MDPI

Review

Literature Review on Ship Localization, Classification, and Detection Methods Based on Optical Sensors and Neural Networks

Eduardo Teixeira [ID], Beatriz Araujo [ID], Victor Costa, Samuel Mafra and Felipe Figueiredo *[ID]

Instituto Nacional de Telecomunicações (INATEL), Santa Rita do Sapucaí 37540-000, MG, Brazil
* Correspondence: felipe.figueiredo@inatel.br (F.F.)

Abstract: Object detection is a common application within the computer vision area. Its tasks include the classic challenges of object localization and classification. As a consequence, object detection is a challenging task. Furthermore, this technique is crucial for maritime applications since situational awareness can bring various benefits to surveillance systems. The literature presents various models to improve automatic target recognition and tracking capabilities that can be applied to and leverage maritime surveillance systems. Therefore, this paper reviews the available models focused on localization, classification, and detection. Moreover, it analyzes several works that apply the discussed models to the maritime surveillance scenario. Finally, it highlights the main opportunities and challenges, encouraging new research in this area.

Keywords: maritime surveillance; classification; localization; detection; artificial intelligence; neural networks

1. Introduction

With the growth in ocean exploration by cruise ships, ocean liners, and other marine ships, the need for monitoring systems has increased considerably. With this, monitoring stations have become increasingly equipped to carry out the identification of possible issues. Among the maritime monitoring applications, one can mention potential collision prediction [1], navigation support, tracking of ships drift [2] target tracking, maritime safety [3]. Visual ship tracking provides crucial kinematic traffic information to maritime traffic participants, which helps to accurately predict ship traveling behaviors in the near future. Each of these applications requires different operating architectures [4].

Automatic maritime surveillance assumes the use of sensors that can provide enough information for automatic situational awareness tasks, such as localization and classification. In localization, a single object is found in an image. In classification, the object is defined as belonging to a specific class. Detection combines the characteristics of these two techniques to locate and classify multiple targets in the scene.

The fusion of different sensing sources can provide a better situational view of the monitored environment and help one take the necessary actions. Sensors based on sound or electromagnetic waves, such as sonar and radar, are generally employed in long-range applications. Optical sensors can be more economical alternatives for applications that require greater detail of the ships and aim for low power consumption. They can be employed in ship tracking and classification tasks, requiring only that these sensors be combined with visual detection techniques that are efficient, fast, and robust to enable the advancement of maritime applications [5].

The number of sensors involved can also vary, and the most common ones for this purpose are thermal cameras, optical cameras, and radar. Choosing the best techniques to obtain maritime situational information is not trivial since the literature offers a huge amount of model options that employ optical data. To make their application even more

complicated, weather or water conditions, such as wind speed, tidal changes, rain, and fog, can, for example, blur or entirely obstruct objects in an image. Additionally, the increasing distance between the monitored object and the sensors can also aggravate visual tasks, as it can cause a high-scale variation.

At the beginning of the research on ship detection, such as the one on object detection in general, methods employing simple and handcrafted features were used, but recently, convolutional neural networks (CNNs) have been added as part of this field of study because of their extraordinary ability to extract and represent visual features [6]. For example, in automatic navigation systems, the role of CNNs is to interpret the visual data collected by the cameras. Thus, the detection information is added to the data from different sensors, allowing the data fusion processing system to have enough information for decision-making.

The paper is organized as follows. Section 2 describes the theoretical background of the techniques, as well as presenting related works. Section 3 explores the different datasets found in the literature, detailing ship classes and image sample characteristics. Section 4 shows the challenges and open questions. Finally, Section 5 concludes this work with some concluding remarks and a discussion of future work, and outlines some lines of future research.

2. Related Works

In marine monitoring scenarios, localization, classification, and detection techniques are applied to data received from sensors to extract information on the location and type of the monitored ship. For example, in the case of optical sensors, commonly called cameras, an analysis of the images is performed within the viewing angle. In this case, images are received frame by frame and processed with the desired algorithm, be it for localization, classification, or detection.

Several works are proposed by autonomous authors or those affiliated with some universities or research centers to perform the tasks in the localization, classification, and detection of ships [7]. Figure 1 shows a generic block diagram for optical sensor image processing systems. This system has four main stages: image acquisition, preprocessing, processing, and information output.
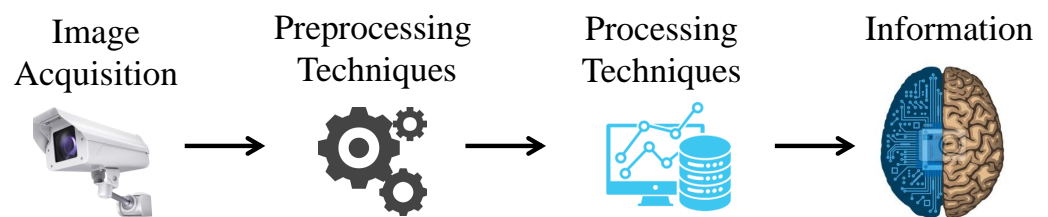
Figure 1. Components of an image processing system.

The first stage is responsible for frame capture, followed by signal preprocessing and the application of corresponding techniques for localization, classification, or detection. In the first stage, individual or joint optical sensors convert the light reflected by the objects into arrays of pixels representing the scene. The next stage, which may or may not exist depending on the system, is preprocessing. Here, the image is prepared for feature extraction using techniques for noise reduction, deblurring, feature intensification, and image quality enhancement [8].

The image enhancement process is usually applied so that the processing models perform better due to higher quality samples. With that, as long as the preprocessing is well implemented, the next step, such as extracting features or object segmentation, performs better [9]. Some jobs do not have a preprocessing stage, while others divide some of these stages into more than one task. The inclusion of the preprocessing step directly influences the parameters related to the performance of the models.

Among the factors that are influenced, it is possible to mention the total time of the image processing system and factors related to accuracy metrics. To calculate the factors related to the accuracy metrics, it is necessary to initially obtain four response parameters from these models: "true positive", "true negative", "false positive", and "false negative".

Once these parameters are obtained, it is possible to calculate the accuracy, which is the total number of correct predictions divided by the total number of predictions made for a dataset, and precision, which quantifies the number of positive class predictions that belong to the positive class. Some works also feature recall, which quantifies the number of positive class predictions made out of all positive examples in the dataset, and the F1-score, which provides a single score that balances the concerns of precision and recall in one number.

2.1. Image Acquisition

During image capture, several types of sensors can be used. However, the focus of this work is on the use of optical sensors, be they remote, installed on satellites or aircraft, or those that observe from a side view, such as those installed in inshore or offshore scenarios, such as on other ships or fixed constructions on land, usually near the coast.

Optical images can still be divided into visible and infrared (IR) spectra, and the range of both is very similar, from the order of meters to at most a few kilometers. The main differences between optical sensors are related to sensitivity to the environment and the quality and quantity of visual information generated by the sensor [7].

Comparing the sensitivity to illumination, both sensor types have problems working outside their respective designations, i.e., while the visible light sensor performs poorly for nighttime applications, the IR sensor presents high saturation in images captured during the day. In addition, the visible light sensor is less robust to the effects of light reflection on ships caused by water dynamics. However, the visual data generated are more detailed when compared to the quality and quantity of elements captured by a visible sensor [7]. Thus, this sensor can lead to the training of detectors with higher reliability.

Optical remote sensing images suffer from weather conditions, such as rain, waves, fog, and clouds, which causes the need, in some cases, for preprocessing of the image to improve the image quality that will be analyzed.

2.2. Preprocessing Techniques

The preprocessing step can be used to improve the quality of images by introducing techniques that allow obtaining a dataset with better quality images through the attenuation of interference caused by elements of the environment, such as extreme brightness and contrast, in addition to the quality of the camera lenses used in the capture process [10].

Among the various techniques that can be used in preprocessing, it is possible to mention super-resolution techniques [11,12] and deblur [13,14]. The main benefit of improving images before they are used in location, classification, or detection models is usually the improvement of accuracy achieved only by increasing the quality of the dataset [15]. An example of detection enhancement can be seen in Figure 2.

Super-resolution techniques are used to recover quality and improve the resolution of an image. With this, instead of receiving low-resolution images, the model starts to operate with more detailed images, leading various situations to an instant performance improvement [16]. The field of image super-resolution has been dominated by methods based on convolutional neural networks in recent years [17]. Among the examples of models related to the super-resolution task, one can mention the models based on training to minimize the mean squared error (MSE), such as super-resolution convolutional neural network (SRCNN) [18], super-resolution residual network (SRResNet) [19], enhanced deep super-resolution network (EDSR) [20], multi-scale deep super-resolution (MDSR) [20], and deep back-projection networks (DBPN) [21], and also models based on generative adversarial networks (GANs), such as super-resolution generative adversarial network

(SRGAN) [22], enhanced super-resolution generative adversarial network (ESRGAN) [23], and rank super-resolution generative adversarial network (RankSRGAN) [24].
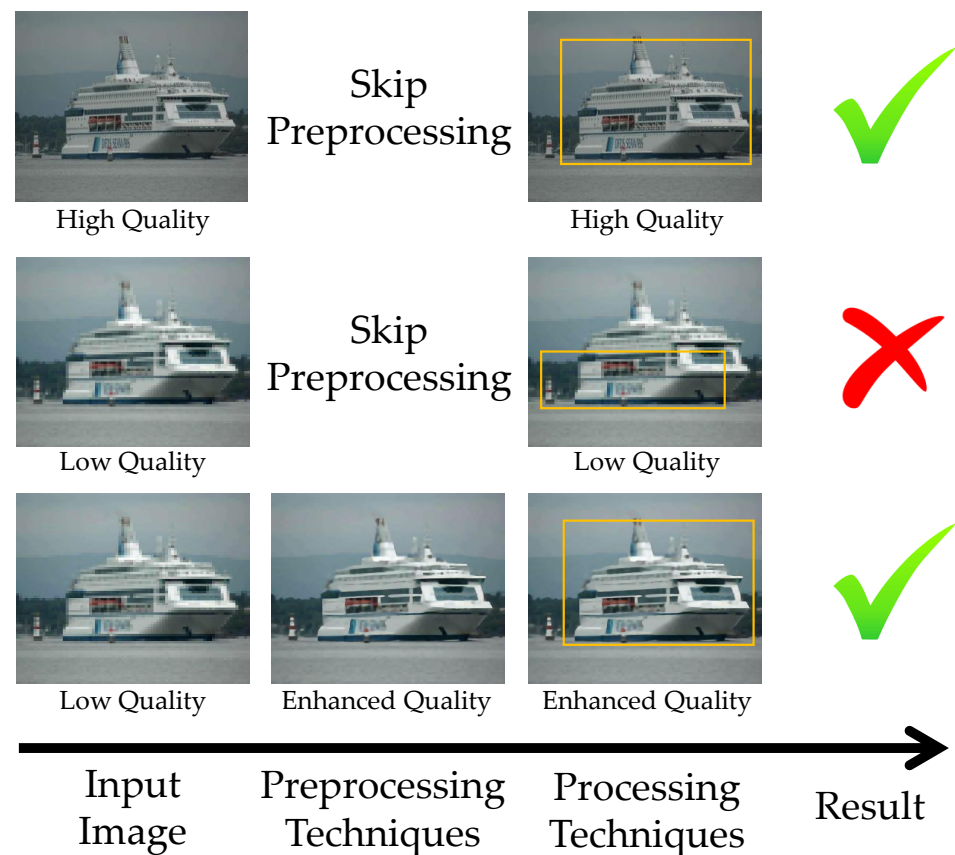
Figure 2. Detection enhancement with preprocessing.

In addition to being based on different forms of training, these models also differ in the layer structures that make up their architectures, which influence their performance, both in aspects related to accuracy and their processing time. Super-resolution models trained through an MSE estimator use the distance between training images and associated predictions as a cost function. As a result, these models tend to produce smoother images. On the other, models based on GANs use generators to create new false images that can mimic the expected result so that the discriminative model has the maximum difficulty distinguishing between the synthesized images of the generator and the actual images. This process generates output images with more realistic detail but can cause some unwanted noise to enter the image during the super-resolution process [25].

2.2.1. SRCNN

The SRCNN model is based on an CNN architecture, trained to learn an end-to-end mapping between low and high-resolution images for the super-resolution (SR) problem. It was one of the first architectures that applied the concept of deep learning in the super-resolution task, achieving one of the best results for this task in 2015. When the authors proposed the use of CNNs, the most common was to use the traditional sparse-coding-based SR methods [18]. The model is divided into three stages; the first is responsible for performing the extraction and representation of low-quality images within the network. The second step is the application of a nonlinear mapping, where the layers of CNN extract as much information about the image. The last step is to rebuild the image at a higher resolution than the image applied to the template entry [18].

2.2.2. SRResNet

The model SRResNet was developed with an architecture based on residual network (ResNet), but with modifications in the optimization of the MSE loss function to achieve high upscaling factors, $4\times$, as quoted in [22]. With this optimization, the model was consolidated in 2017 as the new state-of-the-art, and its performance was evaluated by peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM), two metrics widely used in image quality assessment [26].

Another proposal by the authors of [22] was to change the MSE loss function using model resources visual geometry group (VGG). With this, the authors compared the optimized version of the MSE with the version modified to use the VGG loss metrics. The result was that the model improved the visual metric mean opinion score (MOS) but had lower performance on the metrics of PSNR and SSIM [22].

2.2.3. SRGAN

The model SRGAN is trained through a generative architecture based on the ResNet. This architecture generates images with super-resolution and has its loss analyzed through a second structure with a discriminative function, which only acts during training. The commission's proposal SRGAN is to use a new resource-based loss function of the model VGG, which, when combined with the discriminating network, helps to detect the difference between the image generated about the reference image. According to metric-based tests MOS [27], which evaluates the perceptual quality of an empirically obtained image through a visual classification scale, the trained model achieves results close to the state of the art in the literature [22].

2.2.4. EDSR

The model EDSR, as well as the other SR models already presented, also has its architecture based on ResNet. These models have characteristics similar to SRResNet. However, unnecessary modules are removed from the architecture to optimize the model. Among these changes, one can mention the residual blocks, which have removed the batch normalization (BN). This causes the model to be simplified and memory usage to be reduced [20]. Just as SRResNet achieves $4\times$ upscaling factors, the EDSR model is also capable. In the work [20], the authors trained the model for upscaling of $2\times$, $3\times$, and $4\times$. In addition to achieving model time optimization and simplifying the architecture, the authors also had superior results compared to other networks that were tested, such as SRCNN, SRResNet, and MDSR.

2.2.5. MDSR

The model MDSR was proposed by the same author who developed the EDSR model such that their architectures are described in [20]. The MDSR network has a certain increase in complexity compared to EDSR because it uses extra blocks with different scales at the beginning of the architecture. Removing the BN layers, as suggested in [20], is also adopted in this model [20]. Unlike EDSR, which reconstructs only a super-resolution image scale on the MDSR network, an initial upscaling is applied that operates with parallel image processing structures of different sizes. This allows for reducing various problems caused by variations in image scale. Both models, EDSR and MDSR, were proposed in the NTIRE 2017 Super-Resolution Challenge [28], taking first and second place, respectively. With this, the authors claimed that they managed simultaneously to achieve the state of the art in the topic of super-resolution and simultaneously transformed the architecture ResNet into a more compact model.

2.2.6. ESRGAN

In three respects, the model ESRGAN is an improved version of the network SRGAN. The first was the replacement of residual blocks by residual-in-residual dense block (RRDB) to facilitate training, followed by the exchange of layers of BN by residual scaling and

smaller initialization, as suggested in [20] because it allows the training of a more profound architecture. The second difference was the replacement of a GAN joint for a relativistic average GAN (RaGAN); instead of judging whether an image is true or false, this generative network can identify which image is more realistic. Finally, the perceptual loss was improved using VGG features before and after activation as in the SRGAN. This last change makes the model provide sharper edges and visually more satisfying results [23]. With the modifications made, the model reached the state of the art in 2018, presenting the best results of perceptual quality, with first place in the challenge perceptual image restoration and manipulation—super resolution (PIRM-SR) [17]. The test evaluated several models under the quality of visual perception metrics, PSNR and SSIM. From the evaluation of the performance of the models in this challenge, it was possible to notice that the increasing values of PSNR and SSIM were not always accompanied by an increase in perceptual quality. In many cases, this resulted in increasingly blurred and unnatural outputs, which gives more meaning to the previously cited results of [22].

2.2.7. RankSRGAN

The RankSRGAN model is based on the GANs architecture but adopts a siamese architecture to learn perceptual metrics and rank images according to the quality score found during its training. This model combines different SR algorithms to improve perceptual metrics by combining other models [24]. To train the ranker, the authors used three templates, SRResNet, SRGAN, and ESRGAN. With their combination, RankSRGAN was able to optimize the natural image quality evaluator (NIQE) parameter [29], a visual metric that measures the naturalness of the image in the scene. With this, the model achieved superior performance to the individual models used when applied to the dataset of the PIRM-SR Challenge 2018 [24].

2.2.8. DBPN

The DBPN model is an improved version of the SRCNN network, but instead of using predefined upsampling, it uses interleaved upsampling and downsampling layers. Unlike other methods that build the SR image feed-forwardly, our proposed networks focus on directly increasing SR resources by using multiple stages of ascending and descending sampling that feeds error predictions into each depth. The values of the error feedback of the steps of increase and scale reduction were used to guide the network to obtain a better result. The model performed similarly to the state-of-the-art performance in 2018. In addition, the network was trained with $8\times$ magnification, higher than that used in the creation of SRResNet [19] and EDSR [20].

Unlike super-resolution techniques, deblurring techniques were developed to remove noise and blur present in the image, which hinders the visualization of the image. When noisy images are treated before being inserted into detection and classification systems, the system performance can increase considerably [30]. Some of the techniques that can be applied to the deblur task are deblur generative adversarial network (DeblurGAN), DeblurGAN-V2, and deblurring and shape recovery of fast moving objects (DeFMO).

2.2.9. DeblurGAN

The DeblurGAN template is composed of a GAN architecture, and its purpose is to remove blur in images. The model features an architecture of CNN, composed of residual blocks (ResBlocks) consisting of a convolution layer, instance normalization layer, and ReLU activation [31]. The authors of DeblurGAN validated their results by applying the you only look once (YOLO) model to perform the detection and classification of objects in images with blur and images processed by the deblur model. There is a gain in accuracy in the YOLO results when inserted images are improved by the DeblurGAN model, proving that it significantly contributes to image quality and consequently to the performance of subsequent processing systems [31].

2.2.10. DeblurGAN-V2

The DeblurGAN-V2 model is based on the construction of the original model Deblur-GAN but with some modifications to improve the [32] network. Among them, the generative model in DeblurGAN-V2 integrates the technique feature pyramid network (FPN). This technique was initially developed for object detection purposes [33]. Still, in the case of the DeblurGAN-V2 model, the authors used FPN for the construction of a noisy image [32]. In addition to integrating the FPN technique, the new version allows the selection of different backbones. Each of the different backbones is designed to improve some of the performance parameters. For example, with the Inception-ResNetV2 architecture, you obtain a next-generation blur. In contrast, with the mobile network-depthwise separable convolution (MobileNet-DSC) architecture, you obtain an increase in processing speed, some 10 to 100 times faster than the top competitors in 2019 [32].

2.2.11. DeFMO

Motion blur is one of the existing blur types, and it is caused by the rapid movement of objects when captured by cameras or by the quick scroll of the camera to capture still objects, recording photos or videos, with blur [34]. Thus, DeFMO is designed to act on this type of blur. The proposed network is a novel based on a 'self-supervised' loss function that improves the model's accuracy when applied to images with motion blur. By presenting a good generalization capability, this model can be applied to different areas in computer vision, such as the improvement of security cameras, microscopes, and photos with high noise levels [35]. This model is the first fully neural FMOs deblurring that fills the gap between deblurring, 3D modeling, and FMO subframe tracking for trajectory analysis.

2.3. Processing Techniques

Most previously proposed models for image processing, that is, location, classification, or detection of ships, have focused on using handcrafted resources applied to image processing. These models are built with the expert knowledge of designers. Within the scope of handcraft features models, it is possible to point out several works that employ different techniques, such as Gabor filter in [36], for automatic target detection, discrete cosine transform (DCT) in [37] for maritime surveillance on non-stationary surface platforms, as well as Haar–Cascade [38], scale-invariant feature transform (SIFT) [39], local binary pattern (LBP) [40], support vector machine (SVM) [41], and histograms of oriented gradients (HOG) [42] for the remote sensing of ships.

As a result, the extracted features reflect the limited aspects of the problem, generating a low response accuracy of the models and a low generalization. Thus, deep learning in the computer vision research community, such as CNNs proved to be more suitable for developing and training resource extractors [43].

The techniques based on CNN dominate the most recent works, as shown in Table 1, which details the evolution of the works over the years, pointing out aspects such as the type of image used, applications, and techniques involved in each of the works. They won great strength after winning the ImageNet challenge in 2012 and have been achieving excellent results in several image processing tasks for obtaining visual information [44].

Another point that collaborates with this type of network is the evolution of the sizes of the available datasets, given that CNNs usually require a large number of training samples. With this, the use of detection models based on CNNs has accelerated even more because, according to [45], a good object detector should improve when given more training data.

Within these networks, there is a subclass, the region-based convolutional neural networks (R-CNNs), whose working principle is based on a selective search for object detection, generating region proposals, as shown in Figure 3. Work related to this type of technique began with the R-CNN, proposed by Ross Girshick [46]. Since then, other variations have been proposed, such as fast R-CNN [45], faster R-CNN [47], mask R-CNN [48], single-shot detector (SSD) [49], YOLO [50], YOLOv2/9000 [51], YOLOv3 [52], YOLOv4 [52], and YOLOv5 [52]. These models have some modifications in their topologies to increase

their speed and prediction performances or even to add a new function, as is the case of segmentation in mask R-CNN.
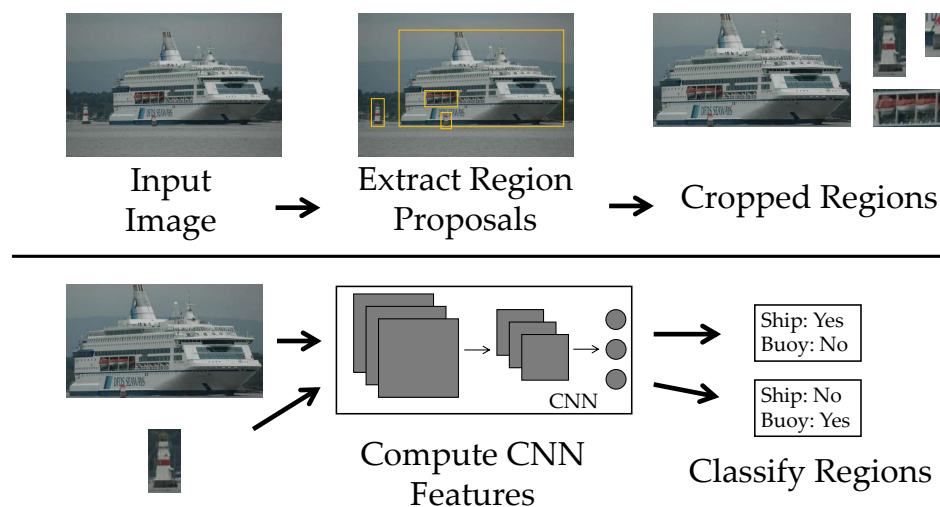
Figure 3. Regions with CNN features.

2.3.1. R-CNN

It emerged with the task of localizing objects through a CNN that could have high detection capability even with a small amount of annotated samples for its training. It is basically divided into three modules. The first is responsible for generating several region proposals without a specific category, by a method called selective search (SS) [53]. The second is an CNN, which extracts a fixed number of features for each of the proposals. Finally, the third module is based on a linear SVMs trained specifically for each possible class. With this, this network can not only locate the object, but also inform which of the possible classes it belongs to. This classification is performed through a score generated by the [46] classifiers.

2.3.2. Fast R-CNN

Fast R-CNN introduces single-stage training with an update of all layers and avoids disk storage for feature caching [45]. Regarding the detection task, it has the advantage of achieving higher mean average precision (mAP) compared to its standard version. In this model, the linear SVMs used in R-CNN is replaced by a softmax classifier. Using the same training algorithm and hyperparameters used in R-CNN, they train a new SVM to be the classifier for fast R-CNN and justify the use of softmax by achieving a slight advantage in mAP over it [45].

2.3.3. Faster R-CNN

This model uses the region proposal network (RPN), which comprises CNNs capable of providing region proposals to fast R-CNN, informing at the same time the object boundaries and the scores of each proposed region. RPN calculates proposal regions much faster and more efficiently compared to SS. Moreover, it brings another advantage by sharing convolutional layers between the proposal generation network and the classification network, optimizing the network training [47].

2.3.4. Mask R-CNN

It follows the same principle as faster R-CNN but has a second output in the model for segmenting objects [48]. The pixel-by-pixel object segmentation is performed through the superposition of an outline, applied by this second output. This overlay mask is applied to each region of interest (RoI) and is based on the fully connected neural network (FCNN) model [54].

Table 1. Models and features on related works.

	Image View		Approaches		
Papers	**Side View**	**Remote**	**Localization**	**Classification**	**Techniques/Models**
2017 [55]	-	x	x	-	FusionNet
2017 [56]	x	-	-	x	VGG16
2018 [57]	x	-	x	-	Faster R-CNN+ResNet
2018 [58]	-	x	x	-	ResNet-50
2018 [59]	-	x	x	-	SNN
2018 [60]	-	x	x	x	Faster R-CNN+Inception-ResNet
2018 [61]	-	x	x	-	RetinaNet
2018 [62]	-	x	x	x	R-CNN
2018 [63]	-	x	x	-	R-CNN
2019 [64]	-	x	-	x	VGG19
2019 [65]	-	x	-	x	VGG16
2019 [66]	x	-	-	x	Skip-ENet
2019 [67]	-	x	x	x	Cascade R-CNN+B2RB
2019 [68]	-	x	-	x	ResNet-34
2019 [69]	x	-	x	-	YOLOv3
2019 [70]	-	x	x	x	VGG16
2019 [71]	x	-	x	-	Faster R-CNN
2020 [72]	-	x	x	x	SSS-Net
2020 [73]	-	x	x	x	YOLOv3
2020 [74]	-	x	x	x	CNN
2020 [75]	x	-	x	-	CNN Segmentation
2020 [76]	-	x	x	-	YOLO
2020 [77]	-	x	x	x	ResNet-50+RNP
2020 [78]	x	-	-	x	CNN
2020 [79]	x	-	x	x	YOLOv4
2020 [80]	-	x	x	-	YOLOv3
2020 [81]	-	x	-	x	VGG16
2020 [82]	x	-	x	-	Mask R-CNN+YOLOv1
2021 [83]	-	x	x	x	Mask RPN+DenseNet
2021 [84]	-	x	x	-	VGG16
2021 [85]	x	-	x	x	SSD MobileNetV2
2021 [86]	x	-	x	x	YOLOv3
2021 [87]	x	-	x	-	Faster R-CNN
2021 [88]	x	-	x	-	R-CNN
2021 [89]	x	-	x	x	BLS

Table 1. *Cont.*

Papers	Image View		Approaches		Techniques/Models
	Side View	Remote	Localization	Classification	
2021 [90]	x	-	x	-	YOLOv5
2021 [3]	x	-	x	x	MobileNet+YOLOv4
2021 [91]	-	x	x	x	Cascade R-CNN
2021 [92]	x	-	x	x	YOLOv3
2021 [93]	-	x	x	x	YOLOv3
2021 [94]	x	-	x	x	YOLOv3
2021 [95]	-	x	x	x	YOLOv4
2021 [96]	x	-	x	x	ResNet-152
2021 [97]	-	x	x	-	Faster R-CNN
2022 [92]	x	-	x	x	YOLOv4
2022 [98]	-	x	x	-	YOLOv3
2022 [99]	x	-	x	x	MobileNetV2+YOLOv4
2022 [100]	-	x	x	x	YOLOv5

2.3.5. SSD

Compared to previous methods that take two stages, SSD is a more straightforward method because it encapsulates all computations in a single deep neural net, eliminating the need to generate object proposals in multiple stages. This increases the speed of the system and facilitates training by providing a unified structure for training and inference. It scores bounding boxes and adjusts to best match the shape of the object and uses boxes of different proportions to handle objects of different sizes [49].

2.3.6. YOLO

Like SSD, this is also a single-stage detector, which can have its optimized performance within its unified detection model. In this method, object detection is performed as a regression task for bounding boxes, which, at the same time, provides the object locations with their respective classes. The primary source of error in this network is in the incorrect location of small objects [50].

3. Datasets

Datasets are structured collections of data that are used by computer vision models during their training and validation stages. Different datasets have been created throughout the literature for visual tasks.

Image databases, in general, whether for ship classification or other purposes, usually have their images divided into classes. The number of images, the number of classes, and the complexity of visual separation of these objects directly affect the training and the results of computer vision systems.

Some datasets have a considerable imbalance concerning the number of images in each class, or even very similar classes. In datasets with many classes, many training iterations may be necessary to achieve good accuracy and other parameters related to system accuracy, such as precision, recall, and F1-score. Even in datasets with few classes, if the similarity between the objects of the two classes is high, it can also require a large number of iterations [101].

In the case of the maritime scenario, for example, architectures generally make many more mistakes when relating classes to ships than when differentiating a ship from a buoy

or even some piece of wood, metal, or rubber lost at sea. This occurs because the ability to visually separate different categories of objects depends on the similarity between them in the image classification process. Therefore, some categories are more difficult to distinguish than others [102].

To suppress this problem, some models also create classes of mountains, trees, buildings, sky, and pier objects, to minimize false ship detection. However, to avoid high pollution of output result elements, some of these networks do not visually deliver the markup of these classes [7].

As long as the images are well curated, the detection and classification performance tends to increase with the expansion of the training data, as long as it is of good quality. For this reason, increasingly more extensive databases are being built, such as the the MARVEL dataset [103], which has more than 2 million images.

Considering general-purpose datasets, it is possible to cite MS COCO [104], CIFAR-10 [105], PASCAL VOC [106], OpenImage [107], and ImageNet [108] as some of the most used datasets containing ship images [109]. These datasets contain thousands and even millions of images divided into different classes, which serve as the basis for the training and validation of object detection and classification models. Each of these datasets contains a class of generic ships. According to Table 2, it is possible to obtain 11,570 images with the sum of the samples of these datasets.

Table 2. Generic datasets.

Dataset	Ship Count
COCO [104]	3146
CIFAR-10 [105]	6000
PASCAL VOC [106]	353
OpenImage [107]	1000
ImageNet [108]	1071

However, specialized ship datasets have many more images of ships and subdivide these ships into sub-classes, giving more detail to the identified object. Table 3 lists some of these specialized datasets, providing the number of classes, the number of images of the ships, and the spatial view of these ships, where the photo datasets are divided into two groups: photos taken from the sides of the ships, in any angle within the 360° of the ship, or photos taken from the top of the ship, usually captured by satellites and classified as remote.

Table 3. Remote and side view ship datasets.

| Dataset	Side View	Remote	Images	Ship Classes
VAIS [110]	x	-	2865	15
ABOShips [109]	x	-	9880	9
MCShips [111]	x	-	14,709	13
Singapore [7]	x	-	17,450	6
SeaShips [112]	x	-	31,455	6
MARVEL [103]	x	-	2,000,000	29
HRSC2016 [113]	-	x	1061	19
Airbus Ship Detection [114]	-	x	208,162	1
BCCT200 [115]	-	x	800	4
ShipRSImageNet [116]	-	x	3435	50

3.1. Dataset Diversity

Going into the diversity of datasets, when analyzing the MARVEL dataset, as shown in Figure 4, there is a significant imbalance between the number of samples in each class, as

this is a natural reflection of a realistic environment, where there are many more ships of one type than of another. In addition, some ship types are pretty similar in size and shape, while others are remarkably different, as shown in Figure 5, which can lead to a better classification between very distinct classes and not so good for similar ones. Finally, there are also pose, brightness, background clutter and scale variations, which can negatively influence the visual system performance.

The diversity of a dataset is based on the visual variation of its samples. In the case of ship images, the most common differences between samples are variations in background, scale, position, illumination, quality, size, viewpoint, and possible occlusions. These variations can be caused by several elements, such as the distance and position chosen for capturing the photo, the capture devices themselves, and the climatic and environmental conditions.

The detection and classification models must maintain a certain sensitivity to these differences, providing stable results, even with the complexities found in maritime environments. Therefore, the data used for training and validation of image processing architectures must have diversity so that the architectures can adapt to all these influences during training [112].
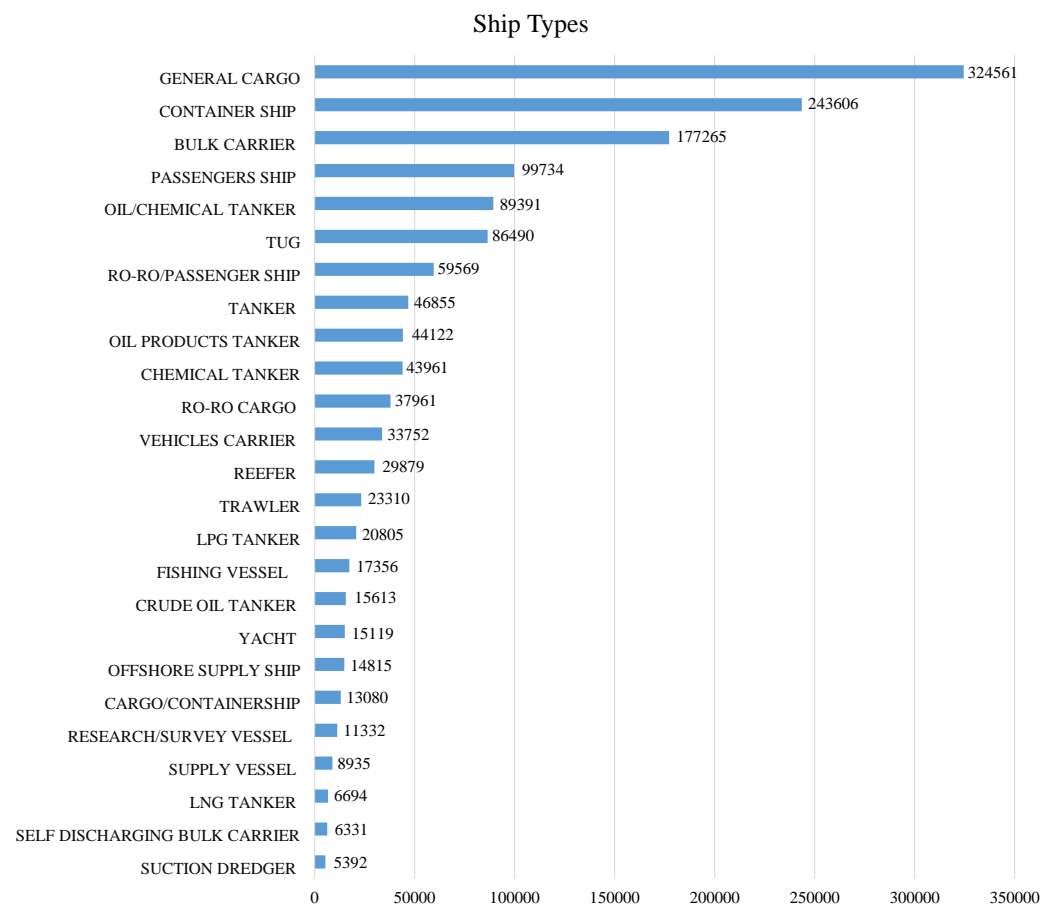
Figure 4. Samples by classes in the MARVEL Dataset.

Figure 5. Examples of classes in MARVEL dataset.

3.1.1. Background and Lighting

The information present in an image is used in the most diverse computer vision tasks. When considering, for example, face recognition, the separation between a front face and the background is easily performed by a background subtraction algorithm and generates a low computational cost due to the standard geometric shape of the face [117].

In the case of ships, there is a greater diversity in formats, sometimes even within the same class. This means that a single ship often can have more than one tag since its characteristics become confused with those of the environment. To solve this problem, there exist some techniques, such as the non-maximum suppression algorithm, that help avoid excessive tags by eliminating overlapping regions [118].

Another factor that can mainly compromise the detection stage is the lighting present in the images, which can sometimes cause objects in the scene to be mistaken as part of the ships and thus generate problems during the training.

3.1.2. Scale and Spatial Vision

During training, the images of the ships collected have their characteristics and patterns used to build models that will make the classification process. Therefore, the scale of the ships within the image is of the utmost importance, as tiny images can contain a limited richness of detail.

Generally, the datasets and applications are separated into two classes of viewpoints: those that work with side view images, i.e., datasets where the pictures were taken from the sides of the ships as shown in Figure 6, and remote sensing images, which work with pictures taken from the top of the ships, usually with images taken by satellites [7].

Regardless of the type of application, whether the side view or remote view is chosen, once the model is trained with images of greater diversity, it is able to better generalize each of the classes, becoming more capable for use in a real scenario. If the system is to identify ships always from the same point of view, the choice of training samples always in the same position may be better to obtain good results.
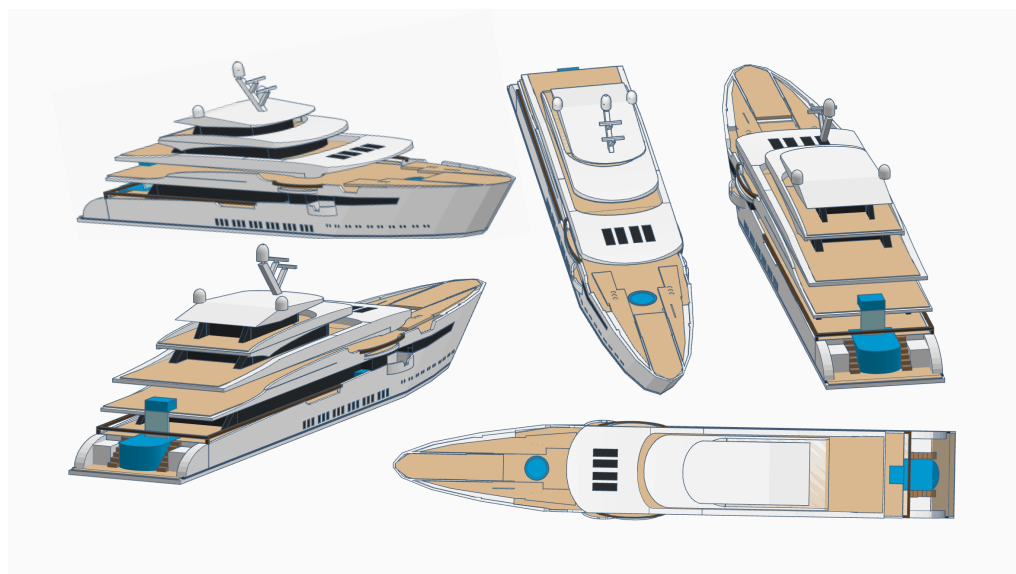
Figure 6. Different types of ship views.

3.1.3. Size, Quality, and Resolution

Generally, most models benefit from the larger size, quality, and resolution of images when the application does not involve storage, time, or processing power. This is because when a model receives higher quality images, it is able to extract more characteristics from the objects in it, which consequently can positively influence the assertiveness of the architecture.

Images with low pixel count, or even blurred images, such as Figure 7, may cause the system to be unable to extract the characteristics necessary to separate the classes. Consequently, the resulting architectures may have a degree of assertiveness lower than expected. This is usually the reason for applying the preprocessing step to the images before the detection or classification process [9].


Figure 7. Comparison of different image resolutions.

3.1.4. Occlusion and Position

Because the image collection is performed both offshore, in harbors or at satellites, a ship may appear partially within the image. This can be caused by the ship or by occlusions, which might be caused by other objects or even by other ships, shown in Figure 8.

Thus, according to the authors of [112], one should not ignore occlusion. Instead, it should be considered so that the trained model handles the occlusions presented in the validation step. At the same time, some care must be taken so that the position of the ship within the frame still preserves features that contribute to the training. Similarly, the partial occlusion of some objects must also preserve features of the original ship. Otherwise, these samples can directly interfere with an architecture's ability to perform good training.
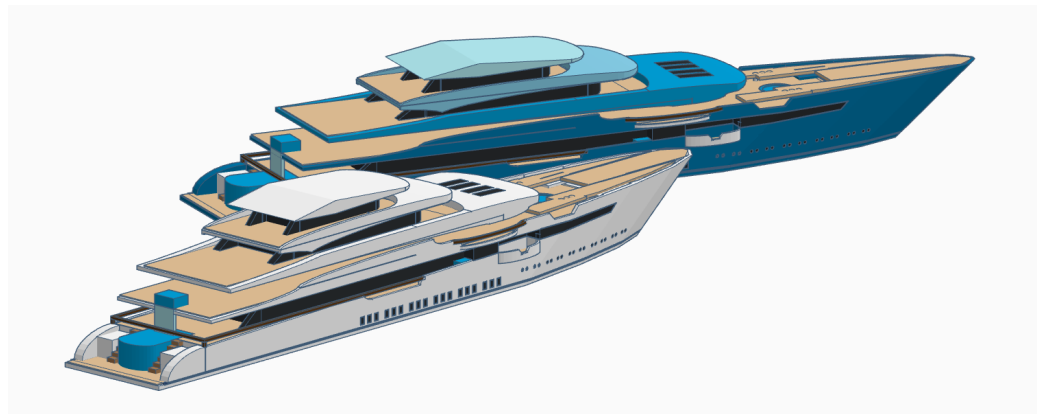
Figure 8. Example of occlusion.

3.1.5. Annotations and Labels

In machine learning and deep learning dataset, an annotation is a file that contains some data information referring to the image. The primary information stored in this file are the coordinates of the object's spatial position within the image and the class to which the object belongs [119].

The labeling process can be carried out using annotation tools, either manually or with some automatic processing. When annotating an image, each image's metadata are added to the dataset. Some of the datasets used in the literature already have annotations that can be used during the model training stage [112].

These annotations are essential because they allow models to understand where an object is positioned within a given image and its classification, so both detection and classification models can use these data as a reference when adjusting their weights during the training phase. With this, the model considers only the area of interest in the image. Then, after the model is trained on the labeled images, it later uses these training weights to identify these classes in new, previously unseen images.

These annotations usually come in separate files that accompany the images. However, the simplest ones are generated with just the four boundary points, which are used to build the box that marks the position of the ship, as shown in Figure 9. With this, the models have access to the object's position within the image and to which class it belongs. Either detection models or simply direct classification models can use these data as a reference when adjusting weights during training. This enables the model to identify the parts of interest in the image. Once the model is trained with the labeled images, it uses these training data to later identify these classes in new and previously unseen images.
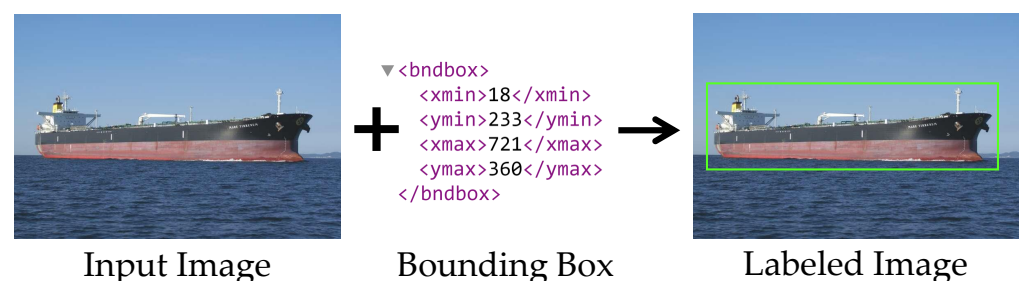
Figure 9. Example of a bounding box.

The annotation files accompanying the images are usually separated in another format, such as ".xml". The simplest ones come with four positions, which are the boundaries used to construct the box that marks the position of the ship, as shown in Figure 9. Some more complex annotations may contain multilevel classifications, segmentation data, and multiple object tags.

Along with the four coordinates, usually, the class to which the ship belongs to is also described so that the system can use these data when building the model. Files with more complex annotations can also bring climate, relative humidity, latitude, and longitude positions of the ship. However, most detection and classification systems disregard that information and only need the object's classes and bounding boxes to perform the training.

Annotations enrich the information about the object in the image so that the detection model is able to learn from this information. To this end, numerous annotation software programs are widely used as crucial tools for preparing images for training. These tools were developed because of the increasing demand for training data and are widely employed.

Table 4 shows a comparison of the tools, which are divided into categories. The working environments being local, those tools that require the software to be installed on a machine, and browser, those that can be only used through a web-browser. The tools can be used either in online or offline modes, as shown in the Table 4. Each tool has different characteristics, such as the processing data that are the input data that the tool annotates, e.g., images and videos, and can even support 3D point cloud annotations, commonly used by radio detection and ranging (RADAR) and light detection and ranging (LiDAR) sensors.

Another essential feature is that these tools offer different types of annotations, being the most used polygons and rectangles. However, some tools can even offer brushes and pencils to draw each object differently. Each tool offers different file formats for saving, that is, the format in which the annotations for each image or video will be saved to be used during training.

Semi-automatic labeling tools delimit objects in an image or video using a pre-trained detection model. The advantage obtained in this process is the time savings compared to manual labeling. The result is a pre-labeled set of images, which allows the user to perform subsequent tasks, such as checking and correcting labels already created or even training new models with the semi-automatically labeled samples. Table 4 compares different labeling tools that use manual and semi-automatic methods, as well as describing the operating characteristics of the different approaches.

The availability in the table refers to either paid or free tools. Furthermore, some remarks can help choose the labeling tool, such as the online support service that some offer.

In the works involving the classification task, the ships are divided into classes, whether directly applied to classify an image or even after a localization. However, within the area of ships monitoring, there are few specific standards and regulations for autonomous marine systems, which already use this detection technology with sensor systems [43]. There are some agencies to assist in the creation, regulation, and control of these systems, such as the International Organization for Standardization (ISO), International Maritime Organization (IMO), International Association for Marine Electronics Companies (CIRM) International Association of Classification Societies (IACS), International Electrotechnical Commission (IEC), International Association of Marine Aids to Navigation and Lighthouse Authorities (IALA), European GNSS Agency (GSA), International Telecommunication Union (ITU) and various classification societies themselves [43].

Table 4. Comparison of tools for labeling images and videos.

Tools	Environment	Conectivity	Processing Data	Annotation Types	Output Data	(Semi)Automatic Labeling	Availability	Remarks
ImgLab [120]	Browser and local	On/Offline	Images	Points, circles, rectangles, and polygons.	dlib XML, dlib pts, VOC, and COCO	No support	Free	-
VoTT [121]	Browser and local	On/Offline	Images and videos	Rectangles and polygons.	CNTK, Azure, VOC, CSV, and VoTT(JSON)	Support	Free	-
CVAT [122]	Browser and local	On/Offline	Images and videos	Points, lines, cuboids, rectangles, and polygons.	VOC, COCO, etc.	Support	Free	-
Labelimg [123]	Local	Offline	Images	Rectangles.	VOC, YOLO, and CSV.	No support	Free	-
Labelme [124]	Local	Offline	Images and videos	Points, circles, lines, rectangles, and polygons.	VOC, COCO, etc.	No support	Free	-
VGG Image Annotator (VIA) [125]	Browser	On/Offline	Images, videos, and audios	Points, circles, lines, ellipses, rectangles, and polygons.	VOC, COCO, and CSV	No support	Free	-
SuperAnnotate [126]	Browser and local	On/Offline	Images, videos, and texts	Points, lines, ellipses, cuboids, rectangles, polygons, and brushes.	JSON and COCO	Support	Paid	Online support
Supervisely [127]	Browser and local	On/Offline	Images, videos, and 3d point cloud	Points, lines, rectangles, polygons, and brushes.	JSON	Support	Paid	Online support
MakeSense [128]	Browser	Online	Images	Points, lines, rectangles, and polygons.	YOLO, VOC, and COCO	Support	Free	-
LabelBox [129]	Browser	Online	Images, videos, and text	Points, lines, rectangles, polygons, and brushes.	JSON and CSV	Support	Paid	Online support
DarkLabel [130]	Local	Offline	Images and videos	Rectangles.	VOC and YOLO	Support	Free	Online support
Autoannotation [131]	Browser	On/offline	Images	Rectangles.	YOLO	Support	Free	-

As the control agencies have not yet created definitive standards, each author follows the class division that best suits their work. Even though they still do not follow a pattern, some of these classes are found more frequently than others in the datasets of related works, as shown in Table 5. In this table, the works that use the separation of ships into classes are presented with their respective classes. The amount of images divided for training and testing in each of the works is also presented.

Table 5. Class division in related works.

Papers	Type	Classes	Train	Test
2008 [132]	2	Aircraft Carrier and Destroyer	-	270
2009 [133]	4	Carrier, Cruiser, Destroyer and Frigate	-	98
2010 [134]	4	Ark Royal, Arizona, Arleigh and Connelly	-	32
2017 [42]	12	Military Ships (Aircraft Carrier, Submarine, San Antonio, Arleigh Burke, Whidbey Island)	200	80
2017 [56]	5	Containers, Fishing Boats, Guards, Tankers, Warships	-	300
2018 [60]	9	Passenger Ship, Leisure Boat, Sailing Boat, Service Vessel, Fishing Boat, Warship, Generic Cargo Ship, Container Carrier and Tanker.	30000	20
2018 [62]	3	Cargo Ship, Cruise and Yacht	-	-
2019 [68]	4	Barge, Cargo, Container and Tanker	-	-
2019 [64]	3	Oil Tankers, Bulk Carriers and Container Ships	-	-
2020 [77]	7	Aircraft Carrier, Destroyer, Cruiser, Cargo Ship, Medical Ship, Cruise Ship and Transport Ship.	24	6
2020 [73]	3	Passenger Ships, General Cargo Ships and Container Ships	-	-
2020 [74]	4	Destroyers, One Bulk Barrier, Submarine and Two Aircraft Carriers	-	500
2020 [78]	2	Fishing Ships and Military	398	16
2020 [81]	23	Non-ship, Aircraft Carrier, Destroyer, Landing Craft, Frigate, Amphibious Transport Dock, Cruiser, Tarawa-Class Amphibious Assault Ship, Amphibious Assault Ship, Command Ship, Submarine, Medical Ship, Combat Boat, Auxiliary Ship, Container Ship, Car Carrier, Hovercraft, Bulk Carrier, Oil Tanker, Fishing Boat, Passenger Ship, Liquefied Gas Ship and Barge	5165	825
2021 [83]	4	Warcraft, Aircraft Carrier, Merchant Ship and Submarine	-	-
2021 [86]	6	Warship, Container Ship, Cruise Ship, Yacht, Sailboat and Fishing Boat	-	-
2021 [91]	15	Aircraft Carrier, Oliver Hazard Perry Class frigate, Ticonderoga-class Cruiser, Arleigh Burke Class Destroyer, Independence-class littoral combat ship, Freedom-class littoral Combat Ship, Amphibious Assault Ship, Tanker, Container Ship, Grocery Ship, Amphibious Transport Ship, Small Military Warship, Supply Ship, Submarine and Other.	4800	1200
2021 [3]	8	Bulk Cargo Ships, Engineering Ships, Armed Ships, Refrigerated Ships, Concrete Ships, Fisheries Vessels, Container Ships and Oil Tankers	-	-

4. Challenges and Issues

This section is established based on proposals for future work from a review of the literature as well as the related works already mentioned above. Among the main problems, challenges, and research opportunities cited are those related to datasets, image processing techniques, data fusion, and practical applications. In addition, some recent works, such as [81], also point to some of these problems, which will be discussed throughout this chapter.

4.1. Datasets

Datasets represent an important part in the construction of object location, classification and detection models. In vessel datasets, it is possible to find problems common to other datasts used in automatic target recognition (ATR) problems, such as overlapping objects. However, other problems encountered, such as the high similarity between different classes of vessels, deterioration of ships and a great variety of models for each class, are inherent problems, or even more common in maritime environments than in computer vision problems in general.

There is some difficulty in accurately finding the object when the image has a considerable background complexity [77]. In [5], a maritime ship tracker is proposed, but the authors state that the proposed tracker can only work in certain weather conditions and

only for some types of ships. In [135], the authors also state that the presented technique shows errors in specific scenarios where the sea color is drastically changed or when the horizon line suffers partial occlusion by other objects.

Based on these problems, the authors of [136] make a practical study of detection with several architectures, such as faster R-CNN, YOLOv2, and YOLOv3 in datasets of images with weather and lighting interference to evaluate the accuracy of the models. In [137], the authors explain that images related to the maritime scene suffer several influences related to weather and lighting factors, resulting in unclear targets in the image.

The proposed solution presented in the study is to attack the problem on three fronts, improving the image acquisition hardware technology, creating an image preprocessing step, and increasing the dataset used for training, with images that have multiple targets and high diversity [137]. Regarding the issue of diversity, the use of datasets in the CNN models must have good image quality and also represent the shapes of the objects, which are taken from multiple sides [78].

Many authors point out that the lack of substantially extensive datasets hinders the construction of their models. The search for high-quality datasets is a shared objective that ranges from authors who develop simpler models to those who seek to validate their systems in more complex environments. For example, the authors of [39] claimed that the popularization of high-resolution remote sensing data could make the proposed method widely applicable. In [65], the authors claimed that they will compare their proposed model with state-of-the-art results while expanding the datasets. In [138], the authors also stated that they will make efforts to expand the dataset to try to obtain a robust detection of the system. Even some work on more recent detector enhancements, such as the YOLOv5, still points out that they intend to perform retraining on large datasets to evaluate the new results [90].

Some works, such as [42], point to good results for the task of automatically locating and recognizing coastal ships in remote sensing images of large scenes. However, they state that they have their efforts in the development of new multimodel methods capable of recognizing more types of ships and that for this, it is necessary to obtain samples of other classes of ships. In this search for an increase in the number of classes capable of being recognized by the models, another problem faced and reported by other works, including those focused on new datasets, is the imbalance of samples in each class. This can be easily seen when when internally analyzing the structure of large databases such as MARVEL, for example [103]. Thus, some authors, such as [83], reaffirm this problem, citing that the efforts of their works have been to reduce the class imbalance, feeding the less favored classes with more samples, thereby decreasing the risk of overfitting that can be caused by the imbalance during the training of the models. The authors of [70] also pointed out the risks of overfitting when the dataset contains many small or poor-quality images.

As an alternative to the difficulties presented by the authors regarding limited databases, adding bad images to the model, or even the imbalance between classes, Ref. [68] suggested leveraging synthetically generated images to compose the training data since they reinforced the idea that CNNs outperform classical object recognition methods when provided with enough data for good training. In their study, they demonstrated that the same ship classifier trained on a bank of real-only images performs worse compared to the same classifier trained on that same dataset with the addition of synthetic images [68].

A second alternative to the limitation posed by the dataset would be transfer learning. Transfer learning is a machine learning technique that stores the knowledge gained from solving a problem and applying it to a different but minimally related problem [139]. For example, Ref. [140] presents the application of this type of technique in a ship recognition task in infrared images. With this, even if there is an imbalance between the samples of each class, it is possible to improve the model's performance. The work presented in [64], which uses visible light remote sensing images, also suggests that transfer learning solves the limitation of the number of images on datasets and improves the convergence speed of the model.

Finally, the last challenge pointed out by the authors in the context of the dataset is about images annotated with precise bounding boxes to provide an effective and available database for training and validation. The idea is to try to reduce as much as possible the images that are in the wrong classes or with no boat present, called negatives. With that, the result of the complete training or transfer learning tends to improve even more [110]. Therefore, a second proposal made in [111] is, instead of discarding negative images, using them during training to explore the effect of these samples within the system in order to develop a more robust algorithm.

4.2. Image Processing Techniques

The techniques chosen to build object localization, classification, and detection systems are another essential part of the system through which the received images are converted into information. In the tests performed in [141], a large dataset of images captured by several optical satellite sensors was submitted to a hierarchical classification architecture, which was able to eliminate candidate regions belonging to objects that did not represent ships. Moreover, other hierarchical classification techniques can be applied with newer architectures. The authors admit the need to improve the model or refine the training parameters to improve the detector. This would be another way to mitigate problems related to complex backgrounds in the received images, now improving the technique instead of dataset changes.

When discussing improvements improvements in detectors and classifiers, it is also possible to find several works that advocate this idea, such as [63,89], which followed the line of improving or adjusting parameters in the model to increase accuracy, instead of improving with the evolution of the dataset only. In [142], the authors advocated increasing the dataset and the number of classes provided for training. Another objective is to adapt the detection models by changing their parameters and the architectures themselves to compare their results to the original ones and verify the accuracy increase.

Another work that aims for future improvements by improving the architecture is the one presented in [143], where the authors used the SSD detection techniques, aiming for automation in the container terminal. In [86], the authors also relied on the optimization of existing methods, where the YOLOv3 architecture is optimized to detect ships at a higher frame rate without sacrificing detection loss. Other work also has as future tasks the optimization of the ship target recognition capability so that the entire model performance can be further improved [74]. The experiment data are part of public Google Earth data and commercial satellite imagery [74]. To select these parameters that optimize the system, Ref [144] stated that the empirical way can work very well, but stated that a more systematic way to select these parameters can be a target of future research.

Furthermore, regarding the techniques, there are some recent works such as [145,146], which operate on a sky-sea basis, i.e., using the dividing line between sea and water to help locate the ship, and have interesting effectiveness for open sea applications. The work [147] suggests research that combines CNNs with handcrafted techniques to perform a sea–land separation, decreasing the application restriction of these systems. The work [148] also intends to deal with the problem of low contrast that sometimes occurs in a dynamic ocean scenario, which generates waves with a different reflected color tone than expected in a pixel analyzed in the image where the ocean is present.

Among the works that present possibilities for combining techniques for future research, there is the case of [149], which introduced a method to exclude confusing samples and thereby reduce the problem of overlapping classes. Future work in this paper aims to integrate this method with other techniques. This is also the research theme in [81], but instead of using features from other models in their own, the authors performed the inverse process. They proposed two feature representation schemes that can be incorporated into most CNN models and bring an increase in the classification performance of the models, taking advantage of the possibility of end-to-end training. The work basis in a new bench-

mark and an attribute guided multilevel feature representation network for fine-grained ship classification in optical remote sensing image [81].

Besides the combination of multiple techniques in a single architecture, it is also possible to find works that point to future tasks, the inclusion of preprocessing models. These tasks range from simpler challenges, such as cropping, aligning and resizing images [115], to even more complex ones, such as noise removal [150–152] and image super-resolution systems [12,16]. Enhanced images, generated by these types of techniques, increase successful detection and reduce false detection [153]. Based on that, Ref. [154] suggested as future work the use of super-resolution GAN models to improve image quality and thus be able to identify attributes over long distances.

The performance of a ship ATR system is not only defined by the chosen algorithm, but also by the image quality and the result generated by the feature extraction technique. By improving the processing model and the image quality, the target recognition rate will be improved [155]. For example, Ref. [66] proposes that in the future, their segmentation-based model be used by another one for image preprocessing so that the final architecture can be used in different weather or light conditions. In parallel, the authors also hope to incorporate distance estimation algorithms to contribute to research on autonomous surface vehicles. Similarly, Ref. [156] also aims to develop detection models for target tracking.

4.3. Data Fusion

Within the literature, there are some works, such as [43], which already explore the situational awareness field. In this type of approach, the optical sensors do not act in isolation, and there is a fusion of data with other sources of information, which allows the generation of a positioning map of ships in an autonomous way.

For this map to be generated, there is a global effort regarding the creation of regulations and standards which allow the reliability and integrity of the information generated. Furthermore, this type of approach aims to design an artificial intelligence (AI) algorithm capable of merging the different types of sensors and information sources. The idea of this fusion is to implement a system capable of obtaining a positioning inference of less than 3 m, as well as collaborating for autonomous navigation [43].

The current availability of sensors capable of collecting information at different levels of an object allows observations made from different acquisition sources to be combined to obtain a more detailed description of the scene. For example, Ref. [157] presents as a focus of future studies the increase in the robustness of the system through multispectral remote sensing, aiming to identify ships that are close to land.

Each of the sensor types has its advantages and disadvantages. Among them, it is possible to mention costs, dependence on the angle of the observation aspect, variation of resolution with distance, susceptibility to atmospheric influences [158]. The differences and advances in the different types of sensors mean that research into new image processing methods and tools never stands still. Just as each data source can generate unique information, combining features generated by these different sources makes sense. The systems developed for this purpose propose the use of fusion techniques during this processing chain in order to obtain at the end a situational awareness of the region or maritime object under analysis [158].

Some works in the literature already bring some of these fusion proposals. For example, in [159], the authors create an autonomous collision avoidance system using a fusion of different sensor sources as global positioning system (GPS), RADAR, LiDAR, automatic identification system (AIS), and optical sensors, but say that more studies are needed in various conditions of real maritime traffic to verify stability and robustness.

In [109], the authors also intend to evaluate the benefits of autonomous navigation and the improvement of navigation safety through tracking techniques. In this line, Ref. [160] reinforces that some more handcrafted image processing methods can reduce computational costs during tracking, that is, the continuous observation of navigation. In [60], the focus of future work is on the use of data involving AIS. In [161], the authors propose the

combination of radar information in addition to AIS and tracking to evaluate suspicious activities within the maritime scenario. The work [162] also considers the integration of the detection method with RADAR and AIS systems for future implementations.

From more recent works, such as [43], to works more than a decade old, such as [163], they all have the exploration of complete surveillance solutions as a common point. Whether they are based on electro-optical and infrared sensors, with the use of multiple image processing techniques, or works that bring together data from various families of sensors, the ultimate focus of the system should be to support the fusion of information to interpret scene activity, associate targets with the offense committed or the threat they correspond to, and generate situational information to a control center [163].

4.4. Practical Applications

Within the entire literature review, several works cited that they aim at the practical implementation of their systems to provide the collection of visual information in the maritime scenario [164]. For this to be possible, it is necessary to evaluate several variables, and computational capacity is usually one of the most important deciding factors when evaluating the feasibility of practical implementations. For this reason, some works that aim to create computer vision systems that run in real time, such as [165], make several analyses related to processing time and the use of devices with different computational capacities.

Some authors, such as [157], conclude that their system responds more efficiently if their candidate region search method is applied offline. However, the algorithm must be employed to create a real-time situational view. Therefore, papers such as [166], which demonstrate operation only for still images, suggest using it on continuous video streams. In a review in [110], the authors suggest performing parallel data collection in addition to processing. They claim that to develop systems for autonomous surface ships, datasets with images collected and annotated from cameras installed on moving ships are needed. With this, the dataset used to train the computer vision model will be extremely close to the maritime scenario in which the autonomous ship will navigate, making the system more robust than if the model only trained it with images of ships docked in ports.

In [56], the authors say that the architecture adopted there can be applied to real-time computer vision problems by installing cameras at harbors. They also say that systems will store the images collected during operation to retrain the system, making it more robust with samples of different degrees of illumination captured throughout the day.

When the scenario of practical implementations is explored, it is possible to find works where an embedded target detection system was implemented using the YOLOv4 architecture [79]. However, the authors themselves say that there is room for further improvement in the detection rate in the experiment and that they intend to retrain the same system with more data and other categories to achieve better results.

Finally, the authors of [6] propose and implement an object detection algorithm for maritime surveillance, where low processing power embedded systems are the focus of the application. The processing architecture was built to reduce the volume of data processed in maritime surveillance systems. In this work, future research directions focus on designing lighter weight detection architectures to achieve good performance, even on computationally limited devices.

5. Conclusions and Future Work

This paper presents a review of ship localization, classification, and detection methods based on optical sensors. This literature review made it possible to find the main challenges and open problems, besides exploring the techniques and architectures used by several authors.

It is possible to state that CNNs have been explored with greater intensity over the years regarding processing techniques. It is possible to recognize the advancement of this type of technique with the observation of high-precision architectures, with the ability to detect small objects, even in scenarios with noise and other sources of interference.

Moreover, the evolution of the computational capacity of devices allows these techniques to be employed in practical applications, replacing the need for human intervention in several tasks [88].

Still exploring the techniques and models already devised, several detection algorithms and practical maritime decision-making systems should be applied to the same dataset to evaluate all works with the same metrics and datasets, just as several face verification works use labeled faces in the wild (LFW) as a standard benchmark [167]. As in [62], several authors have already said that they intend to verify this scalability of the technique on other datasets. This idea is indeed valid since each of the datasets brings some particularities that are new challenges for the model.

Research within datasets has shown that there is no uniformity in their use. Each author works with their database. Therefore, more effort is needed to create large-scale datasets that are readily available, so that the community can begin to have a more reliable standard of comparison. About labeling, many authors claim to have done it manually. Therefore, the ideal would be to explore CNNs that could be adapted to generate the bounding boxes or pixel-by-pixel labeling automatically or semi-automatically.

The combination of deep learning and navigation data has the potential to solve maritime situational awareness problems. This task is quite challenging but of equal or greater importance for applications in maritime environments. The situational awareness of all objects present can bring many benefits to any system. Different architectures are proposed to improve the ability of automatic target recognition and search, each with its advantages and disadvantages.

The research involving practical implementations is mainly based on the fact that all the extra tasks of localization, simple or hierarchical classification, a combination of detection techniques, preprocessing, AIS, and data fusion may increase the amount of input data and results in an additional computational cost. From this arises the focus of analyzing these computational costs based on the available technological conditions.

Therefore, it is possible to conclude that all research efforts within the literature review fall within the following four research lines:

1. Creating large-scale fine-grained datasets with higher diversity and already labeled samples, using synthetic data, and improving the balance between classes.
2. Creating, optimizing, and combining image processing techniques, including preprocessing and the use of transfer learning or similar techniques.
3. Usage of different sensors and data sources to operate in conjunction with the optical sensors, thereby generating a situational awareness of the monitored maritime region.
4. Practical analysis of the systems, indicating their performance and speed in real scenarios, where the complexity may be higher than in the datasets.

With this, it is possible to conclude that this paper describes the main open problems pointed out by the literature, aiming to influence the research of new work and better delimit the challenges to be overcome.

Author Contributions: E.T. conducted the general investigation; E.T. wrote the first draft; B.A. conducted the image labeling investigation, V.C. conducted the investigation of the preprocessing techniques; E.T., S.M. and F.F. reviewed and edited the draft. All authors have read and agreed to the published version of the manuscript.

Abbreviations

The following abbreviations are used in this manuscript:

ATR	Automatic Target Recognition
AIS	Automatic Identification System
B2RB	Bounding-Box to Rotated Bounding-Box
BLS	Broad Learning System
BN	Batch Normalization
CNN	Convolutional Neural Network
DCT	Discrete Cosine Transform
DeFMO	Deblurring and Shape Recovery of Fast Moving Objects
DeblurGAN	Deblur Generative Adversarial Network
EDSR	Enhanced Deep Super-Resolution Network
ESRGAN	Enhanced Super-Resolution Generative Adversarial Network
FPN	Feature Pyramid Network
FusionNet	Fusion Network
GAN	Generative Adversarial Network
HOG	Histograms of Oriented Gradients
IACS	International Association of Classification Societies
IALA	International Association of Marine Aids to Navigation and Lighthouse Authorities
IEC	International Electrotechnical Commission
IEEE	Institute of Electrical Electronic Engineers
IMO	International Maritime Organization
IR	Infrared
ISO	International Organization for Standardization
ITU	International Telecommunication Union
LBP	Local Binary Pattern
LFW	Labeled Faces in the Wild
LiDAR	Light Detection and Ranging
mAP	Mean Average Precision
MobileNet-DSC	Mobile Network-Depthwise Separable Convolution
MDSR	Multi-Scale Deep Super-Resolution
MLP	Multi-Layer Perceptron
MOS	Mean Opinion Score
MSE	Mean Squared Error
NIQE	Natural Image Quality Evaluator
PIRM-SR	Perceptual Image Restoration and Manipulation—Super Resolution
PSNR	Peak Signal-to-Noise Ratio
RADAR	Radio Detection And Ranging
RaGAN	Relativistic average GAN
RankSRGAN	Rank Super-Resolution Generative Adversarial Network
ResNet	Residual Network
ResBlock	Residual Block
RetinaNet	Retina Network
RNN	Recurrent Neural Network
RoI	Region of Interest
RRDB	Residual-in-Residual Dense Block
RPN	Region Proposal Network
R-CNN	Region Based Convolutional Neural Network
LiDAR	Light Detection and Ranging
SIFT	Scale-Invariant Feature Transform
Skip-ENet	Skip Efficient Neural Network
SNN	Spiking Neural Networks
SR	Super-Resolution
SRCNN	Super-Resolution Convolutional Neural Network
SRGAN	Super-Resolution Generative Adversarial Network

SS	Selective Search
SSD	Single-Shot Detector
SSIM	Structural Similarity Index Measure
SSS-Net	Single-Shot Network Structure
SRResNet	Super-Resolution Residual Network
SVM	Support Vector Machine
VGG	Visual Geometry Group
YOLO	You Only Look Once

References

1. Park, J.; Cho, Y.; Yoo, B.; Kim, J. Autonomous collision avoidance for unmanned surface ships using onboard monocular vision. In Proceedings of the OCEANS 2015—MTS/IEEE Washington, Washington, DC, USA, 19–22 October 2015; pp. 1–6. [CrossRef]
2. Dumitriu, A.; Miceli, G.E.; Schito, S.; Vertuani, D.; Ceccheto, P.; Placco, L.; Callegaro, G.; Marazzato, L.; Accattino, F.; Bettio, A.; et al. OCEANS-18: Monitoring undetected vessels in high risk maritime areas. In Proceedings of the 2018 5th IEEE International Workshop on Metrology for AeroSpace (MetroAeroSpace), Rome, Italy, 20–22 June 2018; pp. 669–674. [CrossRef]
3. Yue, T.; Yang, Y.; Niu, J.M. A Light-weight Ship Detection and Recognition Method Based on YOLOv4. In Proceedings of the 2021 4th International Conference on Advanced Electronic Materials, Computers and Software Engineering (AEMCSE), Changsha, China, 26–28 March 2021; pp. 661–670. [CrossRef]
4. Liu, H.; Xu, X.; Chen, X.; Li, C.; Wang, M. Real-Time Ship Tracking under Challenges of Scale Variation and Different Visibility Weather Conditions. *J. Mar. Sci. Eng.* **2022**, *10*, 444. [CrossRef]
5. Shan, Y.; Zhou, X.; Liu, S.; Zhang, Y.; Huang, K. SiamFPN: A Deep Learning Method for Accurate and Real-Time Maritime Ship Tracking. *IEEE Trans. Circuits Syst. Video Technol.* **2021**, *31*, 315–325. [CrossRef]
6. Duan, Y.; Li, Z.; Tao, X.; Li, Q.; Hu, S.; Lu, J. EEG-Based Maritime Object Detection for IoT-Driven Surveillance Systems in Smart Ocean. *IEEE Internet Things J.* **2020**, *7*, 9678–9687. [CrossRef]
7. 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]
8. Mandalapu, H.; Reddy P N, A.; Ramachandra, R.; Rao, K.S.; Mitra, P.; Prasanna, S.R.M.; Busch, C. Audio-Visual Biometric Recognition and Presentation Attack Detection: A Comprehensive Survey. *IEEE Access* **2021**, *9*, 37431–37455. [CrossRef]
9. Zhang, X.; Wang, F.; Dong, H.; Guo, Y. A Deep Encoder-Decoder Networks for Joint Deblurring and Super-Resolution. In Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 15–20 April 2018; pp. 1448–1452. [CrossRef]
10. Talab, M.A.; Awang, S.; Najim, S.A.d.M. Super-Low Resolution Face Recognition using Integrated Efficient Sub-Pixel Convolutional Neural Network (ESPCN) and Convolutional Neural Network (CNN). In Proceedings of the 2019 IEEE International Conference on Automatic Control and Intelligent Systems (I2CACIS), Selangor, Malaysia, 29 June 2019; pp. 331–335. [CrossRef]
11. Robey, A.; Ganapati, V. Optimal physical preprocessing for example-based super-resolution. *Opt. Express* **2018**, *26*, 31333. [CrossRef]
12. Yang, Z.; Shi, P.; Pan, D. A Survey of Super-Resolution Based on Deep Learning. In Proceedings of the 2020 International Conference on Culture-oriented Science Technology (ICCST), Beijing, China, 30–31 October 2020; pp. 514–518. [CrossRef]
13. Xie, J.; Xu, L.; Chen, E. Image Denoising and Inpainting with Deep Neural Networks. In Proceedings of the Advances in Neural Information Processing Systems (NIPS 2012), Lake Tahoe, NV, USA, 3–6 December 2012; Pereira, F., Burges, C.J.C., Bottou, L., Weinberger, K.Q., Eds.; Curran Associates, Inc.: New York, NY, USA, 2012; Volume 25.
14. Sada, M.; Goyani, M. Image Deblurring Techniques—A Detail Review. In Proceedings of the National Conference on Advanced Research Trends in Information and Computing Technologies (NCARTICT-2018), Ahmedabad, Gujarat, India, 20 January 2018; pp. 176–188.
15. Kotsiantis, S.; Kanellopoulos, D.; Pintelas, P. Data Preprocessing for Supervised Learning. *Int. J. Comput. Sci.* **2006**, *1*, 111–117.
16. Yang, W.; Zhang, X.; Tian, Y.; Wang, W.; Xue, J.H.; Liao, Q. Deep Learning for Single Image Super-Resolution: A Brief Review. *IEEE Trans. Multimed.* **2019**, *21*, 3106–3121. [CrossRef]
17. Blau, Y.; Mechrez, R.; Timofte, R.; Michaeli, T.; Zelnik-Manor, L. The 2018 PIRM challenge on perceptual image super-resolution. In Proceedings of the European Conference on Computer Vision (ECCV) Workshops, Munich, Germany, 8–14 September 2018.
18. Dong, C.; Loy, C.C.; He, K.; Tang, X. Image super-resolution using deep convolutional networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *38*, 295–307. [CrossRef]
19. Zabalza, M.; Bernardini, A. Super-Resolution of Sentinel-2 Images Using a Spectral Attention Mechanism. *Remote Sens.* **2022**, *14*, 2890. [CrossRef]
20. Lim, B.; Son, S.; Kim, H.; Nah, S.; Mu Lee, K. Enhanced deep residual networks for single image super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, 21–26 July 2017; pp. 136–144.
21. Haris, M.; Shakhnarovich, G.; Ukita, N. Deep Back-ProjectiNetworks for Single Image Super-Resolution. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *43*, 4323–4337. [CrossRef]

22. Ledig, C.; Theis, L.; Huszár, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z.; et al. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 105–114. [CrossRef]

23. Wang, X.; Yu, K.; Wu, S.; Gu, J.; Liu, Y.; Dong, C.; Qiao, Y.; Change Loy, C. Esrgan: Enhanced super-resolution generative adversarial networks. In Proceedings of the European Conference on Computer Vision (ECCV) Workshops, Munich, Germany, 8–14 September 2018.

24. Zhang, W.; Liu, Y.; Dong, C.; Qiao, Y. Ranksrgan: Generative adversarial networks with ranker for image super-resolution. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 27–28 October 2019; pp. 3096–3105.

25. Chen, W.; Liu, C.; Yan, Y.; Jin, L.; Sun, X.; Peng, X. Guided Dual Networks for Single Image Super-Resolution. *IEEE Access* **2020**, *8*, 93608–93620. [CrossRef]

26. Setiadi, D.R.I.M. PSNR vs SSIM: Imperceptibility quality assessment for image steganography. *Multimed. Tools Appl.* **2021**, *80*, 8423–8444. [CrossRef]

27. Ieremeiev, O.; Lukin, V.; Okarma, K.; Egiazarian, K. Full-reference quality metric based on neural network to assess the visual quality of remote sensing images. *Remote Sens.* **2020**, *12*, 2349. [CrossRef]

28. Timofte, R.; Agustsson, E.; Van Gool, L.; Yang, M.H.; Zhang, L. NTIRE 2017 Challenge on Single Image Super-Resolution: Methods and Results. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Honolulu, HI, USA, 21–26 July 2017.

29. Xu, N.; Ma, D.; Ren, G.; Huang, Y. BM-IQE: An image quality evaluator with block-matching for both real-life scenes and remote sensing scenes. *Sensors* **2020**, *20*, 3472. [CrossRef]

30. Zhang, K.; Ren, W.; Luo, W.; Lai, W.S.; Stenger, B.; Yang, M.H.; Li, H. Deep image deblurring: A survey. *Int. J. Comput. Vis.* **2022**, pp. 1–28. [CrossRef]

31. Kupyn, O.; Budzan, V.; Mykhailych, M.; Mishkin, D.; Matas, J. DeblurGAN: Blind Motion Deblurring Using Conditional Adversarial Networks. In Proceedings of the 2018 IEEE/CVF International Conference on Computer Vision (ICCV), Salt Lake City, UT, USA, 18–23 December 2018; pp. 8183–8192. [CrossRef]

32. Kupyn, O.; Martyniuk, T.; Wu, J.; Wang, Z. DeblurGAN-v2: Deblurring (Orders-of-Magnitude) Faster and Better. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea, 27 October–2 November 2019; pp. 8877–8886. [CrossRef]

33. Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature Pyramid Networks for Object Detection. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 936–944. [CrossRef]

34. Rozumnyi, D.; Oswald, M.R.; Ferrari, V.; Matas, J.; Pollefeys, M. DeFMO: Deblurring and Shape Recovery of Fast Moving Objects. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021. [CrossRef]

35. Sun, J.; Cao, W.; Xu, Z.; Ponce, J. Learning a convolutional neural network for non-uniform motion blur removal. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 769–777.

36. Rahmani, N.; Behrad, A. Automatic marine targets detection using features based on Local Gabor Binary Pattern Histogram Sequence. In Proceedings of the 2011 1st International eConference on Computer and Knowledge Engineering (ICCKE), Mashhad, Iran, 13–14 October 2011; pp. 195–201. [CrossRef]

37. Zhang, Y.; Li, Q.Z.; Zang, F.N. Ship detection for visual maritime surveillance from non-stationary platforms. *Ocean. Eng.* **2017**, *141*, 53–63. [CrossRef]

38. Mutalikdesai, A.; Baskaran, G.; Jadhav, B.; Biyani, M.; Prasad, J.R. Machine learning approach for ship detection using remotely sensed images. In Proceedings of the 2017 2nd International Conference for Convergence in Technology (I2CT), Mumbai, India, 7–9 April 2017; pp. 1064–1068. [CrossRef]

39. Shuai, T.; Sun, K.; Shi, B.; Chen, J. A ship target automatic recognition method for sub-meter remote sensing images. In Proceedings of the 2016 4th International Workshop on Earth Observation and Remote Sensing Applications (EORSA), Guangzhou, China, 4–6 July 2016; pp. 153–156. [CrossRef]

40. Yang, F.; Xu, Q.; Li, B. Ship Detection From Optical Satellite Images Based on Saliency Segmentation and Structure-LBP Feature. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 602–606. [CrossRef]

41. Song, Z.; Sui, H.; Hua, L. How to Quickly Find the Object of Interest in Large Scale Remote Sensing Images. In Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 4843–4845. [CrossRef]

42. Li, W.; Fu, K.; Sun, H.; Sun, X.; Guo, Z.; Yan, M.; Zheng, X. Integrated Localization and Recognition for Inshore Ships in Large Scene Remote Sensing Images. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 936–940. [CrossRef]

43. Thombre, S.; Zhao, Z.; Ramm-Schmidt, H.; Vallet García, J.M.; Malkamäki, T.; Nikolskiy, S.; Hammarberg, T.; Nuortie, H.; Bhuiyan, M.Z.H.; Särkkä, S.; et al. Sensors and AI Techniques for Situational Awareness in Autonomous Ships: A Review. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 64–83. [CrossRef]

44. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. ImageNet Large Scale Visual Recognition Challenge. *Int. J. Comput. Vis. (IJCV)* **2015**, *115*, 211–252. [CrossRef]

45. Girshick, R. Fast R-CNN. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; pp. 1440–1448. [CrossRef]

46. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587. [CrossRef]

47. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [CrossRef]

48. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2980–2988. [CrossRef]

49. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. SSD: Single Shot MultiBox Detector. In *Computer Vision – ECCV 2016*; Springer International Publishing: Berlin/Heidelberg, Germany, 2016; pp. 21–37. [CrossRef]

50. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788. [CrossRef]

51. Redmon, J.; Farhadi, A. YOLO9000: Better, Faster, Stronger. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 6517–6525. [CrossRef]

52. Nepal, U.; Eslamiat, H. Comparing YOLOv3, YOLOv4 and YOLOv5 for Autonomous Landing Spot Detection in Faulty UAVs. *Sensors* **2022**, *22*, 464. [CrossRef]

53. Uijlings, J.R.R.; van de Sande, K.E.A.; Gevers, T.; Smeulders, A.W.M. Selective Search for Object Recognition. *Int. J. Comput. Vis.* **2013**, *104*, 154–171. [CrossRef]

54. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 3431–3440. [CrossRef]

55. Cheng, D.; Meng, G.; Xiang, S.; Pan, C. FusionNet: Edge Aware Deep Convolutional Networks for Semantic Segmentation of Remote Sensing Harbor Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 5769–5783. [CrossRef]

56. Kumar, A.S.; Sherly, E. A convolutional neural network for visual object recognition in marine sector. In Proceedings of the 2017 2nd International Conference for Convergence in Technology (I2CT), Mumbai, India, 7–9 April 2017; pp. 304–307. [CrossRef]

57. Fu, H.; Li, Y.; Wang, Y.; Han, L. Maritime Target Detection Method Based on Deep Learning. In Proceedings of the 2018 IEEE International Conference on Mechatronics and Automation (ICMA), Changchun, China, 5–8 August 2018; pp. 878–883. [CrossRef]

58. Li, M.; Guo, W.; Zhang, Z.; Yu, W.; Zhang, T. Rotated Region Based Fully Convolutional Network for Ship Detection. In Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 673–676. [CrossRef]

59. Liu, Y.; Cai, K.; Zhang, M.h.; Zheng, F.b. Target detection in remote sensing image based on saliency computation of spiking neural network. In Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 2865–2868. [CrossRef]

60. Voinov, S.; Krause, D.; Schwarz, E. Towards Automated Vessel Detection and Type Recognition from VHR Optical Satellite Images. In Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 4823–4826. [CrossRef]

61. Wang, Y.; Li, W.; Li, X.; Sun, X. Ship Detection by Modified RetinaNet. In Proceedings of the 2018 10th IAPR Workshop on Pattern Recognition in Remote Sensing (PRRS), Beijing, China, 19–20 August 2018; pp. 1–5. [CrossRef]

62. Zhang, Y.; You, Y.; Wang, R.; Liu, F.; Liu, J. Nearshore vessel detection based on Scene-mask R-CNN in remote sensing image. In Proceedings of the 2018 International Conference on Network Infrastructure and Digital Content (IC-NIDC), Guiyang, China, 22–24 August 2018; pp. 76–80. [CrossRef]

63. Zhang, Z.; Guo, W.; Zhu, S.; Yu, W. Toward Arbitrary-Oriented Ship Detection With Rotated Region Proposal and Discrimination Networks. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 1745–1749. [CrossRef]

64. Hui, Z.; Na, C.; ZhenYu, L. Combining a Deep Convolutional Neural Network with Transfer Learning for Ship Classification. In Proceedings of the 2019 12th International Conference on Intelligent Computation Technology and Automation (ICICTA), Xiangtan, China, 26–27 October 2019; pp. 16–19. [CrossRef]

65. Jiang, B.; Li, X.; Yin, L.; Yue, W.; Wang, S. Object Recognition in Remote Sensing Images Using Combined Deep Features. In Proceedings of the 2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), Chengdu, China, 15–17 March 2019; pp. 606–610. [CrossRef]

66. Kim, H.; Koo, J.; Kim, D.; Park, B.; Jo, Y.; Myung, H.; Lee, D. Vision-Based Real-Time Obstacle Segmentation Algorithm for Autonomous Surface Vehicle. *IEEE Access* **2019**, *7*, 179420–179428. [CrossRef]

67. Sun, J.; Zou, H.; Deng, Z.; Cao, X.; Li, M.; Ma, Q. Multiclass Oriented Ship Localization and Recognition In High Resolution Remote Sensing Images. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 1288–1291. [CrossRef]

68. Ward, C.M.; Harguess, J.; Hilton, C. Ship Classification from Overhead Imagery using Synthetic Data and Domain Adaptation. In Proceedings of the OCEANS 2018 MTS/IEEE Charleston, Charleston, SC, USA, 22–25 October 2018; pp. 1–5. [CrossRef]

69. Zheng, R.; Zhou, Q.; Wang, C. Inland River Ship Auxiliary Collision Avoidance System. In Proceedings of the 2019 18th International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES), Wuhan, China, 8–10 November 2019; pp. 56–59. [CrossRef]

70. Zong-ling, L.; Lu-yuan, W.; Ji-yang, Y.; Bo-wen, C.; Liang, H.; Shuai, J.; Zhen, L.; Jian-feng, Y. Remote Sensing Ship Target Detection and Recognition System Based on Machine Learning. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 1272–1275. [CrossRef]

71. Zou, J.; Yuan, W.; Yu, M. Maritime Target Detection Of Intelligent Ship Based On Faster R-CNN. In Proceedings of the 2019 Chinese Automation Congress (CAC), Hangzhou, China, 22–24 November 2019; pp. 4113–4117. [CrossRef]

72. Huang, Z.; Sun, S.; Li, R. Fast Single-Shot Ship Instance Segmentation Based on Polar Template Mask in Remote Sensing Images. In Proceedings of the IGARSS 2020—2020 IEEE International Geoscience and Remote Sensing Symposium, Waikoloa, HI, USA, 26 September–2 October 2020; pp. 1236–1239. [CrossRef]

73. Chen, Y.; Yang, S.; Suo, Y.; Chen, W. Research on Recognition of Marine Ships under Complex Conditions. In Proceedings of the 2020 Chinese Automation Congress (CAC), Guangzhou, China, 27–30 August 2020; pp. 5748–5753. [CrossRef]

74. Jin, L.; Liu, G. A Convolutional Neural Network for Ship Targets Detection and Recognition in Remote Sensing Images. In Proceedings of the 2020 IEEE 9th Joint International Information Technology and Artificial Intelligence Conference (ITAIC), Chongqing, China, 11–13 December 2020; Volume 9, pp. 139–143. [CrossRef]

75. Kelm, A.P.; Zölzer, U. Walk the Lines: Object Contour Tracing CNN for Contour Completion of Ships. In Proceedings of the 2020 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 10–15 January 2021; pp. 3993–4000. [CrossRef]

76. Li, X.; Cai, K. Method research on ship detection in remote sensing image based on YOLO algorithm. In Proceedings of the 2020 International Conference on Information Science, Parallel and Distributed Systems (ISPDS), Xi'an, China, 14–16 August 2020; pp. 104–108. [CrossRef]

77. Li, J.; Tian, J.; Gao, P.; Li, L. Ship Detection and Fine-Grained Recognition in Large-Format Remote Sensing Images Based on Convolutional Neural Network. In Proceedings of the IGARSS 2020—2020 IEEE International Geoscience and Remote Sensing Symposium, Waikoloa, HI, USA, 26 September–2 October 2020; pp. 2859–2862. [CrossRef]

78. Syah, A.; Wulandari, M.; Gunawan, D. Fishing and Military Ship Recognition using Parameters of Convolutional Neural Network. In Proceedings of the 2020 3rd International Conference on Information and Communications Technology (ICOIACT), Yogyakarta, Indonesia, 24–25 November 2020; pp. 286–290. [CrossRef]

79. Wang, Y.; Wang, L.; Jiang, Y.; Li, T. Detection of Self-Build Data Set Based on YOLOv4 Network. In Proceedings of the 2020 IEEE 3rd International Conference on Information Systems and Computer Aided Education (ICISCAE), Dalian, China, 27–29 September 2020; pp. 640–642. [CrossRef]

80. Yulin, T.; Jin, S.; Bian, G.; Zhang, Y. Shipwreck Target Recognition in Side-Scan Sonar Images by Improved YOLOv3 Model Based on Transfer Learning. *IEEE Access* **2020**, *8*, 173450–173460. [CrossRef]

81. Zhang, X.; Lv, Y.; Yao, L.; Xiong, W.; Fu, C. A New Benchmark and an Attribute-Guided Multilevel Feature Representation Network for Fine-Grained Ship Classification in Optical Remote Sensing Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 1271–1285. [CrossRef]

82. Zhao, D.; Li, X. Ocean ship detection and recognition algorithm based on aerial image. In Proceedings of the 2020 Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC), Dalian, China, 14–16 April 2020; pp. 218–222. [CrossRef]

83. Han, Y.; Yang, X.; Pu, T.; Peng, Z. Fine-Grained Recognition for Oriented Ship Against Complex Scenes in Optical Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–18. [CrossRef]

84. Gong, P.; Zheng, K.; Jiang, Y.; Liu, J. Water Surface Object Detection Based on Neural Style Learning Algorithm. In Proceedings of the 2021 40th Chinese Control Conference (CCC), Shanghai, China, 26–28 July 2021; pp. 8539–8543. [CrossRef]

85. Boyer, A.; Abiemona, R.; Bolic, M.; Petriu, E. Vessel Identification using Convolutional Neural Network-based Hardware Accelerators. In Proceedings of the 2021 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA), Virtual, 18–20 June 2021; pp. 1–6. [CrossRef]

86. Chang, L.; Chen, Y.T.; Hung, M.H.; Wang, J.H.; Chang, Y.L. YOLOv3 Based Ship Detection in Visible and Infrared Images. In Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, 11–16 July 2021; pp. 3549–3552. [CrossRef]

87. Zhou, J.; Jiang, P.; Zou, A.; Chen, X.; Hu, W. Ship Target Detection Algorithm Based on Improved YOLOv5. *J. Mar. Sci. Eng.* **2021**, *9*, 908. [CrossRef]

88. Sali, S.M.; Manisha, N.L.; King, G.; Vidya Mol, K. A Review on Object Detection Algorithms for Ship Detection. In Proceedings of the 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 19–20 March 2021; Volume 1, pp. 1–5. [CrossRef]

89. Su, H.; Zuo, Y.; Li, T. Ship detection in navigation based on broad learning system. In Proceedings of the 2021 International Conference on Security, Pattern Analysis, and Cybernetics (SPAC), Chengdu, China, 18–20 June 2021; pp. 318–322. [CrossRef]

90. Ting, L.; Baijun, Z.; Yongsheng, Z.; Shun, Y. Ship Detection Algorithm based on Improved YOLO V5. In Proceedings of the 2021 6th International Conference on Automation, Control and Robotics Engineering (CACRE), Dalian, China, 15–17 July 2021; pp. 483–487. [CrossRef]

91. Zhang, C.; Xiong, B.; Kuang, G. Ship Detection and Recognition in Optical Remote Sensing Images Based on Scale Enhancement Rotating Cascade R-CNN Networks. In Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, 11–16 July 2021; pp. 3545–3548. [CrossRef]

92. Li, H.; Deng, L.; Yang, C.; Liu, J.; Gu, Z. Enhanced YOLO v3 Tiny Network for Real-Time Ship Detection From Visual Image. *IEEE Access* **2021**, *9*, 16692–16706. [CrossRef]

93. Chen, L.; Shi, W.; Deng, D. Improved YOLOv3 Based on Attention Mechanism for Fast and Accurate Ship Detection in Optical Remote Sensing Images. *Remote Sens.* **2021**, *13*, 660. [CrossRef]

94. Liu, R.W.; Yuan, W.; Chen, X.; Lu, Y. An enhanced CNN-enabled learning method for promoting ship detection in maritime surveillance system. *Ocean. Eng.* **2021**, *235*, 109435. [CrossRef]

95. Hu, J.; Zhi, X.; Shi, T.; Zhang, W.; Cui, Y.; Zhao, S. PAG-YOLO: A Portable Attention-Guided YOLO Network for Small Ship Detection. *Remote Sens.* **2021**, *13*, 3059. [CrossRef]

96. Leonidas, L.A.; Jie, Y. Ship Classification Based on Improved Convolutional Neural Network Architecture for Intelligent Transport Systems. *Information* **2021**, *12*, 302. [CrossRef]

97. Dong, Y.; Chen, F.; Han, S.; Liu, H. Ship Object Detection of Remote Sensing Image Based on Visual Attention. *Remote Sens.* **2021**, *13*, 3192. [CrossRef]

98. Su, N.; Huang, Z.; Yan, Y.; Zhao, C.; Zhou, S. Detect Larger at Once: Large-Area Remote-Sensing Image Arbitrary-Oriented Ship Detection. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 1–5. [CrossRef]

99. Xie, P.; Tao, R.; Luo, X.; Shi, Y. YOLOv4-MobileNetV2-DW-LCARM: A Real-Time Ship Detection Network. In Proceedings of the Knowledge Management in Organisations, Hagen Germany, 11–14 July 2022; Uden, L., Ting, I.H., Feldmann, B., Eds.; Springer International Publishing: Cham, Switzerland, 2022; pp. 281–293.

100. Li, L.; Jiang, L.; Zhang, J.; Wang, S.; Chen, F. A Complete YOLO-Based Ship Detection Method for Thermal Infrared Remote Sensing Images under Complex Backgrounds. *Remote Sens.* **2022**, *14*, 1534. [CrossRef]

101. Gao, H.; Cheng, B.; Wang, J.; Li, K.; Zhao, J.; Li, D. Object Classification Using CNN-Based Fusion of Vision and LIDAR in Autonomous Vehicle Environment. *IEEE Trans. Ind. Inform.* **2018**, *14*, 4224–4231. [CrossRef]

102. Yan, Z.; Zhang, H.; Piramuthu, R.; Jagadeesh, V.; DeCoste, D.; Di, W.; Yu, Y. HD-CNN: Hierarchical Deep Convolutional Neural Networks for Large Scale Visual Recognition. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015.

103. Gundogdu, E.; Solmaz, B.; Yücesoy, V.; Koç, A. MARVEL: A Large-Scale Image Dataset for Maritime Vessels. In *Computer Vision – ACCV 2016*; Springer International Publishing: Berlin/Heidelberg, Germany, 2017; pp. 165–180. [CrossRef]

104. Lin, T.Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft COCO: Common Objects in Context. In Proceedings of the Computer Vision—ECCV 2014, Zurich, Switzerland, 6–12 September 2014; pp. 740–755.

105. Doon, R.; Kumar Rawat, T.; Gautam, S. Cifar-10 Classification using Deep Convolutional Neural Network. In Proceedings of the 2018 IEEE Punecon, Pune, India, 30 November–2 December 2018; pp. 1–5. [CrossRef]

106. Everingham, M.; Van Gool, L.; Williams, C.K.I.; Winn, J.; Zisserman, A. The Pascal Visual Object Classes (VOC) Challenge. *Int. J. Comput. Vis.* **2010**, *88*, 303–338. [CrossRef]

107. Kuznetsova, A.; Rom, H.; Alldrin, N.; Uijlings, J.; Krasin, I.; Pont-Tuset, J.; Kamali, S.; Popov, S.; Malloci, M.; Kolesnikov, A.; et al. The Open Images Dataset V4: Unified image classification, object detection, and visual relationship detection at scale. *Int. J. Comput. Vis. (IJCV)* **2020**, *128*, 1956–1981. [CrossRef]

108. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Fei-Fei, L. ImageNet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255. [CrossRef]

109. Iancu, B.; Soloviev, V.; Zelioli, L.; Lilius, J. ABOships—An Inshore and Offshore Maritime Vessel Detection Dataset with Precise Annotations. *Remote Sens.* **2021**, *13*, 988. [CrossRef]

110. Zhang, M.M.; Choi, J.; Daniilidis, K.; Wolf, M.T.; Kanan, C. VAIS: A dataset for recognizing maritime imagery in the visible and infrared spectrums. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Boston, MA, USA, 7–12 June 2015; pp. 10–16. [CrossRef]

111. Zheng, Y.; Zhang, S. Mcships: A Large-Scale Ship Dataset For Detection And Fine-Grained Categorization In The Wild. In Proceedings of the 2020 IEEE International Conference on Multimedia and Expo (ICME), London, UK, 6–10 July 2020; pp. 1–6. [CrossRef]

112. Shao, Z.; Wu, W.; Wang, Z.; Du, W.; Li, C. SeaShips: A Large-Scale Precisely Annotated Dataset for Ship Detection. *IEEE Trans. Multimed.* **2018**, *20*, 2593–2604. [CrossRef]

113. Kaggle. High Resolution Ship Collections 2016 (HRSC2016). 2016. Available online: https://www.kaggle.com/datasets/guofeng/hrsc2016 (accessed on 22 December 2021).

114. Kaggle. Airbus Ship Detection Challenge. 2018. Available online: https://www.kaggle.com/c/airbus-ship-detection (accessed on 22 December 2021).

115. Rainey, K.; Parameswaran, S.; Harguess, J.; Stastny, J. Vessel classification in overhead satellite imagery using learned dictionaries. In Proceedings of the Applications of Digital Image Processing XXXV, San Diego, CA, USA, 13–16 August 2012. [CrossRef]

116. Zhang, Z.; Zhang, L.; Wang, Y.; Feng, P.; He, R. ShipRSImageNet: A Large-Scale Fine-Grained Dataset for Ship Detection in High-Resolution Optical Remote Sensing Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 8458–8472. [CrossRef]

117. Zou, W.; Lu, Y.; Chen, M.; Lv, F. Rapid Face Detection in Static Video Using Background Subtraction. In Proceedings of the 2014 Tenth International Conference on Computational Intelligence and Security, Kunming, China, 15–16 November 2014; pp. 252–255. [CrossRef]

118. Hosang, J.; Benenson, R.; Schiele, B. Learning Non-maximum Suppression. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 6469–6477. [CrossRef]

119. You, J.; Hu, Z.; Peng, C.; Wang, Z. Generation and Annotation of Simulation-Real Ship Images for Convolutional Neural Networks Training and Testing. *Appl. Sci.* **2021**, *11*, 5931. [CrossRef]

120. NaturalIntelligence. ImgLab. 2020. Available online: https://github.com/NaturalIntelligence/imglab (accessed on 22 December 2021).

121. Microsoft. Visual Object Tagging Tool (VoTT). 2019. Available online: https://github.com/microsoft/VoTT (accessed on 22 December 2021).

122. Sekachev, B. Computer Vision Annotation Tool (CVAT). 2020. Available online: https://github.com/openvinotoolkit/cvat (accessed on 22 December 2021).

123. Lin, T. LabelImg. 2015. Available online: https://github.com/tzutalin/labelImg (accessed on 22 December 2021).

124. Wada, K. Labelme. 2021. Available online: https://github.com/wkentaro/labelme (accessed on 23 December 2021).

125. Dutta, A.; Zisserman, A. The VIA Annotation Software for Images, Audio and Video. In Proceedings of the 27th ACM International Conference on Multimedia, Nice, France, 21–25 October 2019. [CrossRef]

126. SuperAnnotate. 2018. Available online: https://www.superannotate.com/ (accessed on 23 December 2021).

127. Supervisely. 2021. Available online: https://github.com/supervisely/supervisely (accessed on 23 December 2021).

128. Skalski, P. Make-Sense. 2019. Available online: https://github.com/SkalskiP/make-sense/ (accessed on 23 December 2021).

129. LabelBox. 2021. Available online: https://labelbox.com/ (accessed on 27 December 2021).

130. DarkLabel. 2020. Available online: https://github.com/darkpgmr/DarkLabel (accessed on 27 December 2021).

131. Arunachalam, A.; Ravi, V.; Acharya, V.; Pham, T.D. Toward Data-Model-Agnostic Autonomous Machine-Generated Data Labeling and Annotation Platform: COVID-19 Autoannotation Use Case. *IEEE Trans. Eng. Manag.* 2021, 1–12. [CrossRef]

132. Li, H.; Wang, X. Automatic Recognition of Ship Types from Infrared Images Using Support Vector Machines. In Proceedings of the 2008 International Conference on Computer Science and Software Engineering, Wuhan, China, 12–14 December 2008; Volume 6, pp. 483–486. [CrossRef]

133. Zabidi, M.M.A.; Mustapa, J.; Mokji, M.M.; Marsono, M.N.; Sha'ameri, A.Z. Embedded vision systems for ship recognition. In Proceedings of the TENCON 2009—2009 IEEE Region 10 Conference, Singapore, 23–26 November 2009; pp. 1–5. [CrossRef]

134. Kao, C.H.; Hsieh, S.P.; Peng, C.C. Study of feature-based image capturing and recognition algorithm. In Proceedings of the ICCAS 2010, Gyeonggi-do, Korea, 27–30 October 2010; pp. 1855–1861. [CrossRef]

135. Ganbold, U.; Akashi, T. The real-time reliable detection of the horizon line on high-resolution maritime images for unmanned surface-vehicle. In Proceedings of the 2020 International Conference on Cyberworlds (CW), Caen, France, 29 September–1 October 2020; pp. 204–210.

136. Chen, J.; Wang, J.; Lu, H. Ship Detection in Complex Weather Based on CNN. In Proceedings of the 2021 6th International Conference on Intelligent Computing and Signal Processing (ICSP), Xi'an, China, 9–11 April 2021; pp. 1225–1228.

137. Mu, X.; Lin, Y.; Liu, J.; Cao, Y.; Liu, H. Surface Navigation Target Detection and Recognition based on SSD. In Proceedings of the 2019 3rd International Conference on Electronic Information Technology and Computer Engineering (EITCE), Xiamen, China, 18–20 October 2019; pp. 649–653. [CrossRef]

138. Prayudi, A.; Sulistijono, I.A.; Risnumawan, A.; Darojah, Z. Surveillance System for Illegal Fishing Prevention on UAV Imagery Using Computer Vision. In Proceedings of the 2020 International Electronics Symposium (IES), Surabaya, Indonesia, 29-30 September 2020; pp. 385–391.

139. Patel, K.; Bhatt, C.; Mazzeo, P.L. Deep Learning-Based Automatic Detection of Ships: An Experimental Study Using Satellite Images. *J. Imaging* **2022**, *8*, 182. [CrossRef] [PubMed]

140. Dan, Z.; Sang, N.; Wang, R.; Chen, Y.; Chen, X. A Transductive Transfer Learning Method for Ship Target Recognition. In Proceedings of the 2013 Seventh International Conference on Image and Graphics, Qingdao, China, 26–28 July 2013; pp. 418–422. [CrossRef]

141. Zhu, C.; Zhou, H.; Wang, R.; Guo, J. A Novel Hierarchical Method of Ship Detection from Spaceborne Optical Image Based on Shape and Texture Features. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 3446–3456. [CrossRef]

142. Zerrouk, I.; Moumen, Y.; Khiati, W.; Berrich, J.; Bouchentouf, T. Detection Process of Ships in Aerial Imagery Using Two Convnets. In Proceedings of the 2019 International Conference on Wireless Technologies, Embedded and Intelligent Systems (WITS), Fez, Morocco, 3–4 April 2019; pp. 1–8. [CrossRef]

143. Kitayama, T.; Lu, H.; Li, Y.; Kim, H. Detection of Grasping Position from Video Images Based on SSD. In Proceedings of the 2018 18th International Conference on Control, Automation and Systems (ICCAS), PyeongChang, Korea, 17–20 October, 2018; pp. 1472–1475.

144. Wang, X.; Zhang, T. Clutter-adaptive infrared small target detection in infrared maritime scenarios. *Opt. Eng.* **2011**, *50*, 1–13. [CrossRef]

145. Bian, W.; Zhu, Q. The ship target detection based on panoramic images. In Proceedings of the 2015 IEEE International Conference on Mechatronics and Automation (ICMA), Beijing, China, 2–5 August 2015; pp. 2397–2401. [CrossRef]

146. Liu, W.; Yang, X.; Zhang, J. A Robust Target Detection Algorithm Using MEMS Inertial Sensors for Shipboard Video System. In Proceedings of the 2020 27th Saint Petersburg International Conference on Integrated Navigation Systems (ICINS), Saint Petersburg, Russia, 25–27 May 2020; pp. 1–5. [CrossRef]

147. Zhong, Z.; Li, Y.; Han, Z.; Yang, Z. Ship Target Detection Based on LightGBM Algorithm. In Proceedings of the 2020 International Conference on Computer Information and Big Data Applications (CIBDA), Guiyang, China, 17–19 April 2020; pp. 425–429. [CrossRef]

148. Szpak, Z.L.; Tapamo, J.R. Maritime surveillance: Tracking ships inside a dynamic background using a fast level-set. *Expert Syst. Appl.* **2011**, *38*, 6669–6680. [CrossRef]

149. Chang, J.Y.; Oh, H.; Lee, S.J.; Lee, K.J. Ship Detection for KOMPSAT-3A Optical Images Using Binary Features and Adaboost Classification. In Proceedings of the IGARSS 2020—2020 IEEE International Geoscience and Remote Sensing Symposium, Waikoloa, HI, USA, 26 September–2 October 2020; pp. 968–971. [CrossRef]

150. Sankaraiah, Y.R.; Varadarajan, S. Deblurring techniques—A comprehensive survey. In Proceedings of the 2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI), Chennai, India, 21–22 Sepember 2017; pp. 2032–2035. [CrossRef]

151. Zheng, H. A Survey on Single Image Deblurring. In Proceedings of the 2021 2nd International Conference on Computing and Data Science (CDS), Stanford CA USA, 28–30 January 2021; pp. 448–452. [CrossRef]

152. Mahalakshmi, A.; Shanthini, B. A survey on image deblurring. In Proceedings of the 2016 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 7–9 January 2016; pp. 1–5. [CrossRef]

153. van Valkenburg-van Haarst, T.Y.C.; Scholte, K.A. Polynomial background estimation using visible light video streams for robust automatic detection in a maritime environment. In Proceedings of the Electro-Optical Remote Sensing, Photonic Technologies, and Applications III, Berlin, Germany, 1-3 September 2009; Bishop, G.J., Gonglewski, J.D., Lewis, K.L., Hollins, R.C., Merlet, T.J., Kamerman, G.W., Steinvall, O.K., Eds.; International Society for Optics and Photonics, SPIE Digital Library: Bellingham, WA, USA, 2009; Volume 7482, pp. 94–101. [CrossRef]

154. Pan, M.; Liu, Y.; Cao, J.; Li, Y.; Li, C.; Chen, C.H. Visual Recognition Based on Deep Learning for Navigation Mark Classification. *IEEE Access* **2020**, *8*, 32767–32775. [CrossRef]

155. Li, K.d.; Zhang, Y.y.; Li, Y.j. Researches of Sea Surface Ship Target Auto-recognition Based on Wavelet Transform. In Proceedings of the 2010 International Conference on System Science, Engineering Design and Manufacturing Informatization, Yichang, China, 12–14 November 2010; Volume 1, pp. 193–195. [CrossRef]

156. Shao, Z.; Wang, L.; Wang, Z.; Du, W.; Wu, W. Saliency-Aware Convolution Neural Network for Ship Detection in Surveillance Video. *IEEE Trans. Circuits Syst. Video Technol.* **2020**, *30*, 781–794. [CrossRef]

157. Li, Z.; Yang, D.; Chen, Z. Multi-layer Sparse Coding Based Ship Detection for Remote Sensing Images. In Proceedings of the 2015 IEEE International Conference on Information Reuse and Integration, Francisco, CA, USA, 13–15 August 2015; pp. 122–125. [CrossRef]

158. van den Broek, S.P.; Schwering, P.B.W.; Liem, K.D.; Schleijpen, R. Persistent maritime surveillance using multi-sensor feature association and classification. In Proceedings of the Signal Processing, Sensor Fusion, and Target Recognition XXI, Baltimore, MD, USA, 17 April 2012; Kadar, I., Ed.; International Society for Optics and Photonics, SPIE Digital Library: Bellingham, WA, USA, 2012; Volume 8392, pp. 341–351. [CrossRef]

159. Son, N.S. On an Autonomous Navigation System for Collision Avoidance of Unmanned Surface Vehicle. In Proceedings of the ION 2013 Pacific PNT Meeting, Honolulu, HI, USA, 22–25 April 2013; pp. 470–476.

160. Suzuki, S.; Mitsukura, Y.; Furuya, T. Ship detection based on spatio-temporal features. In Proceedings of the 2014 10th France-Japan/ 8th Europe-Asia Congress on Mecatronics (MECATRONICS2014- Tokyo), Tokyo, Japan, 27–29 November 2014; pp. 93–98. [CrossRef]

161. Patino, L.; Ferryman, J. Loitering Behaviour Detection of Boats at Sea. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA, 21–26 July 2017; pp. 2169–2175. [CrossRef]

162. Kaido, N.; Yamamoto, S.; Hashimoto, T. Examination of automatic detection and tracking of ships on camera image in marine environment. In Proceedings of the 2016 Techno-Ocean (Techno-Ocean), Kobe, Japan, 6–8 October 2016; pp. 58–63. [CrossRef]

163. Wei, H.; Nguyen, H.; Ramu, P.; Raju, C.; Liu, X.; Yadegar, J. Automated intelligent video surveillance system for ships. In Proceedings of the Optics and Photonics in Global Homeland Security V and Biometric Technology for Human Identification VI, Orlando, FL, USA, 13–16 April 2009; Volume 7306, pp. 58–63. [CrossRef]

164. Chen, X.; Ling, J.; Wang, S.; Yang, Y.; Luo, L.; Yan, Y. Ship detection from coastal surveillance videos via an ensemble Canny-Gaussian-morphology framework. *J. Navig.* **2021**, *74*, 1252–1266. [CrossRef]

165. Teixeira, E.H.; Mafra, S.B.; Rodrigues, J.J.P.C.; Da Silveira, W.A.A.N.; Diallo, O. A Review and Construction of a Real-time Facial Recognition System. In Proceedings of the Anais do Simpósio Brasileiro de Computação Ubíqua e Pervasiva (SBCUP), Cuiabá, Mato Grosso, Brazil, 16–20 November 2020; pp. 191–200. [CrossRef]

166. Lee, J.M.; Lee, K.H.; Nam, B.; Wu, Y. Study on Image-Based Ship Detection for AR Navigation. In Proceedings of the 2016 6th International Conference on IT Convergence and Security (ICITCS), Prague, Czech Republic, 26–29 September 2016; pp. 1–4. [CrossRef]

167. Huang, G.B.; Ramesh, M.; Berg, T.; Learned-Miller, E. *Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments*; *Technical Report 07-49*; University of Massachusetts: Amherst, MA, USA, 2007.