Sea Situational Awareness (SeaSAw) Dataset

Parneet Kaur
pkaur@sea-machines.com

Arslan Aziz
aaziz@sea-machines.com

Darshan Jain
djain@sea-machines.com

Harshil Patel
hpatel@sea-machines.com

Jonathan Hirokawa
jhirokawa@sea-machines.com

Lachlan Townsend
ltownsend@sea-machines.com

Christoph Reimers
creimers@sea-machines.com

Fiona Hua
fhua@sea-machines.com

Abstract

Vessels move 90% of international cargo by volume, with the marine economy contributing to 5.1% of global GDP. As one of the oldest industries, the marine industry has yet to embrace innovations in modern technology to safeguard the blue economy. Situational awareness from intelligent vessel systems can enable enhanced safety and decision-making for mariners. As the foundation for these intelligent systems, advanced perception technology requires sufficient real-world operational data to leverage recent AI technologies. In this work, we introduce the Sea Situational Awareness (SeaSAw) dataset - a novel dataset that is comprised of 1.9 million images with 14.6 million objects associated with 20.4 million attributes from 12 object classes, making it the largest maritime dataset for object detection, fine-grained classification and tracking. Furthermore, this dataset consists of 9 sources in combination with various RGB cameras, mounted on different moving vessels, operating in different geographic locations globally, having variations in scenario, weather and illumination conditions. This data collection took place across 4 years with rigorous efforts on data selection, annotation, management and analysis to enhance the marine perception technology.

1. Introduction

Marine navigation for surface vessels requires operational awareness combining true 360° horizontal information with various levels of undersea depth information. Commonly used sensors on surface vessels include AIS, radar and nautical charts. Each of these sensors has its own limited capability and reliability for different weather conditions, sea states and navigational requirements. Even with a complete suite of sensors, the biggest challenge is the limited visual information that humans can easily understand and interoperate to navigate vessels in accordance with COLREGS [14], the leading set of international regulations for maritime navigation and collision avoidance.

To overcome human limitations in vessel navigation under COLREGS, the maritime industry can benefit from advanced perception technology to enable enhanced vessel situational awareness and autonomous operation, following the example in other domains (e.g. autonomous vehicles and drones).

Recent work has proposed using multiple sensors such as infrared cameras, RGB cameras, radar, and LiDAR for vessel detection [3, 4, 7]. Other work has also explored remote sensing approaches such as optical satellite imagery [10] and synthetic-aperture radar (SAR) [9, 12]. For general-purpose detection of objects using on-board sensors, AIS provides limited accuracy due to the loose requirements on installation and usage procedures. Marine RADAR usually focus on long range (i.e., up to 20,000 meters) for varying environmental conditions and marine long range LiDAR has health concerns when required to past a 2000 meters range. Overall, all of these sensors lack the visible wavelength information. However, RGB camera data uniquely presents rich features that can enhance the performance of perception tasks while maintaining easy understanding and interpretation. Additionally, RGB camera data is the most cost-efficient with respect to sensor and compute hardware, and data processing. Thus, RGB cameras are a key source for advanced perception.

In practice, vessel identification from RGB camera images is still a challenging task, primarily due to a lack of available data that captures the heterogeneity of the maritime environment resulting from different weather conditions, sea states, and vessel movements. Additionally, complex variation in the size, category and appearance of vessels is much greater in magnitude as compared to cars, human faces, and other favored objects in computer vision tasks.

(a) Location variability. [Left] Harbor. [Center] Far sea shore. [Right] Open sea.

(b) Weather variability. Rough weather such as rain and fog impacts the visibility of the objects as seen from the camera.

(c) Illumination variability. Sharp glare, dull light or bright lights at night add complexity.

(d) Viewpoint variability. Same object from front and side view: [Left]] Workboat. [Right] Towing vessel.

(e) Camera variability. [left] FoV is 90° with 3840×2160. [Middle] FoV is 180° with 3648×2052. [Right] FoV is 180° with 7840×1408.

(f) Scale variability. The scale of objects decreases as the distance increases. Vessel size in pixels[Left to right]: 140×40, 75×20, 38×12, 12×11. Smaller objects that are only a few pixels in size can be hard to detect, classify and track.

Figure 1. Sea Situational Awareness (SeaSAw) Dataset: a diverse dataset collected in several geographical locations across United States and Europe on 6 different vessels and using 5 different cameras of varying FoV and resolution. It is comprised of 1.9 million images with 14.6 million objects from 12 classes in a variety of locations, weather and illumination conditions.

Dataset	Images	Objects	Classes	Resolution	Scope
Singapore Maritime Dataset [15]	20,367	157,668	10	1920 × 1080	Detection
SeaShips [17]	31,455	40,077	6	1920 × 1080	Detection
Harbor Surveillance [20]	48,966	70,513	1	2048 × 1536	Detection, Tracking
McShips [19]	14,709	26,529	13	random	Detection, Classification
Marvel [6]	2,000,000	2,000,000	109	random	Classification
VIAS [18]	1,623	not available	15	5056 × 5056	Classification
GLSD [16]	100,729	140,616	13	random	Detection, Classification
ABOShips [8]	9,880	41,967	9	1920 × 720	Detection, Classification
SeaSAw (Ours)	**1.9M**	**14.6M**	**12**	**7680 × 1408** **3840 × 2056** **3648 × 2052** **1920 × 1080**	**Detection, Classification, Tracking**

Table 1. Maritime Dataset for object detection and classification using RGB images. In comparison, SeaSAw dataset (ours) is the largest and most diverse dataset collected from several cameras mounted on vessels in dynamic environment.

To address these challenges, we present the Sea Situational Awareness (SeaSAw) Dataset: a diverse RGB imagery dataset comprising of 1.9 million high-resolution images with 14.6 million objects from varying cameras, vessels and geographic locations that can leverage advanced perception technologies for marine vessel detection, classification and tracking.

The rest of the paper is organized as follows: Section 2 discusses prior datasets for maritime object detection and classification. Section 3 discusses the data collection and data processing systems to build the SeaSAw dataset, followed by a statistical analysis and discussion of the dataset. Section 4 concludes this work.

2. Related Work

Water surface vessels have existed as one of the categories in well-known computer vision datasets such as PASCAL VOC2007 [2], CIFAR-10 [11], Caltech-256 [5] and Coco [13], but in limited quantity and quality.

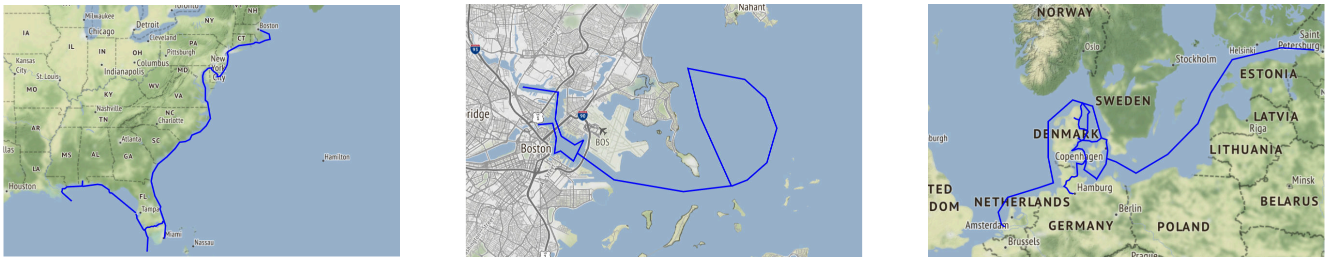
Several datasets have been introduced recently for enhancing performance of computer vision tasks such as detection, classification and tracking in maritime environments (Table 1).

The Singapore Maritime Dataset(SMD) [15] has RGB and near-infrared images, of which the RGB image dataset contains of 20,367 images with 157,668 objects, sampled from 51 videos among which 40 videos are taken from a stationary viewpoint on the shore and 11 videos are taken from on-board a vessel.

The SeaShips dataset [17] is comprised of 31,455 images with 6 ship classes. The images are collected from fixed surveillance cameras mounted on the coastline and thus lack variation in location and camera movement. The Harbor Surveillance dataset [20] contains 70,513 ships in 48,966 images collected with 10 camera viewpoints. For both of these datasets, the use case is limited to surveillance in the harbor area and is difficult to extend to dynamic environments to aid in navigation and collision avoidance.

The McShips dataset [19] has images and videos with at least 500 × 500 resolution, collected using web crawling. The dataset is comprised of 13 classes (7 civilian vessels and 6 warships) in varying illumination, viewpoints and locations. However, the detection models developed using this dataset are not tested in real environments. The ABOShips dataset [8] is comprised of 9,880 images with 41,967 objects from 9 types of vessels and miscellaneous floaters and seamarks. Their dataset is collected for 13 days on a single route in Full HD. In comparison, our dataset is collected for a much longer duration and from multiple locations, at much higher resolution (ranging from 1920 × 1080 to 7680 × 1408). The VIAS dataset [18] comprises paired visible and infrared ship images consisting of 1623 visible images from 15 categories and thus is limited in quantity.

The Marvel dataset [6] comprises 2 million images from 109 vessel classes collected from the *Shipspotting* website. The primary purpose of the dataset is limited to image classification because the images are typically of vessels in an ideal situation (close view, simple background and clear weather). The images are not representative of real scenerios with varying scale of objects and challenging weather and illumination conditions. The GLSD dataset [16] includes 140,616 annotated objects from 100,729 images. Some of the images are collected by deploying a video monitoring system while the rest are collected using web crawling. Although the web images are diverse, they are not similar to the views and challenges observable when collecting on a moving vessel. In contrast, all of the images in our dataset are collected from cameras mounted on moving vessels and contain objects of varying size (less than 10 pixels to greater than 1500 pixels) as well as complex backgrounds and illuminations.



(a) East Coast, USA

(b) Boston Harbor, USA

(c) Europe

Figure 2. Data collection geographical locations.

Source	Images	Total objects	Unique objects	Total attributes	Image resolution	Field of view
Source 1	735,896	5,832,068	147,884	8.04M	1920×1080	60°
Source 2	675,210	6,256,429	118,411	8.26M	3840×2160	90°
Source 3	245,605	689,487	20,221	1.43M	7680×1408	180°
Source 4	105,864	492,166	9,256	0.81M	1920×1080	60°
Source 5	78,111	715,428	15,849	0.95M	3840×2160	90°
Source 6	61,929	299,667	9,105	0.49M	1920×1080	60°
Source 7	14,371	262,483	6,661	0.30M	3840×2160	90°
Source 8	9,706	97,942	2,925	0.13M	3648×2052	180°
Source 9	3,481	7,341	162	0.02M	1920×1080	180°
All	1.9M	14.6M	0.33M	20.42M	variable	variable

Table 2. SeaSAw Dataset using RGB images. The dataset is collected from 9 different sources from 5 different camera types from 6 vessels.

In summary, as compared to the above datasets, the SeaSAw dataset presented in this paper:

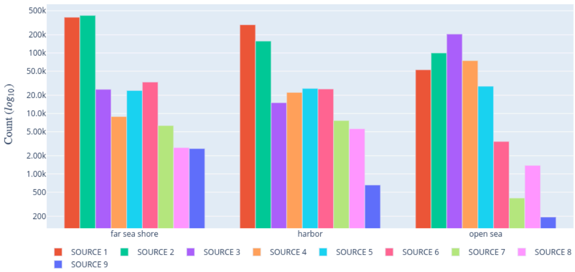
- is much larger: 1.9M images, 14.6M maritime objects (0.33M unique objects) with 12 classes.
- has been collected with cameras mounted on different vessels in various geographical locations on the East Coast, USA and Europe.
- covers 3 different locations: harbor, far sea shore and open sea.
- has 3 varying illuminations: day, dusk/dawn and night.
- includes 5 different weather conditions: sunny, cloudy, rain, fog and snow.
- has high-resolution images that capture small objects at larger distances resulting in much larger scale variations (less than 10 pixels to greater than 1500 pixels).

3. Sea Situational Awareness (SeaSAw) Dataset

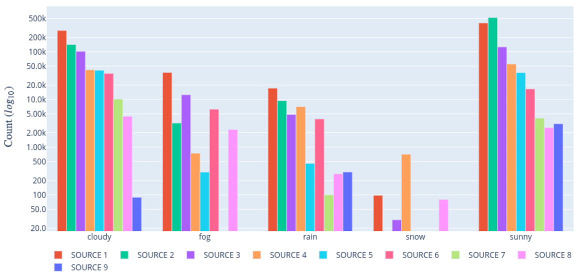
3.1. Data Collection

A data collection system is built and equipped onboard vessels to collect all sensor data. The data collection hardware is designed to connect all sensors to the computing unit (i.e. the server) and data storage unit (i.e. storage hardware). It is also equipped with a 4G connection device to support remote debugging and system status monitoring. The data collection software streams and synchronizes data from the sensors and stores the synchronized result in the storage unit. Finally, the real-time monitoring unit logs the status of sensors, data streaming, and storage, and overall system health with daily notifications sent over the 4G network.

Figure 3. Sea Machines Robotics Fleet used for data collection.

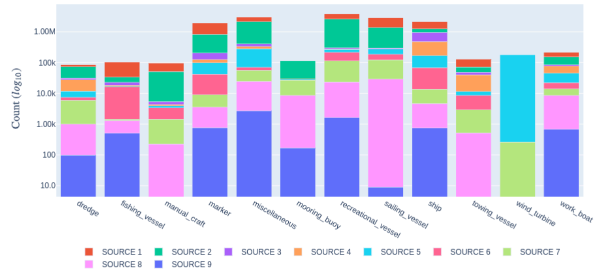


(a) Location distribution

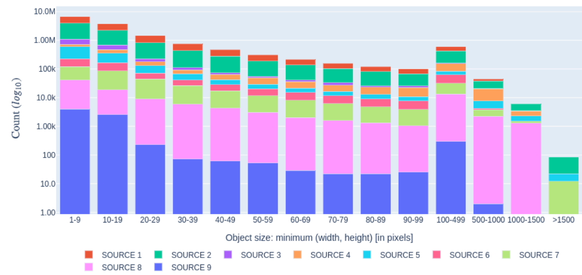


(b) Weather distribution

Figure 4. SeaSAw Dataset: global attributes distribution



(a) Class labels distribution



(b) Object size distribution

Figure 5. SeaSAw Dataset: object distribution

Our data collection fleet consists of 5 different camera types mounted across 6 different vessels. Figure 3 shows 4 of the 6 vessels used for data collection. Table 2 shows the different resolutions and fields of view of these cameras. The vessel fleet is distributed globally across diverse geographic locations (Figure 2). Data collection trips (uninterruped sequences of data collection between data offloads from vessels) varied in duration from days to months. Given constraints posed by limited on-board data storage during longer collection trips, data was collected as both individual frames at a low frame rate and video in a real-time frame rate.

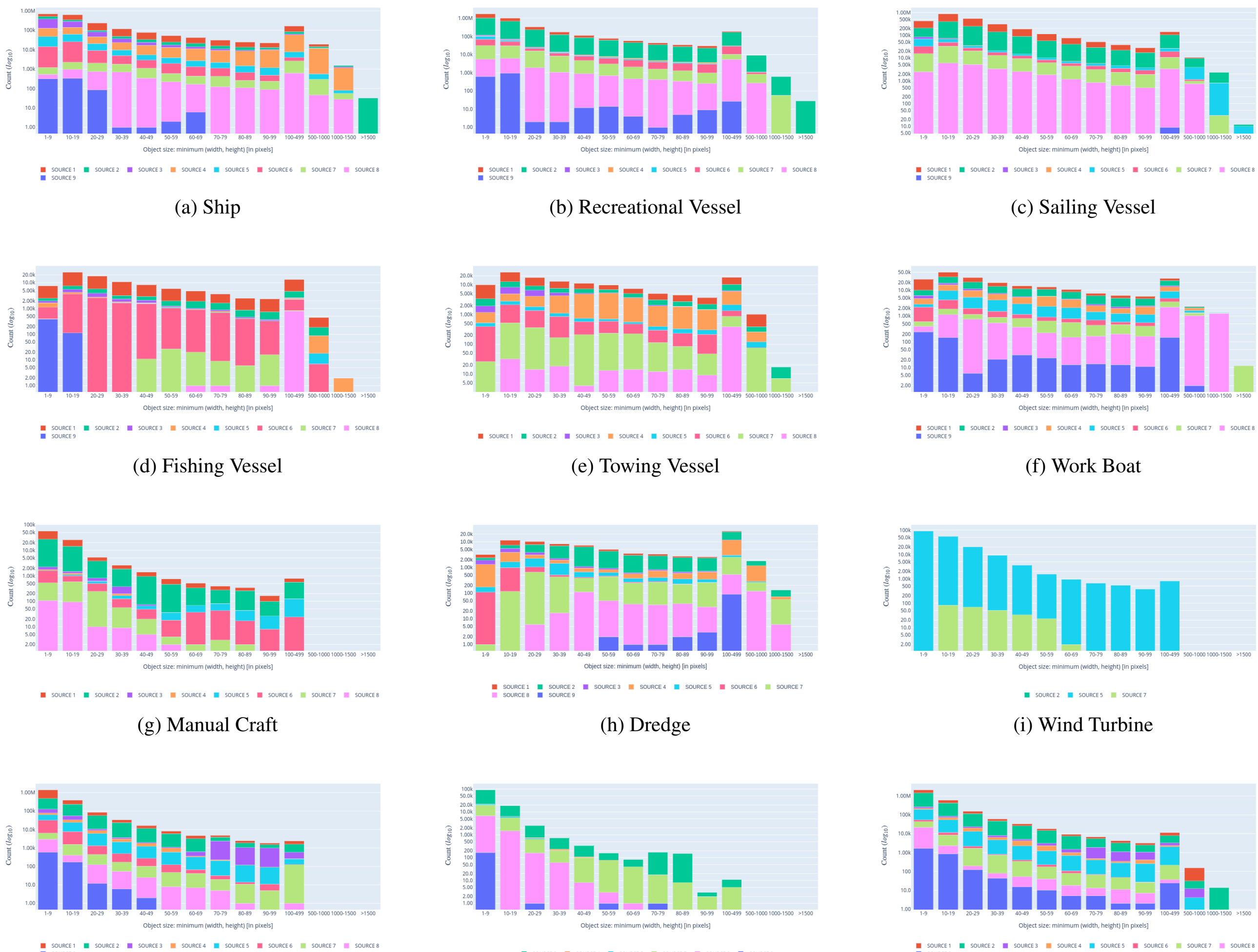
In total, 179 TB of raw image data were collected from our data collection fleet over a period from 2018 to 2022. Subsets from these raw data were coarsely selected for human annotation, and then were further curated for the SeaSAw Dataset. To select the subsets, we performed automated filtering of images in the raw data that were corrupted with aberrations such as banding or illumination artifacts. Additionally, we removed redundant data to avoid overrepresentation of certain locations and/or classes. For instance, subsets are selected with consideration that they are not dominated by images that contain few to no objects.

3.2. Data Management

To aid in the management and selection of data, an internal data management platform is designed and built on Google Cloud Platform (GCP). It consists of cloud-managed storage, automated data pipelines, and associated tooling to support the data annotation process and assist in analyzing, querying, and accessing the data.

During normal vessel operation, network bandwidth is limited (e.g. 4G or satellite connection) and cannot support large volume data transfers. Thus, raw data stored by the data collection system is physically offloaded from the vessel to a data ingestion point (i.e. an internet-enabled machine) from which data is uploaded to the cloud platform. Automated data pipelines serve to perform post-processing steps after data is uploaded to cloud-managed storage.

The pipelines serve to select useful data for annotation using image processing and machine learning-based methods. Given that data collection is performed over days and months in real operational environments, large subsets of the raw data can be redundant or uninformative. Some data can also be corrupt or invalid. For instance, when vessels are shipping in open seas, it is possible to have no other vessels in the camera field of view for hours or days; when a vessel is docked at the port for loading/offloading cargo, the camera will be oriented toward the same scene for several days. These real operational scenarios add more complexity in selecting the most salient data for annotation.



(a) Ship (b) Recreational Vessel (c) Sailing Vessel

(d) Fishing Vessel (e) Towing Vessel (f) Work Boat

(g) Manual Craft (h) Dredge (i) Wind Turbine

(j) Marker (k) Mooring Buoy (l) Miscellaneous

Figure 6. Distribution of object size by class labels

3.3. Data Annotations

Image frames selected by the cloud data pipelines are grouped as batches of 50 to 100 consecutive frames at 1 frame per second (FPS) to submit for annotation. We use a third-party annotation service, ScaleAI [1], for human annotations. The minimum size of objects to be annotated is 5 pixels. The annotations consist of two categories of attributes. First, global image attributes are determined for the entire batch and assigned to each image in the batch. Second, object attributes are assigned to individually annotated objects. Each image is annotated with three global attributes: location, weather and illumination. Location consists of harbor, open sea and far sea shore (Figure 1a); weather consists of fog, rain, snow and sunny; illumination consists of day, night and dusk/dawn. In addition to global attirbutes, each annotated object is assigned: a *bounding box*, a *unique id* representing the same object across multiple frames in a batch and one of the 12 *class labels* (8 vessel classes: dredge, fishing vessel, manual craft, recreational vessel, sailing vessel, ship, towing vessel and work boat; 4 non-vessel classes: wind turbine, marker, mooring buoy and miscellaneous).

In all, the dataset consists of 14.6M objects (0.33M unique objects) from 1.9M images. There are a total of 20.42M attributes, including global attributes and class labels across the entire dataset. The frames that do not have any objects are removed while creating this dataset.

3.4. Discussion

Our annotated data is collected from 9 sources and consists of combinations of geographical locations, vessels and cameras. Table 2 lists the different sources and summary

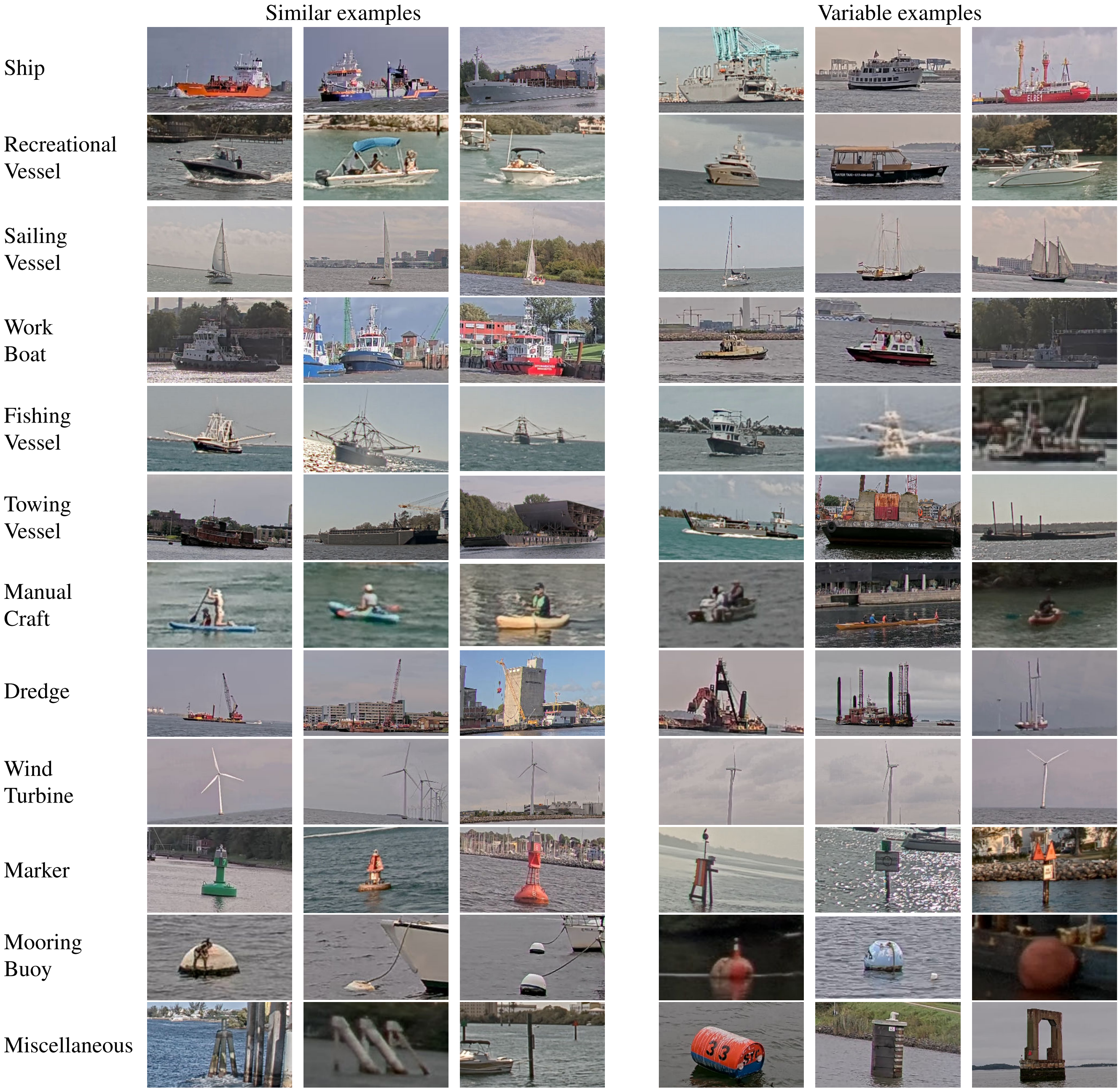


Figure 7. SeaSAw Dataset: Examples of each class. Intra class variability (scale and variety) adds challenges for object detection, classification and tracking.

statistics. Across data from these different sources, we analyze the distributions of different attributes: weather, location, camera field of view, vessel class and size. To account for the large variability in several dimensions across the data sources (Table 1), we use log_{10} scale for better understanding and representation of various distributions.

Location variations: Our data is collected in multiple geographical locations (Figure 2) covering harbor, far sea shore and open sea. Figure 4a shows that the location distribution for the entire dataset varies across different sources. Among the 9 sources, *Source 3* is dominantly in the open sea location, where as *Source 1* and *Source 2* are dominated by harbor and far sea shore locations. Figure 1a shows typical examples of various locations. The harbor location has challenging background behind vessels and may contain a high density of vessels in the frame. On the other extreme, open sea frames tend to be sparse with small vessels on the horizon with simple background.

Figure 8. Example of annotated frame.

Weather variations: Data collection was peformed over multiple years which allowed us to capture data during extreme weather conditions. These real-world navigation challenges are not present in other datasets nor easily found in public media. For all sources, the weather distribution is predominantly in *sunny* (60.5%) and *cloudy* (34%) conditions but also includes 5.5% of the data (105k images) in *rain*, *fog* and *snow* (Figure 4b). Figure 1b shows some examples where fog and rain can occlude objects in the scene.

Illumination variations: In addition to weather being a factor in illumination, sunlight angle, camera direction, sea states and vessel movement also result in varied illumination in images. Figure 1c demonstrates examples of illumination variation in dataset images with sharp glare, dull light and reflections at night.

Camera FoV variations: Our dataset consists of data from different cameras with varying fields of view and resolutions. Figure 1e shows examples of images from 90° and 180° cameras that capture different scales and perspectives of objects.

Class variations: Although our data collection is conducted in several different areas on different vessels and routes, some vessel classes are not easy to collect without specific efforts to locate them. This discrepancy is reflected in the long-tailed nature of the attribute distributions. For example, there are many vessels that are classified and annotated as ship and recreational vessel (>1M) as compared to dredge (<100k) as shown in Figure 5a. We also observe intra-class variability where each object class has diverse appearance that can be difficult to recognize by annotators. Figure 7 shows some examples of each class label and the variability captured by each of them. Futhermore, different viewpoints of vessels are also included (Figure 1d). In addition to the vessels and markers, the dataset also includes a variety of mooring buoys and miscellaneous objects on the water that need to be avoided during navigation.

Size and angle variations: In real operational scenarios, objects can be seen from different angles and distances making it difficult to identify the object class at a distance (see Figure 1f). Figure 5b shows the distribution of all objects by their bounding box size in pixels. Object sizes are binned as the minimum of the object height and width (in pixels). We observe that the distribution is long-tailed with respect to the object size. For instance there are $6.5M$ objects that are <10 pixels in size and $6k$ objects that are >1000 pixels. In all, 70% of the data is <20 pixels in size. Small objects pose several challenges for computer vision tasks such as object detection, classification and tracking and this dataset is valuable for both development and evaluation of these methods. Furthermore, the object size distribution for each class from different sources is presented in Figure 6. Each class label has a different distribution depending on the data source, emphasizing the importance of collecting data in different locations. Some unique class labels such as *wind turbines* are limited to certain sources because of their geographical presence.

In all, the dataset includes a variety of class labels from several different sources in different scenarios (location, weather and illumination) as well as several different viewpoints of vessels at different scales, making it the largest and most diverse maritime dataset for perception tasks.

4. Conclusion

In this paper we presented the Sea Situational Awareness (SeaSAw) dataset, which is the largest maritime dataset for the purpose of object detection, classification and tracking in the marine domain. It is comprised of $1.9M$ images with $14.6M$ objects from 12 classes and a total of $20.4M$ attributes. The dataset is collected using different cameras mounted on moving vessels at resolutions, varying from 1920×1080 to 7680×1480, in several geographical locations in the USA and Europe. The different location, weather and illumination conditions in addition to the long-tailed distribution of the class labels and object sizes make it a challenging dataset relative to existing maritime computer vision datasets.

Acknowledgements

This work has been supported by Sea Machines Robotics Inc.'s crew and its partners through many years of effort. We thank our colleagues in Hamburg (Europe): Peter Holm, James Miller and Oliver Arnold, who developed and installed data collection systems and managed data collection operations in Europe. We thank the captains: Ben Galdo, Lauren Lamm, Steven Turano, Arthur Seaman and William Powers at Sea Machines Robotics for operating the data collection fleet. We thank Vanya Banjac, Moran David and Trevor Vieweg for reviewing the manuscript and providing valuable feedback. Finally, we are immensely grateful to the leadership team at Sea Machines Robotics for their support in publishing this work for the computer vision community.

References

[1] Scale AI: The Data Platform for AI. https://scale.com, accessed 2022-04-20. 6

[2] Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. *International journal of computer vision*, 88(2):303–338, 2010. 3

[3] Fahimeh Farahnakian, Mohammad-Hashem Haghbayan, Jonne Poikonen, Markus Laurinen, Paavo Nevalainen, and Jukka Heikkonen. Object detection based on multi-sensor proposal fusion in maritime environment. In *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 971–976. IEEE, 2018. 1

[4] Fahimeh Farahnakian and Jukka Heikkonen. Deep learning based multi-modal fusion architectures for maritime vessel detection. *Remote Sensing*, 12(16):2509, 2020. 1

[5] Gregory Griffin, Alex Holub, and Pietro Perona. Caltech-256 object category dataset. 2007. 3

[6] Erhan Gundogdu, Berkan Solmaz, Veysel Yücesoy, and Aykut Koc. Marvel: A large-scale image dataset for maritime vessels. In *Asian Conference on Computer Vision*, pages 165–180. Springer, 2016. 3

[7] Mohammad-Hashem Haghbayan, Fahimeh Farahnakian, Jonne Poikonen, Markus Laurinen, Paavo Nevalainen, Juha Plosila, and Jukka Heikkonen. An efficient multi-sensor fusion approach for object detection in maritime environments. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pages 2163–2170. IEEE, 2018. 1

[8] Bogdan Iancu, Valentin Soloviev, Luca Zelioli, and Johan Lilius. Aboships-an inshore and offshore maritime vessel detection dataset with precise annotations. *Remote Sensing*, 13(5):988, 2021. 3

[9] Miao Kang, Kefeng Ji, Xiangguang Leng, and Zhao Lin. Contextual region-based convolutional neural network with multilayer fusion for sar ship detection. *Remote Sensing*, 9(8):860, 2017. 1

[10] Urška Kanjir, Harm Greidanus, and Krištof Oštir. Vessel detection and classification from spaceborne optical images: A literature survey. *Remote sensing of environment*, 207:1–26, 2018. 1

[11] Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. 2009. 3

[12] Sung-Jun Lee, Myung-Il Roh, Hye-Won Lee, Ji-Sang Ha, Il-Guk Woo, et al. Image-based ship detection and classification for unmanned surface vehicle using real-time object detection neural networks. In *The 28th International Ocean and Polar Engineering Conference*. International Society of Offshore and Polar Engineers, 2018. 1

[13] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014. 3

[14] Filippo Lorenzon, Michael Tsimplis, Spyridon Papadas, and Meixian Song. International regulations for preventing collisions at sea 1972, as amended (colreg 1972) and stm. 2017. 1

[15] Dilip K Prasad, Deepu Rajan, Lily Rachmawati, Eshan Rajabally, and Chai Quek. 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, 2017. 3

[16] Zhenfeng Shao, Jiaming Wang, Lianbing Deng, Xiao Huang, Tao Lu, Ruiqian Zhang, Xianwei Lv, Qing Ding, and Zhiqiang Wang. Glsd: The global large-scale ship database and baseline evaluations. *arXiv preprint arXiv:2106.02773*, 2021. 3

[17] Zhenfeng Shao, Wenjing Wu, Zhongyuan Wang, Wan Du, and Chengyuan Li. Seaships: A large-scale precisely annotated dataset for ship detection. *IEEE Transactions on Multimedia*, 20(10):2593–2604, 2018. 3

[18] Mabel M Zhang, Jean Choi, Kostas Daniilidis, Michael T Wolf, and Christopher Kanan. Vais: A dataset for recognizing maritime imagery in the visible and infrared spectrums. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 10–16, 2015. 3

[19] Yitong Zheng and Shun Zhang. Mcships: A large-scale ship dataset for detection and fine-grained categorization in the wild. In *2020 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6. IEEE, 2020. 3

[20] Matthijs H Zwemer, Rob GJ Wijnhoven, and Peter HN de With. Ship detection in harbour surveillance based on large-scale data and cnns. In *VISIGRAPP (5: VISAPP)*, pages 153–160, 2018. 3