 km stretch of waterway that is 80 to 150 meters wide. The dataset consists of 6,743 images with a resolution of 240×800 pixels, captured by multiple cameras from various observing angles and under different weather conditions. It focuses on Venice's unique boat traffic rather than open sea scenarios, as the installed cameras primarily monitors boats navigating the canal. There are 24 specific boat categories, which are grouped into five general categories. In addition, three extra labels- Water, PartialView, and Multiple-Boats- are used in the annotations. The dataset features ground-truth annotations that include foreground masks for evaluating segmentation, bounding boxes for boat detection, and identification numbers to support data association and tracking over time. The boats in the dataset are mostly centered in the images, and the annotations were completed manually by experts in Venice boat navigation. When multiple boats are present in an image, only the one closest to the center is labeled. Each image contains exactly one boat category label, in addition to the Water label and, optionally, one of the additional labels. MarDCT covers complex scenarios where boats are partially visible, and includes challenging cases of multiple boats in high-traffic areas. Additionally, the camera angles, although not directly bird's-eye views, sometimes capture parts of a boat's deck, offering a unique perspective.

The SmartShip Object Detection-HEU dataset (Zhang et al., 2020) is a collection of maritime images (**Figure 15**) designed for detection, classification and segmentation tasks, specifically aimed at improving ship segmentation methods in complex maritime environments.

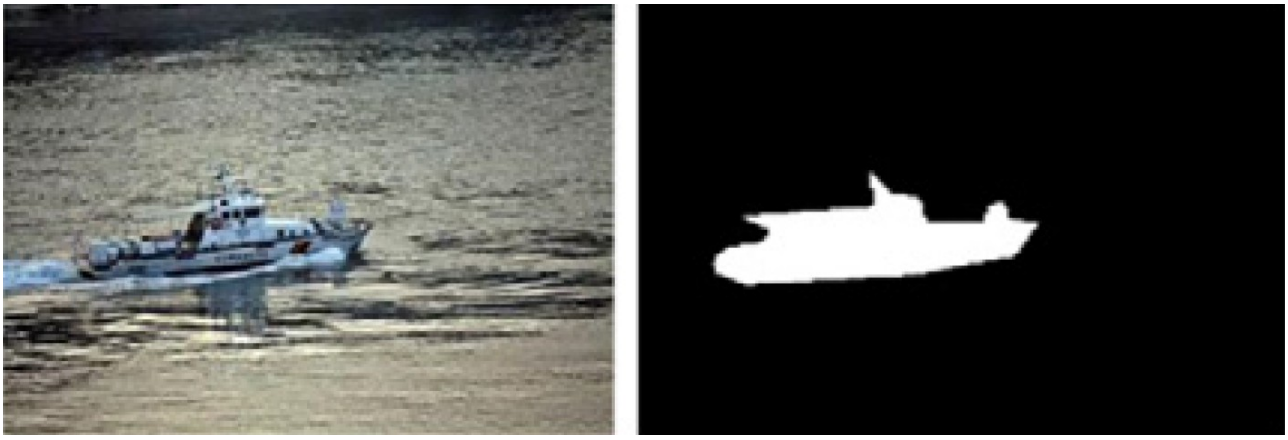


Figure 15 Example image (left) with ground truth segmentation (right) of the SmartShip Object Detection-HEU dataset taken from Zhang et al. (2020).

The dataset consists of 3,560 images, sourced from various datasets and the Internet. Approximately 10 % of the images are from the Singapore Maritime Dataset (SMD) and25 % are from MarDCT, and the remaining images were collected from Internet sources. The dataset also incorporates simulated sea fog in 745 images to address the challenges posed by environmental factors such as fog. By accounting for the presence of sea fog, which typically reduces segmentation accuracy (Sun et al., 2022), this dataset provides a valuable resource for improving the accuracy and robustness of ship detection algorithms. Because every image contains only one instance, and segmentation masks are prevalent, this dataset is considered to deal with detection too, since bounding boxes can be calculated from segmentation masks.

The Maritime Obstacle Detection Stereo (MODS) dataset (Bovcon et al., 2021) is a comprehensive dataset specifically designed for USV obstacle detection, classification, and segmentation. In order to create the final MODS dataset, the MODD, the MODD2, and the SMD were merged. For the self-recorded videos, every 10th frame was annotated with precise bounding boxes, classifying dynamic obstacles into three semantic categories: vessel, person, and other. The MODS dataset consists of 24,090 images captured by multiple cameras, resulting in 145,334 annotated objects. Images without obstacles were excluded, except those containing water reflections or glitter. It is designed for stereo vision-based object detection and consists of various real-world maritime scenes captured under diverse weather conditions, lighting situations, and sea states (**Figure 16**). The dataset includes stereo image pairs and corresponding depth maps, as well as pixel-level semantic annotations of obstacles such as boats, buoys, and other hazards. By providing these annotated stereo images, MODS enables researchers to develop and validate algorithms for tasks such as depth estimation, object detection, and semantic segmentation specifically tailored for the maritime domain.

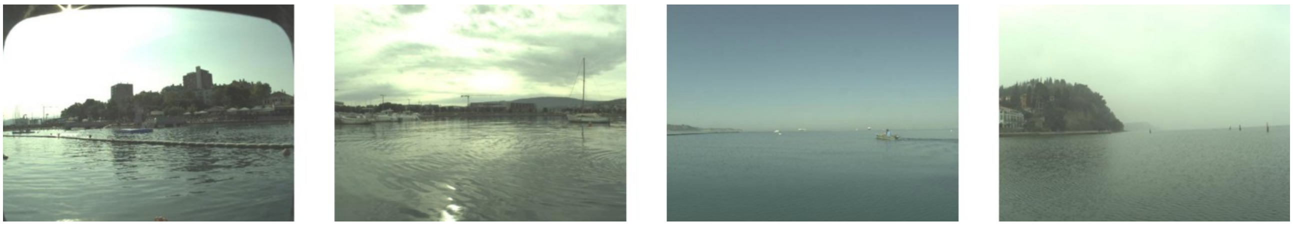


Figure 16 MODS example images taken from Teršek et al. (2023)

The LaRS dataset (Žust et al., 2023) is a diverse maritime dataset specifically designed for maritime obstacle detection, classification, and panoptic segmentation, covering lake, river, and sea domains. It consists of 4,006 keyframes, where for each keyframe there are also 9 previous frames in the data set. The dataset incorporates a variety of scene conditions (**Figure 17**), including different environments, illumination levels, and water surface conditions, making it ideal for USV-centric obstacle detection. The images in LaRS were collected from a mix of online video scenes, self-recorded footage from various locations, and carefully selected samples from existing maritime datasets, offering a wide range of geographic and environmental settings. Professional labeling ensures the annotations are both accurate and detailed. Each keyframe is annotated with three “stuff” categories- water, sky, and static obstacles like shores and piers- and eight “thing” categories, which include dynamic obstacles (**Figure 18**) such as boats, row boats, paddle boards, buoys, swimmers, animals, floating platforms, and an open-world “other” class for less common obstacles. In addition to segmenting objects, 19 global scene attributes (**Figure 17**) were assigned to the keyframes to capture features such as environment type, lighting conditions, water surface conditions, and reflections. Due to the additional 9 preceding frames for each keyframe, resulting in over 40k images in total, this dataset can be used for tracking or time series analysis. Challenging scenes with occluded objects such as swimmers and animals, as well as floating obstacles, make LaRS a valuable asset for developing robust obstacle detection models in complex maritime scenarios.

3. Comparison of datasets and discussion

In this section, the primary characteristics, annotation quality, scenario diversity, size, and limitations of the previously evaluated key maritime datasets commonly used in computer vision for vessel detection, classification, segmentation, and tracking are compared. The purpose of these comparisons is to highlight the strengths and weaknesses of the datasets, guiding the selection of suitable datasets for specific maritime vision applications, with special focus on autonomous navigation and obstacle detection. It is difficult to explicitly quantify dataset biases, such as regional dominance or imbalances in vessel types, because most diverse maritime datasets typically combine images from multiple datasets or are collected from the Internet, and thus lack consistent geographic metadata. Therefore, due to these inherent biases and annotation inconsistencies, current datasets might encounter limitations in scalability when supporting increasingly complex models.

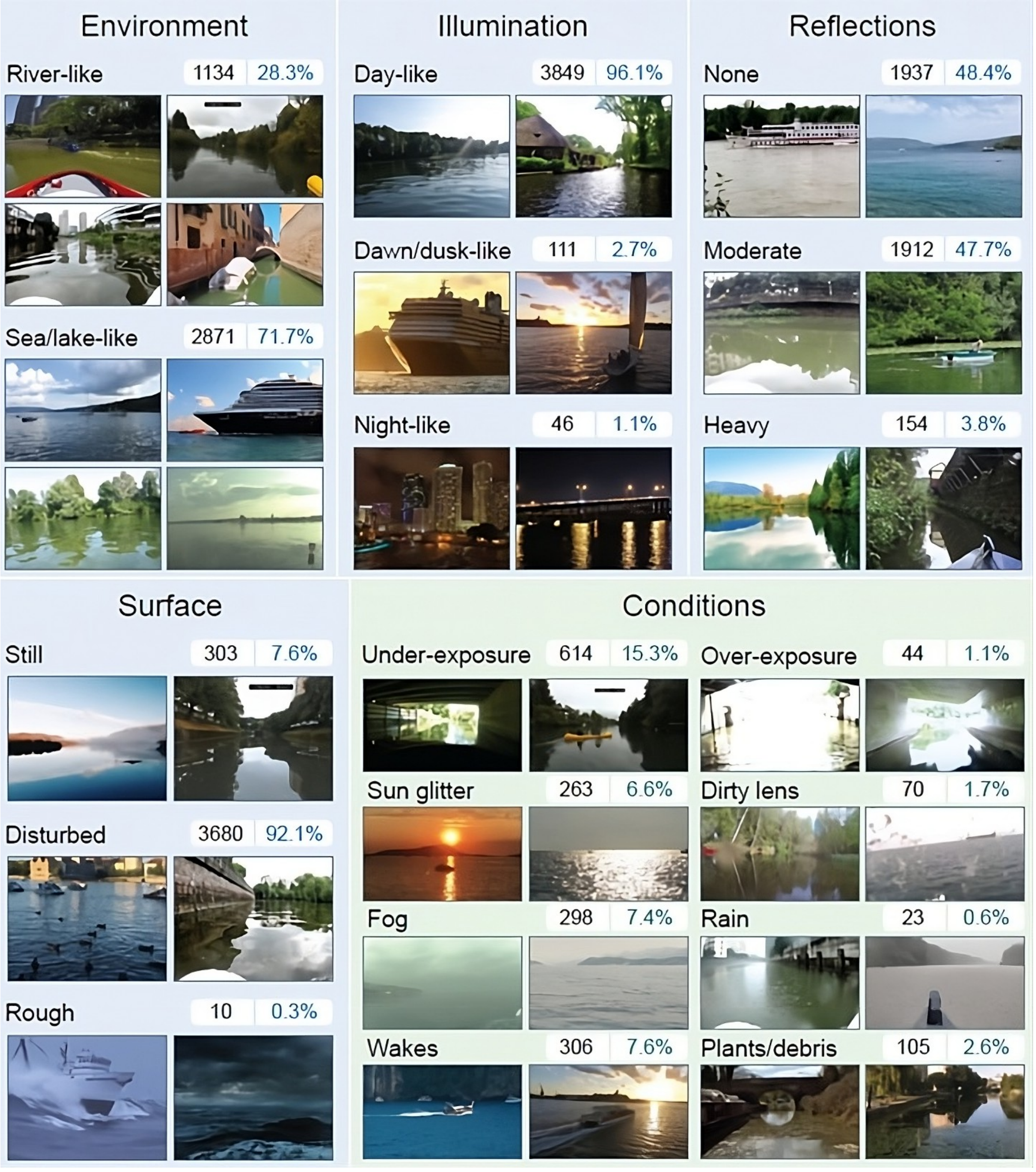
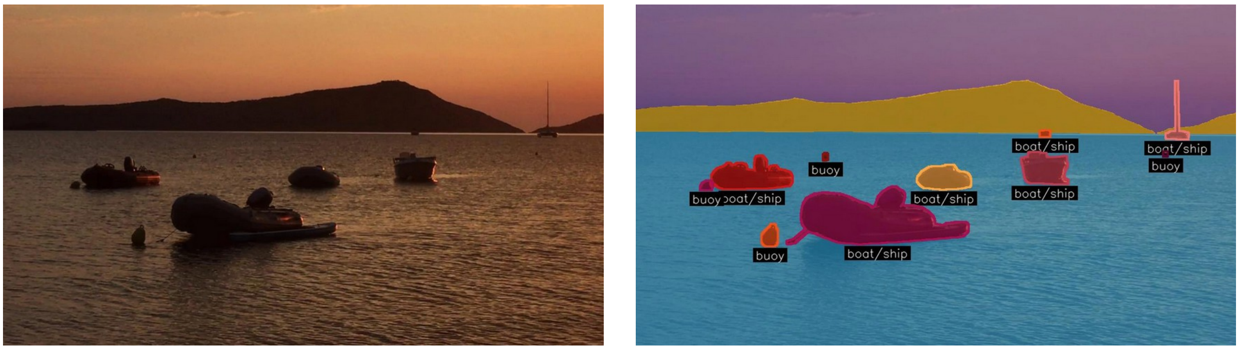


Figure 17 LaRS scene variety with the corresponding number of images and percentages with respect to all keyframes. All images contain condition attributes such as fog or wakes. Images taken from Žust et al. (2023).



(a) (b)

Figure 18 (a) Original example image of the LaRS dataset and (b) corresponding ground truth data taken from Žust et al. (2023).

3.1 Annotation quality and class diversity

Annotation quality is a crucial factor for accurate model training and performance evaluation. The McShips dataset (Zheng & Zhang, 2020), for example, stands out for its meticulous annotation process, with each image being labeled at least five times by expert annotators. This ensures high accuracy and consistency, making McShips suitable for fine-grained classification tasks. In contrast, the Singapore Maritime Dataset (SMD) (Prasad et al., 2017) has problems with inconsistent bounding boxes and inaccurate class labeling, while SMD-plus (Kim et al., 2022) addresses these shortcomings with refined annotations and balanced classes.

The MODS dataset (Bovcon et al., 2021) is valuable for obstacle detection research due to its extensive annotations but is somewhat limited by its categorization, which includes only three basic classes: vessel, person, and other. Additionally, it is imbalanced due to the inclusion of non-vessel images containing reflections, which may limit its applicability for navigation tasks.

Some datasets focus on a broad set of vessel types and environmental conditions. MVDD13 (Wang et al., 2024) excels in this regard, offering 13 distinct vessel categories with detailed annotations for various viewing angles and weather conditions.

MVDD13's exclusion of aerial images makes it generally suitable for navigation tasks, and its detailed consideration of real-world challenges such as scale variations and occlusions enhances its applicability to USVs. Nevertheless, notable class imbalances, particularly the limited representation of military vessels like submarines and warships, could somewhat constrain its effectiveness for tasks specifically requiring robust detection of these vessel categories.

3.2 Scenario diversity and geographic coverage

The diversity of scenarios and geographic coverage greatly influences the utility of a dataset for developing models that generalize well across different maritime environments. The SeaShips dataset (Shao et al., 2018) provides broad coverage of maritime scenarios with high-resolution images, although it is limited by its use of fixed surveillance cameras mounted along the coastline, resulting in a limited variation in location and camera angles. The same limitation holds for the Harbor Surveillance dataset (Zwemer et al., 2018), which focuses solely on coastal surveillance without covering open sea conditions or moving viewpoints. Another dataset with very limited geographic coverage is MarDCT (Bloisi et al., 2015), which solely contains images of Venice canal boat traffic.

The Global Large-Scale Ship Database (GLSD) (Shao et al., 2024) claims extensive geographic coverage, spanning multiple ports in China, the United States, and Europe. GLSD includes images with watermarks and mosaic patterns, which the authors suggest may introduce noise. While this could potentially improve the robustness of detection and classification of arbitrary images sourced from the Internet, it remains unclear whether such augmentation specifically improves robustness to camera-induced noise. Consequently, GLSD may not be well-suited for USV guidance or fine-grained maritime navigation tasks.

3.3 Dataset size and temporal context

The MU-SID dataset (Hashmani & Umair, 2022) is unique in its focus on capturing different states of the sea near Malaysia. However, its narrow geographical scope may limit its generalizability to other maritime conditions. Larger datasets generally allow for more robust model training, particularly when diverse environmental and temporal conditions are represented. Detecting the sea horizon line, as supported by MU-SID, is beneficial for autonomous navigation since it offers a reliable visual reference for maintaining a vessel's heading relative to the horizon. However, horizon detection alone does not provide sufficient semantic information about maritime obstacles, vessels, or other relevant features necessary for comprehensive navigation. Therefore, additional datasets incorporating diverse maritime objects and environmental annotations are required to enable fully autonomous maritime navigation. Similarly, MaSTr1325 (Bovcon et al., 2019) has too few images and limited geographical diversity, restricting its usefulness to specific tasks and making it unsuitable for more general applications. The LaRS dataset offers a unique advantage by incorporating substantial temporal context through sequences capturing maritime traffic over time. Nevertheless, its comparatively small size may limit its effectiveness for training robust models by reducing its generalizability to different maritime conditions. However, temporal data remains particularly valuable for sequential analysis tasks comprising vessel tracking and movement prediction, especially when supplemented by high-quality annotations.

Object tracking over longer periods can be done based on the FVessel dataset (Guo et al., 2023). Videos are up to 25 minutes long, with annotations provided for every second.

3.4 Use-case limitations and practical applicability

Several datasets are constrained by their use case focus, limiting their applicability to general navigation or surveillance tasks. For instance, SPSCD (Petkovic et al., 2023) is designed for passenger port traffic in the Mediterranean, lacking diversity concerning vessel types and geographic coverage, excluding military vessels altogether.

For navigation and surveillance tasks that require comprehensive coverage of vessel classes, the SeaShips dataset (Shao et al., 2018) is not suited due to its exclusion of warships and its primary focus on coastal areas. Similarly, the VAIS dataset (Zhang et al., 2015) contains only larger vessels, with no instances of vessels under 200 pixels, making it less adequate for applications requiring detection of small objects.

3.5 Recommendations for dataset selection for autonomous navigation

Selecting an appropriate dataset is essential for the successful deployment of computer vision models in various maritime applications. The following recommendations are based on the extensive analysis in Section 2 and are specific to autonomous navigation tasks. Key factors include robustness to different environmental conditions, annotation quality, scene diversity, and class variety. Generally, the choice of dataset should align with the specific requirements of the intended application, balancing these factors to optimize performance.

For small USVs performing advanced tasks, such as environmental monitoring, data collection, or detailed object recognition, beyond fundamental navigational functions like obstacle avoidance and waypoint following, the LaRS dataset (Žust et al., 2023) is particularly valuable. This is because it belongs to the few datasets providing panoptic annotations and global scene attributes, essential for such complex maritime operations. As a result of pixel-wise exhaustive annotations, LaRS supports a wide range of tasks. Compared to the other evaluated datasets, LaRS does not only include detailed object annotations, but also offers additional attributes, such as illumination conditions and reflections, making it especially useful for applications that need to take into account environmental features. LaRS's diverse scenes, including multiple maritime contexts, and its coverage of non-vessel objects such as marine animals, make it a unique resource for environmental protection efforts and small USV applications that require situational awareness beyond basic vessel detection.

For use cases where environmental robustness is critical, such as maritime surveillance or navigation, datasets collected at a variety of locations and under various conditions are ideal. In this respect, the recent MVDD13 dataset (Wang et al., 2024) stands out due to comprehensive coverage of maritime scenarios. With scenes taken from multiple viewing angles, under various weather conditions such as fog and rain, and at different times of day, MVDD13 provides one of the most holistic views of maritime environments among all analyzed datasets. Moreover, potential challenges arising from diverse weather scenarios can be partially mitigated through oversampling, emphasizing underrepresented conditions during model training. Its thorough, multi-expert annotation process ensures high-quality data, leading to reliable models.

It is finally noted that the datasets recommended here were published only after the release of Su et al. (2023) and could, therefore, not be included in its analysis. Ultimately, the selection of a dataset should reflect the specific needs of the application and, thus, should consider annotation quality, class diversity, scene variation, size, and the environmental conditions the model is expected to encounter.

4. Conclusions

In this survey, open-source maritime vision datasets published between 2015 and 2024 have been examined by evaluating their relevance and applicability to computer vision tasks for autonomous surface vessel navigation. This study offers one of the most comprehensive overviews of publicly available maritime datasets. The objective is to provide researchers new to the field with the available resources and their suitability for autonomous navigation, thereby facilitating future advancements in the field. However, several datasets are constrained by their use case focus, limiting their applicability to general navigation or surveillance tasks.

To ensure a higher level of reliability of maritime computer vision models, future research should focus on merging datasets with diverse environmental factors, such as different weather conditions, times of day and, especially, geographic locations. Additionally, it is crucial to prioritize datasets that include objects beyond vessels, such as humans and marine animals, to account for environmental considerations and safety aspects. Expanding coverage to include open-sea scenarios and various coastal environments will further enhance the robustness of vision models and allow for safer and more comprehensive navigation capabilities.

To effectively merge and utilize datasets from diverse sources, future work should strive for standardization by employing, e.g., unified data formats, consistent labeling conventions, and standardized annotation frameworks. Cross-dataset label mapping and annotation refinement protocols can mitigate inconsistencies, leading to better interoperability among datasets. Establishing standardized evaluation benchmarks will further improve comparability and reliability for maritime vision research.

In conclusion, the selection of a dataset should be based on the specific requirements of the intended application, balancing annotation quality, class diversity, scenario variation, size, and geographic relevance. With the appropriate datasets, the field of maritime computer vision can continue to advance, driving innovation in autonomous systems and enhancing the safety, efficiency, and environmental responsibility of maritime operations.

Acknowledgment

The authors would like to express their gratitude to TKMS ATLAS ELEKTRONIK GmbH for their valuable insights and collaboration, which significantly contributed to the completion of this work.

References

Abdelsalam, H. E. B., & Elnabawi, M. N. (2024). The transformative potential of artificial intelligence in the maritime transport and its impact on port industry. *Maritime Research and Technology, 3*(1), 19-31. https://doi.org/10.21622/MRT.2024.03.1.752

Bird, J. J., & Lotfi, A. (2024). CIFAKE: Image classification and explainable identification of aigenerated synthetic images. *IEEE Access, 12*, 15642-15650. https://doi.org/10.1109/ACCESS.2024.3356122

Bloisi, D.D., Iocchi, L., Pennisi, A., & Tombolini, L. (2015). *Argos-venice boat classification* (pp. 1-6). In Proceedings of the IEEE International Conference on Advanced Video and Signal Based Surveillance. https://doi.org/10.1109/AVSS.2015.7301727

Bovcon, B., Muhovič, J., Perš, J., & Kristan, M. (2019). *The mastr1325 dataset for training deep USV obstacle detection models* (pp. 3431-3438). In Proceedings of the International Conference on Intelligent Robots and Systems. https://doi.org/10.1109/IROS40897.2019.8967909

Bovcon, B., Muhovič, J., Vranac, D., Mozetič, D., Perš, J., & Kristan, M. (2021). MODS: A USV-oriented object detection and obstacle segmentation benchmark. *IEEE Transactions on Intelligent Transportation Systems, 23*(8), 13403-13418.

https://doi.org/10.1109/TITS.2021.3124192

Bovcon, B., Mandeljc, R., Perš, J., & Kristan, M. (2018). Stereo obstacle detection for unmanned surface vehicles by IMU-assisted semantic segmentation. *Robotics and Autonomous Systems, 104,* 1-13. https://doi.org/10.1016/j.robot.2018.02.017

Bradski, G. (2000). *The OpenCV Library*. Dr. Dobb's Journal of Software Tools.

Cheng, Y., Zhu, J., Jiang, M., Fu, J., Pang, C., Wang, P., Sankaran, K., Onabola, O., Liu, Y., Liu, D., & Bengio, Y. (2021). *FloW: A dataset and benchmark for floating waste detection in Inland Waters* (pp. 10953-10962). In Proceedings of the International Conference on Computer Vision. https://doi.org/10.1109/ICCV48922.2021.01077

Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009). *ImageNet: A large-scale hierarchical image database* (pp. 248-255). In Proceedings of the Computer Vision and Pattern Recognition. https://doi.org/10.1109/CVPR.2009.5206848

Dosovitskiy, A., Dosovitskiy, A., Beyer, L., Beyer, L., Kolesnikov, A., Kolesnikov, A., Weissenborn, D., Weissenborn, D., Zhai, X., Zhai, X., Unterthiner, T., Unterthiner, T., Dehghani, M., Dehghani, M., Minderer, M., Minderer, M., Heigold, G., Heigold, G., Gelly, S., Gelly, S., Uszkoreit, J., Uszkoreit, J., Houlsby, N., & Houlsby, N. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *Computer Vision and Pattern Recognition.* https://doi.org/10.48550/arXiv.2010.11929

Everingham, M., Eslami, S. A., Van Gool, L., Williams, C. K., Winn, J., & Zisserman, A. (2015). The PASCAL visual object classes challenge: A retrospective. *International Journal of Computer Vision, 111*, 98-136 https://doi.org/10.1007/s11263-014-0733-5

Gribbestad, M., Hassan, M. U., & Hameed, I. A. (2021). Transfer learning for prognostics and health management (PHM) of marine air compressors. *Journal of Marine Science and Engineering, 9*(1), 47. https://doi.org/10.3390/jmse9010047

Gundogdu, E., Solmaz, B., Yücesoy, V., & Koc, A. (2017). *MARVEL: A large-scale image dataset for maritime vessels* (pp. 165-180). In Proceedings of the Asian Conference on Computer Vision. https://doi.org/10.1007/978-3-319-54193-8_11 . Springer

Guo, Y., Liu, R. W., Qu, J., Lu, Y., Zhu, F., & Lv, Y. (2023). Asynchronous trajectory matching-based multimodal maritime data fusion for vessel traffic surveillance in Inland Waterways. *IEEE Transactions on Intelligent Transportation Systems, 24*(11), 12779-12792. https://doi.org/10.1109/TITS.2023.3285415

Hashmani, M. A., & Umair, M. (2022). A novel visual-range sea image dataset for sea horizon line detection in changing maritime scenes. *Journal of Marine Science and Engineering, 10*(2), 193. https://doi.org/10.3390/jmse10020193

Iancu, B., Soloviev, V., Zelioli, L., & Lilius, J. (2021). ABOships: An inshore and offshore maritime vessel detection dataset with precise annotations. *Remote Sensing, 13*(5), 988. https://doi.org/10.3390/rs13050988

Iancu, B., Winsten, J., Soloviev, V., & Lilius, J. (2023). A benchmark for maritime object detection with centernet on an improved dataset, ABOships-PLUS. *Journal of Marine Science and Engineering, 11*(9), 1638. https://doi.org/10.3390/jmse11091638

International Maritime Organisation: Autonomous Shipping. (2024). *Autonomous shipping*. Retrieved from https://www.imo.org/en/MediaCentre/HotTopics/Pages/Autonomous-shipping.aspx

Johnson, J. M., & Khoshgoftaar, T. M. (2019). Survey on deep learning with class imbalance. *Journal of Big Data, 6*(1), 1-54. https://doi.org/10.1186/s40537-019-0192-5

Khan, M. M., Schneidereit, T., Mansouri Yarahmadi, A., & Breuß, M. (2024). Investigating training datasets of real and synthetic images for outdoor swimmer localisation with YOLO. *AI, 5*(2), 576-593. https://doi.org/10.3390/ai5020030

Kim, J. H., Kim, N., Park, Y. W., & Won, C. S. (2022). Object detection and classification based on YOLO-v5 with improved maritime dataset. *Journal of Marine Science and Engineering,* *10*(3), 377. https://doi.org/10.3390/jmse10030377

Kristan, M., Kenk, V. S., Kovačič, & S., Perš, J. (2015). Fast image-based obstacle detection from unmanned surface vehicles. *IEEE Transactions on Cybernetics, 46*(3), 641-654. https://doi.org/10.1109/TCYB.2015.2412251

Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014). *Microsoft COCO: Common objects in context* (pp. 740-755). In Proceedings of the 13th European Conference, Zurich, Switzerland. https://doi.org/10.1007/ 978-3-319-10602-1_48

Liu, J., Li, H., Luo, J., Xie, S., & Sun, Y. (2021). Efficient obstacle detection based on prior estimation network and spatially constrained mixture model for unmanned surface vehicles. *Journal of Field Robotics, 38*(2), 212-228. https://doi.org/10.1002/ rob.21983

Liu, T., Pang, B., Ai, S., & Sun, X. (2020). Study on visual detection algorithm of sea surface targets based on improved YOLOv3. *Sensors, 20*(24), 7263. https://doi.org/10.3390/ s20247263

Nirgudkar, S., DeFilippo, M., Sacarny, M., Benjamin, M., & Robinette, P. (2023). Massmind: massachusetts maritime infrared dataset. *The International Journal of Robotics Research, 42*(1-2), 21-32. https://doi.org/10.1177/02783649231153020

Petković, M., Vujović, I., Lušić, Z., & Šoda, J. (2023). Image dataset for neural network performance estimation with application to maritime ports. *Journal of Marine Science and Engineering, 11*(3), 578. https://doi.org/10.3390/jmse11030578

Prasad, D. K., Rajan, D., Rachmawati, L., Rajabally, E., & Quek, C. (2017). Video processing from electro-optical sensors for object detection and tracking in a maritime environment: A survey. *IEEE Transactions on Intelligent Transportation Systems, 18*(8), 1993-2016. https://doi.org/10.1109/TITS.2016.2634580

Sambasivan, N., Kapania, S., Highfill, H., Akrong, D., Paritosh, P., & Aroyo, L. M. (2021). *Everyone wants to do the model work, not the data work: Data cascades in high-stakes AI* (pp. 1-15). In Proceedings of the Chi Conference on Human Factors in Computing Systems. https://doi.org/10.1145/3411764.3445518

Shao, Z., Wang, Y., Wang, J., Deng, L., Huang, X., Lu, T., Luo, F., & Zhang, R. (2024). GLSD: A global large-scale ship database with baseline evaluations. *Geo-Spatial Information Science, 2024*; 1-15. https://doi.org/10.1080/10095020.2024.2416896

Shao, Z., Wu, W., Wang, Z., Du, W., & Li, C. (2018). SeaShips: A large-scale precisely annotated dataset for ship detection. *IEEE Transactions on Multimedia, 20*(10), 2593-2604. https://doi.org/10.1109/TMM.2018.2865686

Su, L., Chen, Y., Song, H., & Li, W. (2023). A survey of maritime vision datasets. *Multimedia Tools and Applications, 82*(19), 28873-28893. https://doi.org/10.1007/s11042-023-14756-9

Sun, Y., Su, L., Luo, Y., Meng, H., Li, W., Zhang, Z., Wang, P., & Zhang, W. (2022). Global mask R-CNN for marine ship instance segmentation. *Neurocomputing, 480*, 257-270. https://doi.org/10.1016/j.neucom.2022.01.017

Sun, Y., Su, L., Luo, Y., Meng, H., Zhang, Z., Zhang, W., & Yuan, S. (2022). IRDCLNet: Instance segmentation of ship images based on interference reduction and dynamic contour learning in foggy scenes. *IEEE Transactions on Circuits and Systems for Video Technology, 32*(9), 6029-6043. https://doi.org/10.1109/tcsvt.2022.3155182

Sun, Z., Hu, X., Qi, Y., Huang, Y., & Li, S. (2023). MCMOD: The multi-category large-scale dataset for maritime object detection. *Computers, Materials and Continua, 75*(1), 1657-1669. https://doi.org/10.32604/cmc.2023.036558

Teršek, M., Žust, L., & Kristan, M. (2023). eWaSR: An embedded-compute-ready maritime obstacle detection network. *Sensors, 23*(12), 5386. https://doi.org/10.3390/ s23125386

Wang, N., Wang, Y., Wei, Y., Han, B., & Feng, Y. (2024). Marine vessel detection dataset and benchmark for unmanned surface vehicles. *Applied Ocean Research, 142*, 103835. https://doi.org/10.1016/j.apor.2023.103835

Wang, P. (2021). Research on comparison of LiDAR and camera in autonomous driving. *Journal of Physics: Conference Series, 2093*(1), 012032. https://doi.org/10. 1088/1742-6596/2093/1/012032

Zhang, M. M., Choi, J., Daniilidis, K., Wolf, M. T., & Kanan, C. (2015*). VAIS: A dataset for recognizing maritime imagery in the visible and infrared spectrums* (pp. 10-16). In Proceedings of the Conference on Computer Vision and Pattern Recognition Workshops. https://doi.org/10.1109/CVPRW.2015.7301291

Zhang, W., He, X., Li, W., Zhang, Z., Luo, Y., Su, L., & Wang, P. (2020). An integrated ship segmentation method based on discriminator and extractor. *Image and Vision Computing, 93*, 103824. https://doi.org/10.1016/j.imavis.2019.11.002

Zheng, Y., & Zhang, S. (2020). *McShips: A large-scale ship dataset for detection and fine-grained categorization in the wild* (pp. 1-6). In Proceedings of the 2020 IEEE International Conference on Multimedia and Expo. https://doi.org/10.1109/ICME46284. 2020.9102907

Žust, L., & Kristan, M. (2022). *Temporal context for robust maritime obstacle detection* (pp. 6340-6346). In Proceedings of the International Conference on Intelligent Robots and Systems. https://doi.org/10.1109/IROS47612.2022.9982043 . IEEE

Žust, L., Perš, J., & Kristan, M. (2023). *LaRS: A diverse panoptic maritime obstacle detection dataset and benchmark* (pp. 20304-20314). In Proceedings of the International Conference on Computer Vision. https://doi.org/10.1109/ICCV51070.2023.01857

Zwemer, M. H., Wijnhoven, R. G., & With, P. H. (2018). Ship detection in harbour surveillance based on large-scale data and CNNs. *Visigrapp, 5*, 153-160. https://doi.org/10.5220/0006541501530160