



Article

# Maritime Object Detection by Exploiting Electro-Optical and Near-Infrared Sensors Using Ensemble Learning

Muhammad Furqan Javed <sup>1</sup>, Muhammad Osama Imam <sup>1</sup>, Muhammad Adnan <sup>2,\*</sup>, Iqbal Murtza <sup>1,3</sup>  
and Jin-Young Kim <sup>4</sup>

<sup>1</sup> Department of Creative Technologies, Faculty of Computing & AI, Air University Islamabad, Islamabad 44230, Pakistan

<sup>2</sup> Department of Technology and Safety, UiT The Arctic University of Norway, 9019 Tromsø, Norway

<sup>3</sup> Education and Research Center for IoT Convergence Intelligent City Safety Platform, Chonnam National University, Gwangju 61186, Republic of Korea

<sup>4</sup> Department of Intelligent Electronics & Compute Engineering, Chonnam National University, Gwangju 61186, Republic of Korea

\* Correspondence: muhammad.adnan@uit.no

**Abstract:** Object detection in maritime environments is a challenging problem because of the continuously changing background and moving objects resulting in shearing, occlusion, noise, etc. Unluckily, this problem is of critical importance since such failure may result in significant loss of human lives and economic loss. The available object detection methods rely on radar and sonar sensors. Even with the advances in electro-optical sensors, their employment in maritime object detection is rarely considered. The proposed research aims to employ both electro-optical and near-infrared sensors for effective maritime object detection. For this, dedicated deep learning detection models are trained on electro-optical and near-infrared (NIR) sensor datasets. For this, (ResNet-50, ResNet-101, and SSD MobileNet) are utilized in both electro-optical and near-infrared space. Then, dedicated ensemble classifications are constructed on each collection of base learners from electro-optical and near-infrared spaces. After this, decisions about object detection from these spaces are combined using logical-disjunction-based final ensemble classification. This strategy is utilized to reduce false negatives effectively. To evaluate the performance of the proposed methodology, the publicly available standard Singapore Maritime Dataset is used and the results show that the proposed methodology outperforms the contemporary maritime object detection techniques with a significantly improved mean average precision.

**Keywords:** maritime object detection; ensemble learning; transfer learning; Singapore maritime dataset; electro-optical images; near-infrared images



**Citation:** Javed, M.F.; Imam, M.O.; Adnan, M.; Murtza, I.; Kim, J.-Y. Maritime Object Detection by Exploiting Electro-Optical and Near-Infrared Sensors Using Ensemble Learning. *Electronics* **2024**, *13*, 3615. <https://doi.org/10.3390/electronics13183615>

Academic Editor: Heung-Il Suk

Received: 5 August 2024

Revised: 3 September 2024

Accepted: 9 September 2024

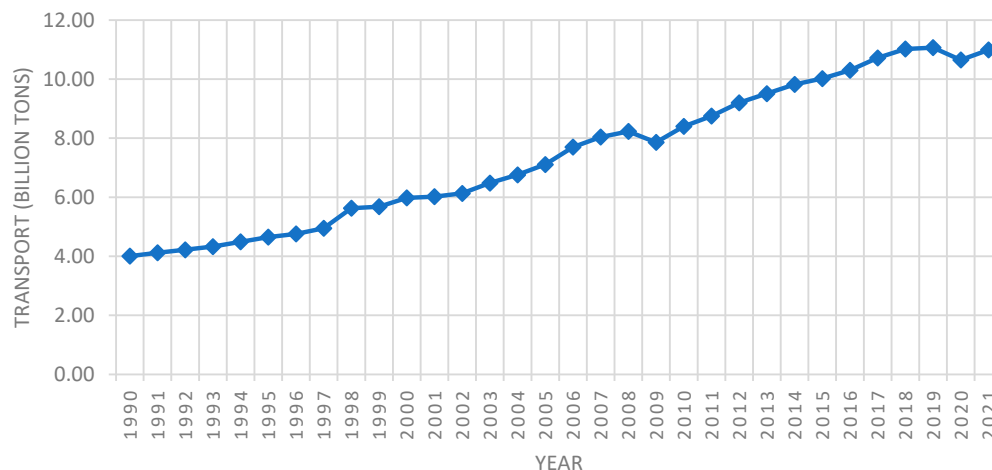
Published: 11 September 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The world's oceans, seas, and waterways serve as crucial tools for global commerce, communication, and resource exploration. However, the surge in global maritime traffic, escalating demand for offshore resources, and heightened risks associated with navigation underscore the critical need for robust maritime safety and security measures [1,2]. The volume of global seaborne trade has witnessed remarkable growth, doubling between 1990 and 2021, with transported goods increasing from around 4 billion tons to an estimated 11 billion tons [3]. This surge underscores the pivotal role of maritime transportation in global trade, as shown in Figure 1. Effective maritime object detection and tracking systems are necessary to ensure safe vessel passage and coastal area protection.



**Figure 1.** Trade volume from 2010 to 2020 [3].

Traditionally, maritime surveillance heavily relied on human operators and radar-based technologies. While effective, these methods face limitations in handling the complexities of the evolving maritime environment, especially in bad weather circumstances, increased sea traffic, low visibility, and when considering the necessity for continuous, real-time monitoring. Integrating deep learning and traditional computer vision techniques is crucial in maritime object detection and tracking to handle these challenges and enhance maritime navigation [4,5]. Some challenges found in our literature review are as follows:

- Accurate detection and classification of relatively small objects.
- Classification of maritime objects in different weather conditions.

In the problem of object detection, deep learning has gained attention for its capability to learn and extract complex features from data automatically. Coupled with computer vision, which enables machines to interpret visual information, deep learning has the potential to revolutionize the identification and tracking of maritime objects, including ships, buoys, fishing vessels, and potential threats such as pirate ships or drifting debris even with the aforementioned challenges [6,7].

To resolve this problem, this study uses the Singapore Maritime Dataset (SMD), a rich and diverse maritime data collection, to explore advanced object detection and tracking approaches. Our study explores various methodologies, models, and algorithms by using artificial intelligence to detect and track objects in maritime environments, contributing to improving operational and navigational aids, maritime security, and environmental impact on navigation.

This study investigates the challenges associated with maritime object detection and tracking, encompassing environmental influences, electro-optical and NIR sensors, object diversity, and the demand for real-time processing. Deep learning (DL) models like SSD Mobile Net and Faster RCNN are explored. It also investigates how different sensors can enhance the precision of object detection approaches.

In conclusion, this research study comprehensively explores the advancements in maritime object detection and tracking using DL and computer vision techniques. By using the potential of these technologies, the aim is to contribute to developing more effective and reliable systems for enhancing maritime safety, security, and efficiency, ultimately ensuring the sustainable use of our oceans and waterways.

## 2. Related Work

Maritime object detection is fundamental for safely navigation of vessels in maritime environment. In order to obtain a comprehensive understanding of this field, the relevant traditional and deep learning techniques are presented as in the following subsections.

### 2.1. Traditional Computer Vision Approaches

In the initial stages of maritime object detection, traditional computer vision techniques have been the main focus in research and development. Various studies have contributed towards this line of research. For example, Negahdaripour et al. [8] anticipated a vision-based system for Autonomous Underwater Vehicles (AUVs) and Remotely Operated Vehicles (ROVs), with features such as real-time navigation, positioning, and video mosaicking of seafloor images obtained. Similarly, Cozman et al. [9] utilized sun altitude measurements for robot localization in completely unfamiliar environments. This proposed approach uses celestial observations in order to enable a robot to determine its position on earth, even when conventional methods are not available or functional.

Moreover, Prasad et al. [4] examined the key problems encountered in maritime object detection and computer vision for video produced by cameras. Even well-documented problems in videos, like detection of the horizon line and frame registration, proved to be quite challenging in maritime scenarios. More advanced issues like background subtraction and object detection in video streams were also complicated due to the continuously changing background, the absence of static signals, the existence of small objects in distant backgrounds, and the effects of changing illumination conditions. The paper discussed these challenges, emphasizing their significance and implications for maritime computer vision applications.

Another study by González-Sabbagh and Robles-Kelly [10] presented a comprehensive survey of the evolving domain of maritime computer vision. The authors reviewed approaches involving image formation and image processing approaches to enhance or correct underwater imagery. The primary motivation behind these efforts is to attain photometric invariance, like shape retrieval and object identification, even in the face of the challenging and variable imaging conditions encountered in underwater scenes.

### 2.2. Deep Learning Advancements

With the advent of DL, the maritime surveillance landscape underwent a transformative shift. Numerous studies have used deep neural networks to address the limitations of traditional techniques. For example, the paper by Prasad et al. [11] explored the technical challenges inherent in object detection within the maritime context using videos, employing computer vision techniques. Image processing and computer vision techniques were applied to video streams to detect an array of obstacles. Notably, even fundamental problems like horizon detection and frame registration become challenging within maritime scenarios. More advanced challenges include background subtraction and object detection within these video streams. These difficulties stem from the changing nature of maritime backgrounds, absence of static reference points, presence of small objects against distant backgrounds, and influence of varying illumination conditions, all of which are thoroughly explored in this paper.

Furthermore, Moosbauer et al. [12] addressed the relatively underexplored field of object detection in the maritime domain, particularly within the field of computer vision. In this context, comprehensive public benchmarks were absent, starkly contrasting with the well-established benchmarks available for object detection in automotive applications. The paper introduced a benchmark framework built upon the Singapore Maritime Dataset (SMD) to fill this gap. The Faster R-CNN and Mask R-CNN models were applied to the dataset for testing and training. Both of these models performed good on this dataset; however, Mask RCNN outperformed Faster RCNN.

Kim et al. [13] applied state-of-the-art YOLO-V5 architecture on an improved maritime dataset, SMD-Plus. SMD suffers from noisy labels and imprecise bounding box annotations. The authors created a more accurate and refined dataset called SMD-Plus to address the issues found in SMD. Authors first corrected the annotations, and then an augmentation technique utilizing the "Online Copy & Paste" method was used for balancing class data. The experimental results showed an increase in the performance for detection and classification using YOLO-V5 on the SMD-Plus dataset.

Zhao et al.'s [14] paper, titled "YOLOv7-Sea: Object Detection of Maritime UAV Images Based on Improved YOLOv7", addressed the role of object detection algorithms in maritime search-and-rescue missions, particularly in detecting individuals, ships, and various items in open water environments. Conversely, these presented unique challenges due to the SeaDronese dataset's characteristics, such as small targets and substantial sea surface interference. The researchers proposed an improved object detector called YOLOv7-sea. They incorporate the Simple, Parameter-Free Attention Module (SimAM) to identify interest regions within the view. The experimental results, conducted on the ODv2 challenge dataset, demonstrate the effectiveness of YOLOv7-sea, with an average precision (AP) result of 59.00%. This represents a notable improvement of approximately 7% compared to the baseline model (YOLOv7).

Yu et al. [15] aimed for an approach to automating the finding of underwater objects using side-scan sonar (SSS) pictures. Integration of a Transformer module and YOLO v5, referred to as "TR-YOLOv5s", was performed for automatic target recognition. Similarly, Shin et al. [16] used a data augmentation technique and employed horizon information for extracting ship data for object recognition. For classification, they utilized ResNet, RetinaNet, and Mask R-CNN models pre-trained on the MSCOCO.

Moreover, Haghbayan et al. [17] consider the multi-sensor combination method to address the challenge in object detection. In addition, they introduce an efficient multi-sensor fusion method grounded in probabilistic data association. This method ensembles data from sensors—radar, LiDAR, RGB camera, and infrared camera—for enhanced object recognition and tracking precision. It generates object region proposals through combined detection outcomes and employs a CNN model trained on real maritime data to classify object categories within these regions. Experimental results based on datasets collected from a ferry affirm the efficacy of the proposed approach. The fusion technique improves object detection rates and mitigates false positives, even under demanding maritime conditions characterized by noisy radar data and LiDAR reflections. Furthermore, the CNN-based classification exhibits high accuracy across all object categories, underscoring the strength of the anticipated approach.

Rekavandi et al. [6] discussed the issue of smaller object detection from the video and optical images data in their study, a task where advanced general object detection techniques struggle to locate and recognize such items accurately. Minor objects typically arise in real practical scenarios owing to a substantial camera-object space, resulting in these objects occupying only a tiny portion of the entered image (often less than 10%). As a result, the data available from a limited spot may not be sufficient for effective decision-making. This paper offers a comprehensive survey of this evolving area. They also summarized the existing literature and presented a taxonomy to overview current research trends comprehensively. The focus is on improving minor object detection in the maritime domain, where precision detection is of utmost importance. The paper establishes connections between general and maritime research and identifies future research guidance. It also discusses popular datasets used for SOD in both generic and maritime contexts and provides assessment metrics for advanced methods on these datasets.

Cane et al. [18] proposed a comprehensive evaluation of deep semantic segmentation networks for maritime observation. This investigation addressed the inherent challenges associated with visually detecting objects in maritime environments, characterized by their diverse and unpredictable nature and the requirement for real-time performance. They trained various semantic segmentation network architectures using the ADE20k scene parsing dataset and assessed their capabilities across publicly available maritime surveillance datasets.

The maritime environment is dynamic in nature particularly due to changing weather conditions, waves, and a lack of point of reference. Unmanned aerial vehicles (UAVs) present real-time video streams of areas in which they are flown. Vasilopoulos et al. [19] used machine learning algorithms to detect and track objects. This study experimented using a UAV's on-board computation unit. These experiments were used to find the

efficiency, accuracy, and speed of the used algorithms. Furthermore, they proposed the solution for training datasets for optimization of both efficiency and accuracy of object detection using a UAV at sea.

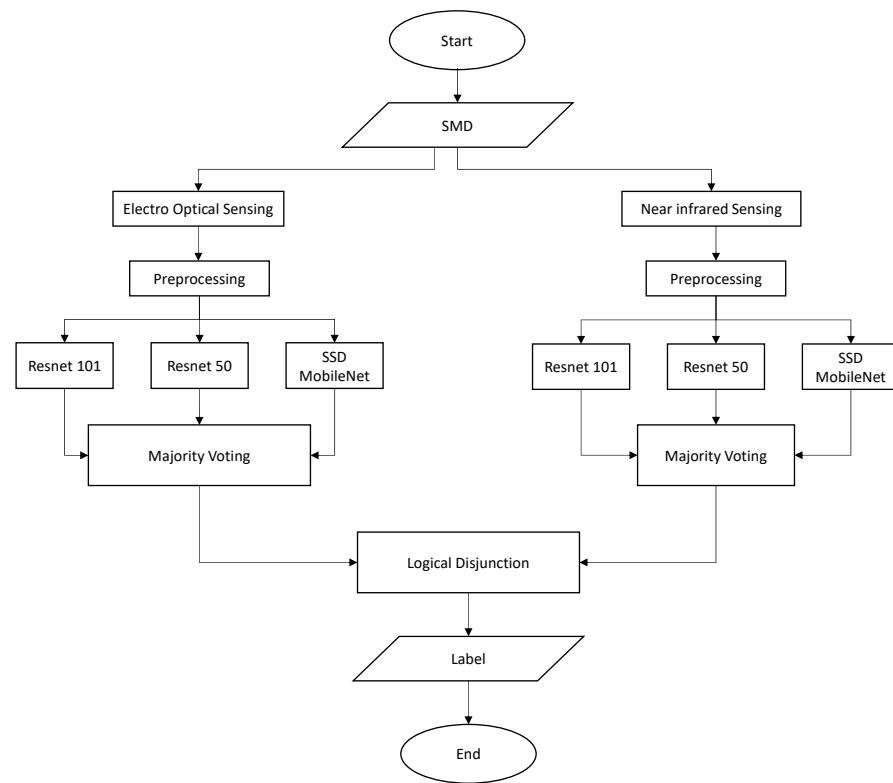
Similarly, Iancu et al. [20] also worked on improving the object detection using water-based imagery. This study was aimed at the development of efficient and accurate autonomous systems using improved object detection. Furthermore, domain-specific datasets for maritime domain object detection are scarce and typically have a restricted number of images and annotations. This paper fill this gap by re-labeling the ABOships dataset and conducting a benchmarking study using a proposed detector called Centernet on the recently annotated dataset, ABOships-PLUS. The study explored Centernet's performance with unique feature separators and investigated how the size of the object and inter-class variability affect recognition precision. The results demonstrate the suitability of the ABOships-PLUS dataset for supervised domain adaptation, with Centernet utilizing DLA (Deep Layer Aggregation) as a feature extraction method, attaining improved precision in sensing maritime objects, achieving a mAP of 74.4%.

Notably, Rahman et al. [21] evaluated 13 deep learning models based on one-stage and two-stage object detectors, which included EfficientDet, RetinaNet, YOLOv5, Fast RCNN, and Faster RCNN, on a three-class weed dataset. RetinaNet (R101-FPN) demonstrated better performance, attaining a (mAP @ 0.50) of 79.98%, emphasizing its efficacy in discerning weeds amidst cotton plants. Some authors have also used various deep learning algorithms in different types of object detection tasks [22,23]. Similarly, most recently, lightweight convolutional neural networks such as WearNet are showing remarkable classification accuracy, with WearNet achieving 94.16% [24].

In conclusion, traditional image processing techniques have laid the groundwork for research in the maritime computer vision field, but deep learning offers more sophisticated and dynamic approaches. Models such as YOLO and RetinaNet offered remarkable effectiveness in detecting and tracking objects in maritime environments. Much work has been performed to refine datasets and methodologies for maritime object detection. The majority of the work in this field was performed using images captured by UAVs and satellites. However, limited research has been performed using electro-optical and near-infrared sensory data. Similarly, integration of above sensors can also offer improved performance. This study aims to utilize electro-optical and near-infrared images separately for object detection in a maritime context. Furthermore, an ensemble learning methodology is used for enhanced performance.

### 3. The Proposed Methodology

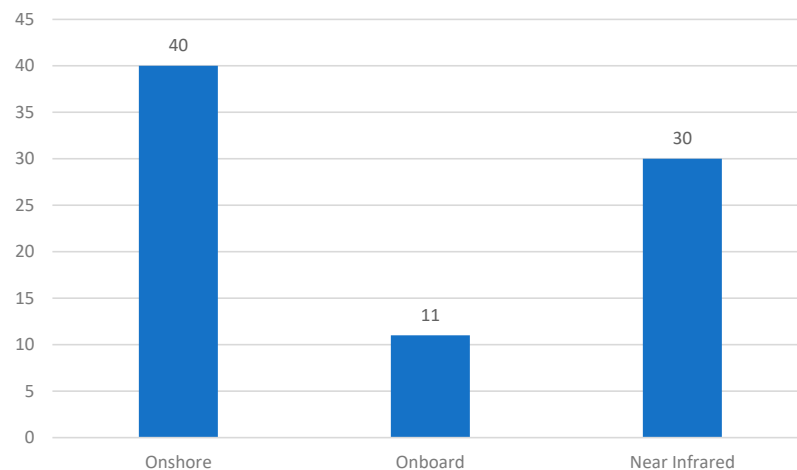
This research focuses on solving the challenge of maritime object detection under diverse weather conditions by using visible light (camera) sensors and near-infrared (NIR) imagery, representing distinct data collection modalities. A flowchart of the proposed model is illustrated in Figure 2. In the preprocessing stage, data from each modality is prepared for analysis using noise reduction and normalization. The processed data is then forwarded to these selected models—ResNet 101 and ResNet 50 for image recognition, and SSD MobileNet, optimized for mobile and embedded vision applications. Predictions from these models are combined through Majority Voting, an ensemble technique that enhances decision robustness and accuracy. The outputs from both pathways are integrated using a logical disjunction (logical OR), resulting in a final decision. Each module in this flowchart is interconnected, ensuring a systematic approach to sensor-based data analysis to solve the abovementioned challenge. These models are precisely trained separately on datasets comprising camera and NIR data, allowing them to become proficient in detecting maritime objects within their respective modalities.



**Figure 2.** Flowchart of the proposed model. From the Singapore Maritime Dataset, each sensor input is learned using ensemble classification separately, and then the output of these ensemble classifications is combined using logical disjunction.

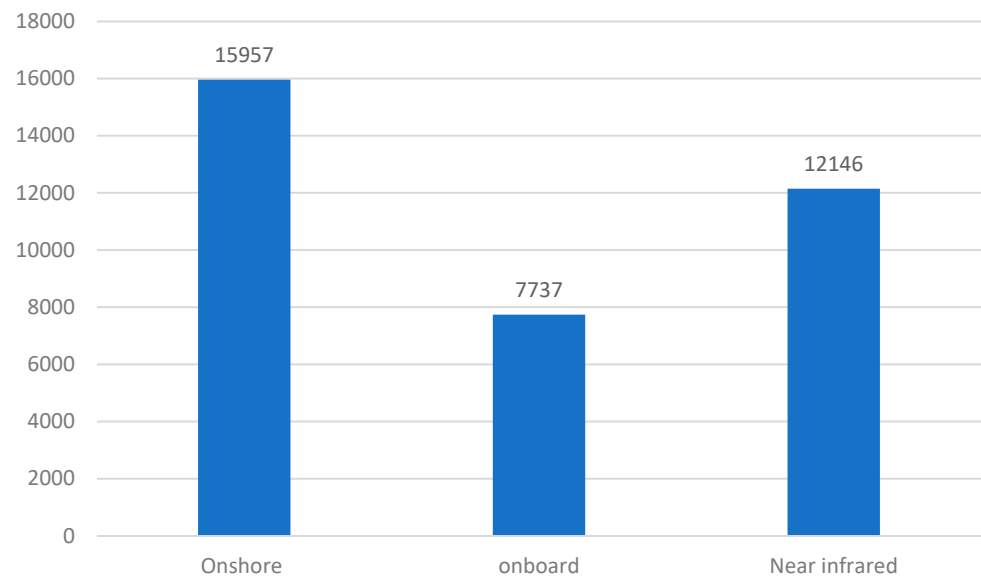
3.1. Dataset Description

The Singapore Maritime Dataset (SMD) is a comprehensive collection of video data collected using Canon 70D cameras positioned in the vicinity of Singapore’s waters [4]. The distribution of videos and frames captured through these videos, including onshore, onboard, and near-infrared environments, are shown in Figures 3 and 4. This dataset offers valuable resources for various applications and research endeavors. Here are the key features of the SMD: (1) All videos in the dataset are in high-definition, using a resolution of 1080 × 1920 pixels. (2) The SMD is divided into three distinct parts, each offering unique insights and perspectives. Sample onshore and onboard pictures from the dataset are provided in Figures 5 and 6, respectively.



**Figure 3.** Distribution of videos.





**Figure 4.** Distribution of the frames gathered from videos.

Additionally, this dataset comprises videos captured using another Canon 70D camera, depicted in Figure 7. The camera has its hot mirror removed and a Mid-Opt BP800 Near-IR Bandpass filter. NIR videos offer insights into the near-infrared spectrum, which can be valuable for various applications. The videos within the SMD are acquired at various locations and along different routes. Consequently, they capture a diverse range of maritime scenes and scenarios. Researchers and practitioners can leverage the Singapore Maritime Dataset for various applications, including object detection, tracking, and environmental analysis. The availability of onshore, onboard, and NIR videos ensures versatility in the dataset and experimentation.



**Figure 5.** Example onshore image.



**Figure 6.** Example onboard image.



**Figure 7.** Example NIR image.

### 3.2. SSD MobileNet (Single-Shot MultiBox Detector with MobileNet Backbone)

SSD MobileNet was trained separately on datasets from both domains, electro-optical and near-infrared, which ensures the in-depth training of the model. This model uses the lightweight MobileNet architecture combined with the Single-Shot MultiBox Detector (SSD) framework [25]. Using depth-wise separable convolutions, MobileNet reduces computational complexity while effectively maintaining the capacity to extract crucial image features. The SSD framework, integrated into SSD MobileNet, facilitates rapid object detection by partitioning input images into a grid of default bounding boxes at various aspect ratios and scales, allowing for the efficient detection of objects with diverse sizes and aspect ratios—training involved fine-tuning the model’s loss function to minimize classification and localization errors specific to maritime scenarios. In our methodology, SSD MobileNet was trained separately on both camera and near-infrared (NIR) data to adeptly handle maritime object detection in varying modalities. SSD MobileNet’s role in our methodology significantly boosted detection accuracy, making it an essential component for maritime object detection across diverse weather conditions and modalities.



### 3.3. Faster R-CNN (ResNet 50)

Faster R-CNN (ResNet 50) was trained separately on datasets of both domains, as electro-optical and near-infrared, which ensures the in-depth training of the model [26]. Furthermore, a layered architecture was used to take advantage of residual blocks to tackle the gradient problems; ResNet 50 provides features to use a residual block which plays a vital role in handling problems like gradients by bypassing certain layers during training in both forward and backward passes with the help of skip connections. The ResNet 50 architecture is a deep convolutional network which extracts important features and helps Faster R-CNN to construct a regional proposal network. The trainable features are from images taken in distinct weather environments, such as haze, fog, and rainy weather. The ResNet 50 model's depth increases computation complexity. However, it extracts features that help a lot in efficiently detecting and classifying objects in the maritime domain.

### 3.4. Faster R-CNN (ResNet 101)

Similarly, Faster R-CNN with ResNet-101 was also used in our methodology separately for both datasets. This incorporation of regional proposal networks (RPNs) within the Faster R-CNN architecture enables rapid generation of region proposals, providing potential boxes for objects in the image. The proposed approach focuses on a two-stage process as refining these proposals for accurate object localization and classification. Incorporating region-of-interest (RoI) pooling ensures constant feature extraction aligns region proposals to a fixed size. The Faster R-CNN model is finely adjusted to lower the classification and localization errors specific to a maritime scenario. The model detection and accuracy results make it necessary for effective maritime object detection across constantly changing weather conditions and modalities in this case.

### 3.5. Models Hyperparameter Tuning

In the Faster R-CNN case, Hyperopt was incorporated to automatically search for the optimal selection of hyperparameters [27], for example, learning rate and momentum coefficients, improving the model effectiveness, which proved to increase the performance of each individual model. The Hyperopt function selected a learning rate of 0.0003 and momentum of 0.9 with a batch size of 8 in the Faster R-CNN case. In the case of SSD MobileNet, the learning is adjusted to 0.004 and momentum to 0.9 with a batch size of 24.

### 3.6. Computational Complexity Reduction

Computational complexity is of great importance in AI tasks which are meant to be deployed in real-time scenarios or cutting-edge devices. To serve that purpose, this study experimented with some techniques like model caching which can help to reduce computational cost. Moreover, it was established that using the model caching technique resulted in the storage of the most recent output during inference [28]. This method enhances the real-time performance and lowers latency, ensuring our proposed system can seamlessly operate in challenging and constantly changing maritime scenarios. By implementing these strategies, computational overhead can be significantly reduced.

### 3.7. Ensemble Methodology

Leveraging ensemble learning techniques has been exhibited to notably enrich precision across a spectrum of systems and applications, as highlighted in [29,30]. This study adopted an ensemble method to enhance the precision and reliability of object detection in two distinct imaging domains: near-infrared (NIR) and electro-optical images. This ensemble methodology involved several vital steps.

#### 3.7.1. Dedicated Majority Voting for Each Sensor

Initially, three base object detection models are selected, namely SSD MobileNet v1 FPN, Faster R-CNN with a ResNet-50 backbone, and Faster R-CNN with ResNet-101

support. These three classifiers have been used as base learners for each sensor and thus generate a total of six bases. For each sensor, its votes are aggregated separately and these votes are combined using dedicated majority voting as shown in Figure 2.

### 3.7.2. Logical Disjunction

In the problem at hand, false negatives are associated with a worse cost compared to false positives. Thereby, output decisions of the individual ensemble for each sensor are then combined using logical disjunction. The logical disjunction has been chosen to ensure that the final decision is positive if any of the base ensemble classifications predict that it is a positive sample. Thereby, this strategy reduces false negatives with a compromise of an increase in false positives, as evident by the work of Murtza et al. [31] and as shown in Equations (1) and (2). In these equations,  $X_{FN}$  denotes the number of false negatives for the first dedicated ensemble classifier, whereas  $Y_{FN}$  denotes the sum of false negatives for the second dedicated ensemble classifier.

$$p_{LD}(FN) = \left( \frac{X_{FN}}{X_{TP} + X_{FN}} \right) \frac{Y_{FN}}{Y} \quad (1)$$

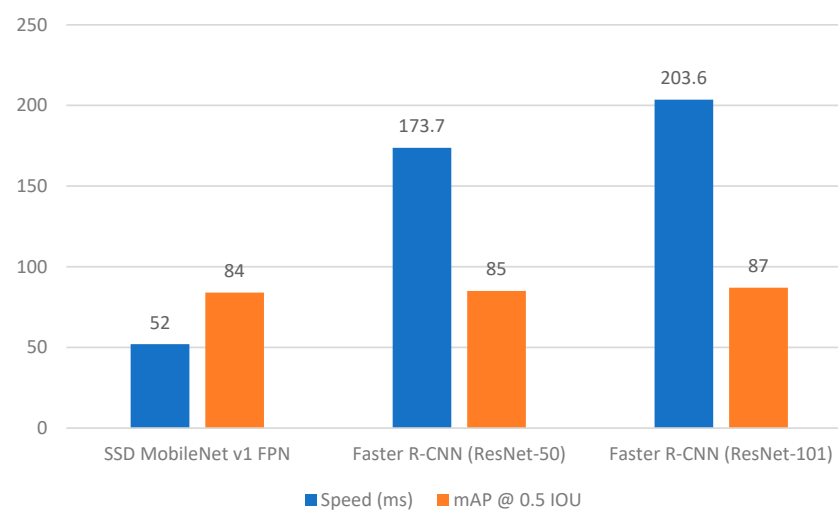
$$p_{LD}(FP) = \left( \frac{X_{FP}}{X_{FP} + X_{TN}} \right) \frac{Y_{FP}}{Y} + \left( \frac{X_{TN}}{X_{FP} + X_{TN}} \right) \frac{Y_{FP}}{Y} + \left( \frac{X_{FP}}{X_{FP} + X_{TN}} \right) \left( \frac{Y_{TN}}{Y_{TP} + Y_{FN} + Y_{TN}} \right) \left( 1 - \frac{Y_{FP}}{Y} \right) \quad (2)$$

## 4. Results and Discussion

In our comprehensive assessment of object detection models, three key models have been explored, and their performance has been evaluated across two pivotal imaging domains: near-infrared (NIR) and electro-optical. The evaluation included speed in milliseconds (ms) and mean average precision (mAP) at a 0.5 intersection over union (IOU) threshold as performance metrics.

### 4.1. Base Learners Performance on NIR Sensor

Using NIR images, all three base models demonstrated good performance. The SSD MobileNet v1 FPN model displayed a higher speed of 52 ms and higher mAP of 84. Faster R-CNN models with ResNet-50 demonstrated an mAP of 85 but took 173.7 ms per detection. This makes it suitable for applications where accuracy is essential and real-time constraints are less critical. Similarly, the Faster R-CNN model with a ResNet-101 backbone excelled in mAP with a score of 87 but had a speed of 203.6 ms, positioning it as an ideal choice for tasks demanding the highest accuracy. The results are showed in Table 1 and Figure 8 below.



**Figure 8.** Speed and mAP for NIR images.

**Table 1.** Average speed and mAP for NIR images.

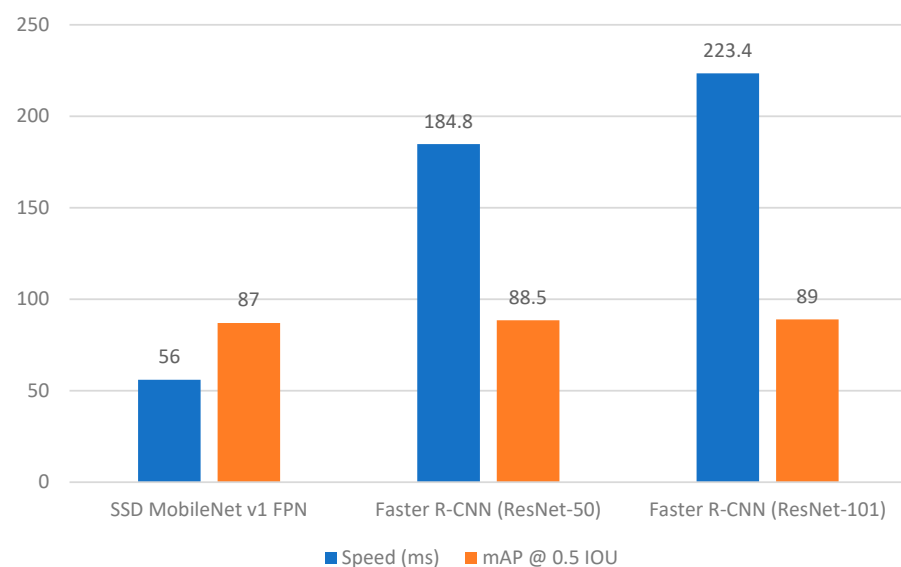
Sr. #	Model	Speed (ms)	mAP @ 0.5 IOU
01	SSD MobileNet v1 FPN	52	84
02	Faster R-CNN (ResNet-50)	173.7	85
03	Faster R-CNN (ResNet-101)	203.6	87

#### 4.2. Base Learners Performance on Electro-Optical Sensor

For electro-optical images, SSD MobileNet v1 FPN model, with a speed of 56 ms, managed to maintain a robust mAP of 87, presenting a well-balanced choice for object detection tasks in this domain. In the electro-optical domain, the Faster R-CNN model with a ResNet-50 support exhibited a speed of 184.8 ms and excelled in mAP, achieving a score of 88.5, making it a suitable choice for scenarios prioritizing accuracy over speed. Similarly, the Faster R-CNN model with a ResNet-101 backbone, although slower at 223.4 ms, further elevated accuracy with a mAP of 89, rendering it a valuable model for applications demanding precision. These results are collectively displayed in Table 2 and Figure 9.

**Table 2.** Average speed and mAP for EO images.

Sr. #	Model	Speed (ms)	mAP @ 0.5 IOU
01	SSD MobileNet v1 FPN	56	87
02	Faster R-CNN (ResNet-50)	184.8	88.5
03	Faster R-CNN (ResNet-101)	223.4	89

**Figure 9.** Speed and mAP for EO images.

#### 4.3. Ensemble Results

The problem at hand is a 08-class maritime object detection classification problem. This problem is achieved as eight different binary (one versus all) classification problems. Thereby, each binary classification problem has training data in which the number of positive samples are significantly small as compared to the number of negative (all other classes) samples. This data imbalance problem is likely to affect the prediction of the binary classifier. Unluckily, because of the smaller number of positive samples, this bias is against the prediction of positive samples.

In addition to this, a false negative means to miss a maritime object and false positive means to falsely identify a maritime object. Thereby, there is a significant cost difference

between a false negative and false positive, i.e., the cost of a false negative is significantly greater than the cost of a false positive. The ensemble classification mechanism employed in the proposed technique is basically designed to discourage false negatives with the compromise of an increase in false positives.

Table 3 demonstrates the performance metrics of the ensemble learning model for both near-infrared (NIR) and electro-optical images. The speed (ms) column indicates the average time in milliseconds required for the ensemble model to process and detect objects in images from each domain. The mAP @ 0.5 IOU column shows the mean average precision at a 0.5 intersection over union threshold, which reflects the model’s accuracy in object detection. Whereas, classification results are given in Table 4.

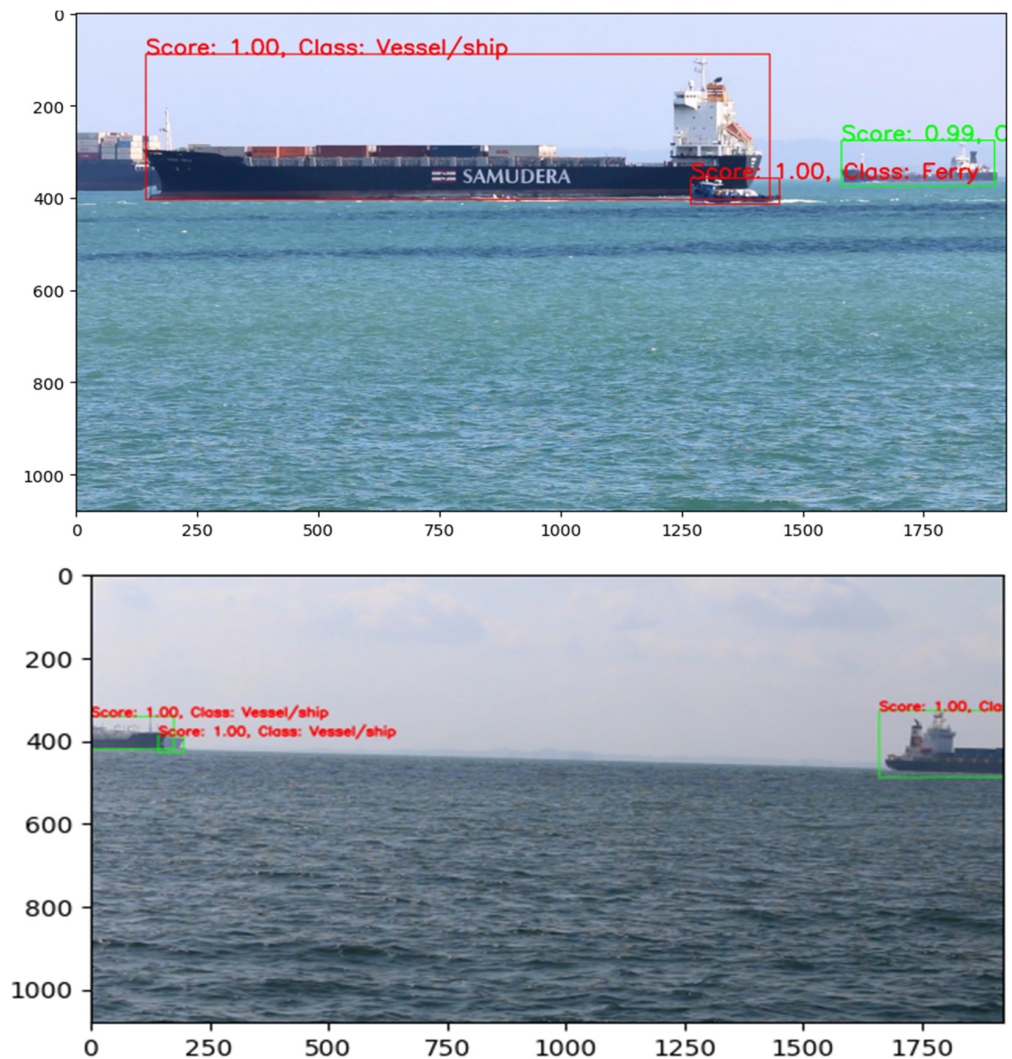
**Table 3.** Speed and mAP using ensemble learning.

Domain	Metric
NIR images	243 Speed (ms)
	91.5 mAP @ 0.5 IOU
Electro-optical images	256 Speed (ms)
	92.2 mAP @ 0.5 IOU

**Table 4.** Performance measures (confusion matrices, precision, recall, and accuracy) for each maritime object for the final logical-disjunction-based ensemble classification which is built upon individual EO and NIR ensemble classification models.

	EO Ensemble	NIR Ensemble	Logical Disjunction	EO Ensemble	NIR Ensemble	Logical Disjunction
FN	74	73	6	112	112	35
TP	651	652	719	261	261	338
FP	120	72	192	110	165	270
TN	2287	2335	2215	2649	2594	2489
Precision	84.44	90.06	78.92	70.35	61.27	55.59
Recall	89.79	89.93	99.17	69.97	69.97	90.62
Accuracy	93.81	95.37	93.68	92.91	91.16	90.26
(a) ferry			(b) Buoy			
	EO Ensemble	NIR Ensemble	Logical Disjunction	EO Ensemble	NIR Ensemble	Logical Disjunction
FN	184	340	19	38	38	0
TP	2714	2558	2879	340	340	378
FP	218	202	233	165	110	266
TN	16	32	1	2589	2644	2488
Precision	92.56	92.68	92.51	67.33	75.56	58.70
Recall	93.65	88.27	99.34	89.95	89.95	100.00
Accuracy	87.16	82.69	91.95	93.52	95.27	91.51
(c) Vessel/Ship			(d) Speed Boat			
	EO Ensemble	NIR Ensemble	Logical Disjunction	EO Ensemble	NIR Ensemble	Logical Disjunction
FN	173	173	119	66	66	3
TP	73	73	127	586	586	649
FP	86	202	284	74	173	242
TN	2800	2684	2602	2406	2307	2238
Precision	45.91	26.55	30.90	88.79	77.21	72.84
Recall	29.67	29.67	51.63	89.88	89.88	99.54
Accuracy	91.73	88.03	87.13	95.53	92.37	92.18
(e) Boat			(f) Kayak			
	EO Ensemble	NIR Ensemble	Logical Disjunction	EO Ensemble	NIR Ensemble	Logical Disjunction
FN	194	194	59	93	91	52
TP	450	450	585	66	68	107
FP	124	74	194	118	144	256
TN	2364	2414	2294	2855	2829	2717
Precision	78.40	85.88	75.10	35.87	32.08	29.48
Recall	69.88	69.88	90.84	41.51	42.77	67.30
Accuracy	89.85	91.44	91.92	93.26	92.50	90.17
(g) Sail Boat			(h) Flying Plane			

These results underscore the usefulness of the ensemble method in combining the strengths of individual models to attain a synergy of speed and precision in object recognition tasks, making it a valuable asset for various real-world applications. In addition to this, it also fixes the data imbalance problem very effectively. From these experiments, the NIR, onshore, and onboard images in Figure 10 show confidence scores along with bounding boxes of the ensemble method on the sample captured images.



**Figure 10.** Confidence score on different scenes.

#### 4.4. Discussion

In our thorough evaluation of object detection models for near-infrared (NIR) and electro-optical (EO) images, we uncovered significant insights into their performance characteristics, guiding their applicability in practical scenarios.

The results clearly demonstrate that a balance between speed and accuracy was achieved in NIR images by utilizing different base learners. The SSD MobileNet v1 FPN showed to be a more favorable choice for applications that prioritize precision and speed in NIR object detection by achieving a slightly higher speed of 52 ms and mAP of 84. The Faster R-CNN models were slower, but they achieved a good mAP score. The Faster R-CNN with ResNet 101 backbone achieved a mAP of 87, but was slower for both of them. In the EO domain, similar trends were seen in model performance. Meanwhile, the SSD MobileNet v1 FPN model maintained a robust mAP of 87 with a speed of 56 ms, making it a balanced and good choice for object recognition tasks in the maritime domain. The Faster



R-CNN models continued to perform exceptionally well, achieving an mAP score of 89 with the ResNet-101 variant, but were slower in terms of speed.

An ensemble approach was utilized to enhance the mAP of approximately 91.5% in NIR and 92.2% in electro-optical images, providing a significant improvement in detection accuracy as compared to individual models. Moreover, the ensemble approach demonstrated its effectiveness in combining the strengths of unique models to achieve a balanced combination of speed and accuracy. Choosing an object detection and classification model for a specific task requires several considerations; while some models prioritize speed, others excel in accuracy, while an ensemble approach provides a solution to balance these trade-offs. Our analysis provides practitioners with valuable insights for informed decision making in deploying object detection in NIR and EO imaging.

The fusion of electro-optical and near-infrared (NIR) sensor data presents both challenges and significant advantages. Electro-optical sensors are good at capturing high-resolution visible-spectrum images, which are effective in well-lit conditions but can struggle with occlusions and varying light conditions. In contrast, NIR sensors excel in low-light environments and can penetrate through certain types of occlusions, providing valuable information where visible-spectrum data might be obscured by smoke, fog, or other obstructions. Combining these modalities leverages their complementary strengths. Electro-optical sensors offer detailed texture and color information, while NIR sensors provide enhanced visibility in challenging lighting conditions.

Although the proposed methodology has achieved successful results, it is essential to mention that Singapore Maritime Dataset (SMD) does not completely represent the real-world environment or the dynamic nature of the sea. In this regard, the proposed model may perform differently when deployed in real-time applications. This requires further fine-tuning and experimentation of the model in real-world scenarios. Similarly, it was also observed that some objects within the dataset were not labeled, such as human beings and birds. Before deployment, it is essential to incorporate these objects in the model for better performance and to avoid any accidents at sea. Moreover, our dataset lacks videos in adverse weather conditions such as storms, cyclones, and hail. For real-world applications, it is essential that models must be trained on both calm weather and adverse conditions. Finally, ensemble techniques require more computational resources than a single model because multiple models are initially trained and then combined.

Future research could place emphasis on numerous crucial areas to advance the field of sensor fusion and object detection in maritime environments. Firstly, improving sensor fusion techniques could involve the development of more sophisticated algorithms that better integrate electro-optical and NIR data. Advances in machine learning and data fusion methods may enhance the accuracy and robustness of these systems, particularly in challenging conditions.

## 5. Conclusions

In this research, an ensemble of advanced models were successfully trained on the challenging Singapore Maritime Dataset (SMD). The proposed ensemble model achieved an increase in mean average precision and also maintains a balance between computational speed and precision. The development resulted in increased sensitivity in maritime object detection as compared to the base learners. This is possible because of the unique employment of a logical-disjunction-based final ensemble classification mechanism using the dedicated sub-ensemble classifiers for each NIR and electro-optical sensor object detection and classification. Because of this logical-disjunction-based decision combination, the proposed model resulted in discouraging false negatives with an appropriate and acceptable tradeoff with false positives and thus, the proposed model is basically designed to ensure that if an object is present, the chances of missing its presence are minimized.

In conclusion, while the maritime object detection system presented in this research shows promise, ongoing work aims to enhance its adaptability, accuracy, and applicability

in dynamic maritime environments, contributing to the continual evolution of maritime surveillance and navigation technologies.

Further along this line of research, there are several potential areas that can be explored for optimizing ensemble configuration and model weights for specific maritime applications depending on the operational requirements. Another area that can be explored is the integration of physical phenomena and environmental factors that can possibly enhance the adaptability. Ongoing research can focus on continually upgrading the dataset, addressing evolving maritime challenges. In particular, combining visible light and infrared images will expand the adaptability of the system in diverse weather conditions.

**Author Contributions:** Conceptualization, M.F.J., M.O.I., M.A. and I.M.; Methodology, M.F.J., M.O.I., M.A. and J.-Y.K.; Software, M.F.J. and M.O.I.; Validation, M.F.J., M.O.I. and I.M.; Formal analysis, M.F.J., M.O.I. and I.M.; Investigation, M.F.J. and I.M.; Resources, J.-Y.K.; Data curation, M.F.J.; Writing—original draft, M.F.J. and M.O.I.; Writing—review & editing, M.A., I.M. and J.-Y.K.; Supervision, M.A. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** The authors declared that the datasets used in this research are publicly available.

**Conflicts of Interest:** The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

## References

1. Meersman, H. Maritime traffic and the world economy. In *Future Challenges for the Port and Shipping Sector*; Informa Law from Routledge: Boca Raton, FL, USA, 2014; pp. 1–25.
2. Dalaklis, D. Safety and security in shipping operations. In *Shipping Operations Management*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 197–213.
3. United Nations. Review of Maritime Transport. In Proceedings of the United Nations Conference on Trade and Development. 2022. Available online: <https://www.un-ilibrary.org/content/books/9789210021470> (accessed on 3 September 2024).
4. Prasad, D.K.; Rajan, D.; Rachmawati, L.; Rajabally, E.; Quek, C. Video processing from electro-optical sensors for object detection and tracking in a maritime environment: A survey. *IEEE Trans. Intell. Transp. Syst.* **2017**, *18*, 1993–2016. [CrossRef]
5. Wang, N.; Wang, Y.; Er, M.J. Review on deep learning techniques for marine object recognition: Architectures and algorithms. *Control Eng. Pract.* **2022**, *118*, 104458. [CrossRef]
6. Rekavandi, A.M.; Xu, L.; Boussaid, F.; Seghouane, A.-K.; Hoefs, S.; Bennamoun, M. A guide to image and video based small object detection using deep learning: Case study of maritime surveillance. *arXiv* **2022**, arXiv:2207.12926.
7. Zhao, C.; Liu, R.W.; Qu, J.; Gao, R. Deep learning-based object detection in maritime unmanned aerial vehicle imagery: Review and experimental comparisons. *Eng. Appl. Artif. Intell.* **2024**, *128*, 107513. [CrossRef]
8. Negahdaripour, S.; Xu, X.; Khamene, A. A vision system for real-time positioning, navigation, and video mosaicing of sea floor imagery in the application of ROVs/AUVs. In Proceedings of the Fourth IEEE Workshop on Applications of Computer Vision, WACV'98 (Cat. No. 98EX201), Princeton, NJ, USA, 19–21 October 1998; pp. 248–249.
9. Cozman, F.; Krotkov, E. Robot localization using a computer vision sextant. In Proceedings of the 1995 IEEE International Conference on Robotics and Automation, Aichi, Japan, 21–27 May 1995; Volume 1, pp. 106–111.
10. González-Sabbagh, S.P.; Robles-Kelly, A. A survey on underwater computer vision. *ACM Comput. Surv.* **2023**, *55*, 1–39. [CrossRef]
11. Prasad, D.K.; Prasath, C.K.; Rajan, D.; Rachmawati, L.; Rajabaly, E.; Quek, C. Challenges in video based object detection in maritime scenario using computer vision. *arXiv* **2016**, arXiv:1608.01079.
12. Moosbauer, S.; Konig, D.; Jakel, J.; Teutsch, M. A benchmark for deep learning based object detection in maritime environments. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Long Beach, CA, USA, 16–17 June 2019.
13. Kim, J.-H.; Kim, N.; Park, Y.W.; Won, C.S. Object detection and classification based on YOLO-V5 with improved maritime dataset. *J. Mar. Sci. Eng.* **2022**, *10*, 377. [CrossRef]
14. Zhao, H.; Zhang, H.; Zhao, Y. Yolov7-sea: Object detection of maritime uav images based on improved yolov7. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 2–7 January 2023; pp. 233–238.
15. Yu, Y.; Zhao, J.; Gong, Q.; Huang, C.; Zheng, G.; Ma, J. Real-time underwater maritime object detection in side-scan sonar images based on transformer-YOLOv5. *Remote Sens.* **2021**, *13*, 3555. [CrossRef]
16. Shin, H.-C.; Lee, K.-I.; Lee, C.-E. Data augmentation method of object detection for deep learning in maritime image. In Proceedings of the 2020 IEEE International Conference on Big Data and Smart Computing (BigComp), Busan, Republic of Korea, 19–22 February 2020; pp. 463–466.

17. Haghbayan, M.-H.; Farahnakian, F.; Poikonen, J.; Laurinen, M.; Nevalainen, P.; Plosila, J.; Heikkonen, J. An efficient multi-sensor fusion approach for object detection in maritime environments. In Proceedings of the 2018 21st International Conference on Intelligent Transportation Systems (ITSC), Maui, HI, USA, 4–7 November 2018; pp. 2163–2170.
18. Cane, T.; Ferryman, J. Evaluating deep semantic segmentation networks for object detection in maritime surveillance. In Proceedings of the 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Auckland, New Zealand, 27–30 November 2018; pp. 1–6.
19. Vasilopoulos, E.; Vosinakis, G.; Krommyda, M.; Karagiannidis, L.; Ouzounoglou, E.; Amditis, A. A comparative study of autonomous object detection algorithms in the maritime environment using a UAV platform. *Computation* **2022**, *10*, 42. [[CrossRef](#)]
20. Iancu, B.; Winsten, J.; Soloviev, V.; Lilius, J. A Benchmark for Maritime Object Detection with Centernet on an Improved Dataset, ABOships-PLUS. *J. Mar. Sci. Eng.* **2023**, *11*, 1638. [[CrossRef](#)]
21. Rahman, A.; Lu, Y.; Wang, H. Performance evaluation of deep learning object detectors for weed detection for cotton. *Smart Agric. Technol.* **2023**, *3*, 100126. [[CrossRef](#)]
22. Khosravi, B.; Mickley, J.P.; Rouzrokh, P.; Taunton, M.J.; Larson, A.N.; Erickson, B.J.; Wyles, C.C. Anonymizing radiographs using an object detection deep learning algorithm. *Radiol. Artif. Intell.* **2023**, *5*, e230085. [[CrossRef](#)] [[PubMed](#)]
23. Ma, P.; Li, C.; Rahaman, M.M.; Yao, Y.; Zhang, J.; Zou, S.; Zhao, X.; Grzegorzec, M. A state-of-the-art survey of object detection techniques in microorganism image analysis: From classical methods to deep learning approaches. *Artif. Intell. Rev.* **2023**, *56*, 1627–1698. [[CrossRef](#)] [[PubMed](#)]
24. Li, W.; Zhang, L.; Wu, C.; Cui, Z.; Niu, C. A new lightweight deep neural network for surface scratch detection. *Int. J. Adv. Manuf. Technol.* **2022**, *123*, 1999–2015. [[CrossRef](#)] [[PubMed](#)]
25. Howard, A.G. MobileNets: Efficient convolutional neural networks for mobile vision applications. *arXiv* **2017**, arXiv:1704.04861.
26. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *39*, 1137–1149. [[CrossRef](#)] [[PubMed](#)]
27. Hutter, F.; Kotthoff, L.; Vanschoren, J. *Automated Machine Learning: Methods, Systems, Challenges*; Springer Nature: Berlin/Heidelberg, Germany, 2019.
28. Shuja, J.; Bilal, K.; Alasmary, W.; Sinky, H.; Alanazi, E. Applying machine learning techniques for caching in next-generation edge networks: A comprehensive survey. *J. Netw. Comput. Appl.* **2021**, *181*, 103005. [[CrossRef](#)]
29. Ganaie, M.A.; Hu, M.; Malik, A.K.; Tanveer, M.; Suganthan, P.N. Ensemble deep learning: A review. *Eng. Appl. Artif. Intell.* **2022**, *115*, 105151. [[CrossRef](#)]
30. Polikar, R. Ensemble learning. In *Ensemble Machine Learning: Methods and Applications*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 1–34.
31. Murtza, I.; Kim, J.-Y.; Adnan, M. Predicting the Performance of Ensemble Classification Using Conditional Joint Probability. *Mathematics* **2024**, *12*, 2586. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.