Knowledge-Transfer-Based Bidirectional Vessel Monitoring System for Remote and Nearshore Images

Jiawen Li 1,2,3, Yun Yang 1,4, Xin Li 1, Jiahua Sun 1 and Ronghui Li 1,2,3,*

1 Naval Architecture and Shipping College, Guangdong Ocean University, Zhanjiang 524005, China; jiawen-li@gdou.edu.cn (J.L.); yangyun5@stu.gdou.edu.cn (Y.Y.)
2 Technical Research Center for Ship Intelligence and Safety Engineering of Guangdong Province, Zhanjiang 524005, China
3 Guangdong Provincial Key Laboratory of Intelligent Equipment for South China Sea Marine Ranching, Zhanjiang 524005, China
4 College of Civil and Transportation Engineering, Shenzhen University, Shenzhen 518060, China
* Correspondence: lironghui@163.com

Abstract: Vessel monitoring technology involves the application of remote sensing technologies to detect and identify vessels in various environments, which is critical for monitoring vessel traffic, identifying potential threats, and facilitating maritime safety and security to achieve real-time maritime awareness in military and civilian domains. However, most existing vessel monitoring models tend to focus on a single remote sensing information source, leading to limited detection functionality and underutilization of available information. In light of these limitations, this paper proposes a comprehensive ship monitoring system that integrates remote satellite devices and nearshore detection equipment. The system employs ResNet, a deep learning model, along with data augmentation and transfer learning techniques to enable bidirectional detection of satellite cloud images and nearshore outboard profile images, thereby alleviating prevailing issues such as low detection accuracy, homogeneous functionality, and poor image recognition applicability. Empirical findings based on two real-world vessel monitoring datasets demonstrate that the proposed system consistently performs best in both nearshore identification and remote detection. Additionally, extensive supplementary experiments were conducted to evaluate the effectiveness of different modules and discuss the constraints of current deep learning-based vessel monitoring models.

Keywords: ship classification; synthetic-aperture radar (SAR); deep learning; transfer learning; vessel monitoring system

1. Introduction

The activities of vessels immediately affect all aspects of the maritime domain, including security, economy, and environment [1,2]. The vessel surveillance system has tremendous value for maritime management. A real-time and accurate vessel monitoring system contributes to safeguarding vessel security and upgrading the efficiency of vessel turnover in port management. Furthermore, vessel monitoring systems with accurate identification performance can greatly aid in salvage at sea, marine traffic monitoring, fishery management, and environmental protection [3]. Thus, how to construct an efficient and practical vessel monitoring system has attracted the attention of many researchers.

With increasing maritime activities and frequent vessel movements, the large-scale information on vessel activities poses an urgent challenge for the maritime surveillance system [4–7]. Current existing traditional vessel monitoring systems have security risks and cannot meet the demands of the maritime industry [8–10]. Traditional vessel monitoring systems can only identify vessels equipped with radio communication equipment such as AIS (automatic identification system), GPS (global positioning system), and VHF (very high frequency) that are considered intelligent, while small vessels or other non-intelligent
vessels, such as fishing boats or sailing boats that lack such radio communication equipment, are imperceptible. In addition, current monitoring systems are also unable to detect vessels with weak signals or spotty satellite reception and interference as well as those with intentionally or unintentionally disabled GPS, AIS, or VHF systems.

Unmonitored vessels pose a significant risk to maritime security and can facilitate illicit activities that threaten the economic stability and political security of a nation. In November 2023, Chinese authorities apprehended a group of criminals who deliberately disabled their vessel’s automatic identification system (AIS) and smuggled 768 tons of frozen products illegally. The limitations of current monitoring systems, which rely heavily on electronic signaling devices, provide an easy opportunity for rogue actors to evade detection and engage in lawless activities. These vulnerabilities underscore the urgent need to develop more effective monitoring systems that are capable of detecting vessels that are intentionally or unintentionally operating without electronic signaling devices.

To address the issue of radio dependence in the current vessel monitoring system, researchers have sought to optimize it using artificial intelligence (AI) methods. Specifically, with the emergence of deep learning and the ability of convolutional neural networks to autonomously learn structured features, scholars have turned their attention to synthetic-aperture radar (SAR) vessel image classification [11,12]. However, most of these methods are focused on remote detection from SAR images, and few have been developed for nearshore detection. This is because existing algorithms are often hindered by onshore buildings, making it difficult to distinguish the background from the vessel and causing inconvenience to vessel monitoring systems. To address this issue, some scholars have attempted to conduct research on nearshore vessel monitoring [13,14]. Despite the widespread use of SAR image classification, research on combined remote and nearshore maritime monitoring remains scarce.

The significance of vessel image monitoring systems in ensuring maritime safety cannot be understated. These systems enable rapid counting of vessels in the monitored waters and effective differentiation between various types of vessels, thereby providing a comprehensive view of vessel activity at sea. However, the full potential of these advanced technologies has not been harnessed by the maritime industry. The current monitoring systems can only perform a single task of either vessel detection or classification, necessitating multiple systems to work in conjunction for simultaneous multiple tasks. This approach results in inefficient and impractical vessel image monitoring tasks. Consequently, the existing single-function and poorly practical vessel monitoring and classification systems are not suitable for real-world applications. There is a need to develop efficient and practical vessel image monitoring systems capable of performing multiple tasks simultaneously to enhance maritime safety.

Hence, in order to address the shortcomings of existing vessel monitoring systems in terms of low accuracy, too homogeneous functionality and poor applicability in image recognition, we propose a bi-directional vessel monitoring system based on knowledge transfer for remote and nearshore to achieve high accuracy and realistic scenarios that can be applied to maritime monitoring, as shown in Figure 1. This innovative system integrates remote satellite equipment and nearshore detection equipment, resulting in a comprehensive vessel surveillance system. The KB-VMS system facilitates knowledge transfer between the remote satellite equipment and nearshore detection equipment, enabling bidirectional data flow. This data flow enhances the monitoring capabilities of the system, resulting in enhanced surveillance of vessels. The KB-VMS system has the potential to improve maritime safety and security by enabling the real-time monitoring of vessel traffic in a comprehensive manner. When vessels entering within KB-VMS-monitored waters, the satellite detection cloud map acquires remote images and uploads them to the remote detection module for feature fusion, and the fused features are predicted and filtered through the ResNet network to output the number of vessels within those waters. When a vessel approaches, the nearshore camera captures the image of a single vessel in real time and the nearshore monitoring module obtains that image information, again using
ResNet to extract the image features, fuses the features of different depths using the feature mechanism, and makes predictions on the fused target image to output the type of that vessel. Finally, in terms of data processing, this article incorporates training strategies such as data enhancement and knowledge transfer, thus improving the accuracy of detection and classification.

Figure 1. The concept graphic of knowledge-transfer-based bidirectional vessel monitoring system (KB-VMS).

The experimental results demonstrate that the KB-VMS outperforms both the baseline monitoring systems and the state-of-the-art monitoring systems. With the ability to enable real-time joint remote and nearshore monitoring without the need for smart devices, our system can adapt to monitoring scenarios at different distances using a single algorithm, making it highly suitable for practical applications.

In summary, the main contributions of this article are the following:

• To the best of our knowledge, our study is the first to empirically integrate both remote satellite equipment and nearshore detection equipment into vessel surveillance systems. Our model is capable of identifying vessels even in situations where GPS, AIS, and other equipment are not being used, providing a higher level of sea safety and enhancing defense capabilities.

• Experiments based on two real-world vessel monitoring datasets (nearshore dataset and remote dataset) show that our model achieves the highest accuracy and outperforms the baselines and state-of-the-art models, achieving 97.18% and 94.43% accuracy, respectively. Compared to the original deep learning model, our model has shown significant improvement, with a nearshore detection accuracy increase of 18% and a remote detection accuracy increase of 38%.

• The experimental findings reported herein establish the efficacy of transfer learning and data augmentation in the realm of vessel detection and recognition. In addition, the investigations presented herein have uncovered several flawed tendencies in extant deep learning models for vessel monitoring, thereby advancing the underpinnings of future research endeavors aimed at deep-learning-based vessel detection.

The remainder of this article is organized as follows. Section 2 reviews the related work. Section 3 describes our integrated remote and nearshore monitoring system in detail. Section 4 shows the experimental situation and analysis results. Finally, Section 5 provides a summary and an outlook for future work.

2. Related Work

This section provides a review of related works, which can be classified into two main types: (1) traditional marine monitoring methods and (2) modern CNN-based vessel monitoring methods.
2.1. Traditional Vessel Monitoring Methods

The most prominent feature of traditional vessel monitoring systems is their reliance on vesselboard radio communication equipment, which often only monitors vessels that are normally enabled with these devices. Most of the previous paper studied how to use GPS, AIS, and VHF equipment for maritime monitoring, generally optimizing on AIS data and SAR. In earlier years, T. Eriksen et al. [15] used space-based AIS receivers for vessel detection, and space-based AIS receivers have a communication range of more than 1000 nautical miles in low-Earth orbit, creating a good opportunity for wide-range maritime monitoring. While Hannevik et al. [16] found that using only space-based AIS could only monitor specific vessel types and not all vessels, they proposed to combine SAR imagery with space-based AIS, where AIS can monitor remote areas and identify vessels detected in SAR imagery, which in turn can be used to detect vessels that do not send mandatory AIS reports, i.e., SAR and Chaturvedi et al. [17] conducted research in maritime security, supported by AIS data, to further identify “friend” and “foe” vessel targets in the TerraSAR-X images. The early warning system is useful for discerning vessel information in a large ocean area as “enemy” vessels can be identified in advance to determine if the vessel poses a threat to the area. In addition, radar also plays an important role in early maritime surveillance. When comparing AIS and HF radar, John F et al. [18] found that HF radar can miss detection and AIS can only identify some of the vessels, so they combined AIS and HF radar with a Bayesian network so they could prioritize the presence of vessels in the area, reduce environmental interference, and improve the area target detection efficiency. Hong et al. [19] take advantage of the low power consumption and all-weather operation of FMCW radar to monitor the vessel activity information in real time. Compared with AIS, some small vessels not equipped with AIS can also be detected by FMCW radar. This method expands the range of maritime monitoring and helps to monitor the vast ocean in real time and improve maritime safety. Generally speaking, most of the images of small vessels are low-resolution images, which cause great inconvenience to the maritime monitoring. Stephan et al. [20] adopt TerraSAR-X, which can provide high resolution over a wide range, in order to overcome the limitations of AIS systems for vessel detection, to improve the resolution of images and further reduce the false alarm of small vessel detection. In addition, TerraSAR-X works with SatAIS to improve the accuracy of vessel detection in bad weather, providing a better mode for monitoring vessels at sea and bridging the monitoring gap of the coverage of ground-based AIS in the sea beyond 40 km from the coastline.

To summarize, although the above methods achieve satisfactory results in some cases, these systems are very dependent on the intelligent communication equipment on board, and the types of vessels that can be monitored are very limited. When the radio communication equipment on board does not work properly, it is difficult for the maritime authorities to obtain timely information about the activities of the vessel at sea, which brings inconvenience to the maritime monitoring.

2.2. Deep-Learning-Based Vessel Monitoring Methods

Deep learning has powerful learning ability and efficient feature representation. Modern CNN-based vessel monitoring methods do not require signals transmitted back from radio communication devices for vessel monitoring. In recent years, the popularity of deep learning in image recognition has led to the widespread use of convolutional neural networks for vessel classification and target detection. CNN-based classification models have almost dominated the field of deep learning image classification, and their accuracy rates surpass those of traditional methods.

In addition, the unique penetrating power of SAR sensors, which can work around the clock, makes vessel detection under SAR images play a crucial role in ocean monitoring. Therefore, scholars have started to focus on SAR vessel classification and designed several CNN-based SAR vessel classifiers.
In 2018, Jiao et al. [13] proposed a densely connected multi-scale neural network (DCMSNN) based on the top-down dense connection of a feature map with other feature maps to address multi-scale and multi-scene SAR vessel detection under complex background interference near shore. Finally, their experimental results show that different levels of feature maps can adapt to vessels of different sizes and scales, resulting in higher detection accuracy. Zhang et al. [21] obtain support from the OpenSARvessel database to improve vessel recognition through migration learning. Thus, inspired by their work, knowledge migration will be fused into our deep network ResNet to form a new KT-ResNet for joint classification for vessel classification. They used CNN and MLP to extract image features during classification. In their experiments, they conducted training tests on RADARSAT-2 and Sentinel-1 datasets. Their models have better performance than the current models.

In 2019, Chen et al. [14] proposed an object detection network incorporating an attention mechanism to address the problem that vessel recognition is susceptible to complex background interference, enabling them to focus on target vessels in different scenarios. In addition, loss functions were constructed to reduce the sensitivity of vessel scales in order to address the detection accuracy reduction caused by different scales of vessels. Some joint optimization work on generator and classifier design was investigated by Wu et al. [23]. They proposed a joint CNN framework to improve the resolution of images and optimize the image quality to distinguish different types of vessels in high-resolution SAR images.

In the vast ocean, the monitoring of small-sized vessels is particularly difficult, and how to efficiently monitor small-sized vessels has become a research challenge. In 2020, Zhao et al. [24] conducted a study on solving small-scale vessel detection and proposed DDNet, which combines stacked convolutional layers and dense connectivity to efficiently solve the monitoring problem of more small-sized vessels and finally obtain the target of more accurate detection results. A densely connected triple CNN was proposed by He et al. [25]. They used a DML scheme to deepen the depth between the same classes and expand the distance between different classes. In their report, they conducted sufficient experiments on the MRSAR vessel dataset, and their model has superior performance in the three- and five-class vessel identification datasets compared to the original CNN. Inspired by the fact that the human vision system can quickly focus on the area of interest in the image, Xu et al. [26] designed a cascaded CNN, which is divided into a front-end shallow CNN and a back-end deep CNN, which can be used to quickly exclude non-vessel areas and identify areas with vessels so as to improve the accuracy and efficiency of vessel detection.

In 2021, Tang et al. [27] designed a YOLO (You Only Look Once)-based N-YOLO based on the vulnerability of vessels to different levels of noise using a noise-level classifier to preclassify SAR images according to the noise level and then using CA-CFAR to extract target regions; the above two steps aim to reduce the interference of noise and nearshore buildings on the images. They found that the N-YOLO model is more competitive than the traditional CNN-based target detection methods. Zeng et al. [28] investigated a vessel grain classification for dual-polarization SAR, and their model was able to effectively classify vessels into eight accurate classes, such as cargo, tanker, carrier, container, fishing dredger, tug, passenger, etc., using VV and VH dual-polarization channels to enhance the classification performance on the OpenSARvessel dataset. The application of the dual-polarization idea also provides an important support for later research.

In 2022, He et al. [29] found that single-polarization SAR vessel classification was limited in practical applications, so they actively explored SAR vessel classification related to dual polarization and designed a GBCNN model to extract vessel targets by combining vertical/horizontal polarization and vertical/vertical schemes and constructed a multi-polarization fusion loss function (MFPL) to train the model using dual-polarization information. In their work, they improved the classification accuracy of three and five types of dual-polarized vessels on the OpenSAR vessel dataset. Huang et al. [30] found that in addition to having different types of vessels, there are also vessels with similar
hull structures but different superstructures and equipment. Therefore, they designed a CNN-Swin model for the classification of military vessels. In their experiments, their model has great potential in vessel classification. Polap et al. [31] provided an artificial intelligence technique with image classifier to perform automatic ship classification for a riverbank monitoring system using cascading and a reward and punishment mechanism. The cascading approach with multiple classifiers and the reward and punishment mechanism is undoubtedly a fabulous idea and has been shown to be effective in ship classification. However, we proposed an integrated ship monitoring system that integrates remote satellite equipment and nearshore detection equipment, using the deep learning model ResNet, combined with data augmentation and migration learning techniques, to achieve bi-directional detection of satellite cloud images and offshore outboard profiles. Our approach utilizes more information sources, making it more practical and having stronger usability in real-world scenarios.

2.3. Comparison of the Existing Models and Our Approach

Most existing vessel monitoring methods focus only on single nearshore or remote detection, lacking the ability to provide unified monitoring of far and inshore maritime vessel activity information. Such a single-function system may not be practical for real-world maritime surveillance. Therefore, our proposed KB-VMS is an integrated remote and inshore vessel monitoring system that incorporates knowledge transfer and is capable of rapid learning and application. It aims to provide better all-around remote and nearshore monitoring.

Table 1 provides a comparative analysis of traditional and modern deep learning methods across various parameters, including remote detection, inshore identification, stability, visibility, security, practicability, multi-functionality, and precision. Notably, nearly all models are capable of presenting collected data in image form, thereby facilitating visualization. In terms of stability and security, while the automatic identification system (AIS) is limited in its ability to identify vessels not using AIS, radar and convolutional neural network (CNN) models exhibit greater capacity for detecting such vessels and ensuring maritime safety. Our proposed system represents a pioneering effort in integrating remote and nearshore monitoring, thereby enabling adaptability to varying distance-based monitoring scenarios. In contrast to previous approaches, our system is uniquely equipped to address realistic scenarios by leveraging remote satellite and nearshore surveillance stations to achieve comprehensive monitoring capabilities. Our approach emphasizes bidirectional information processing, thereby contributing to stronger identification capabilities.

Table 1. Comparison of recent related studies.

<table>
<thead>
<tr>
<th>Model</th>
<th>Remote Detection</th>
<th>Inshore Identification</th>
<th>Stability</th>
<th>Visibility</th>
<th>Security</th>
<th>Practicability</th>
<th>Multi Function</th>
<th>High Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space-AIS [15]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ground-AIS [20]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HF [18]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FMCW [19]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CNN-MLP [22]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cascade CNN [26]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DCMSNN [13]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DDNET [24]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GBCNN [29]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KB-VMS</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
</tbody>
</table>

The majority of current vessel monitoring approaches are confined to singular nearshore or remote detection capabilities, limiting their potential to provide comprehensive coverage of both offshore and inshore maritime vessel activity. This unifunctional approach may lack the practicality required for effective real-world maritime surveillance. In response, we
propose an integrated remote and inshore vessel monitoring system, the KB-VMS, which incorporates knowledge transfer mechanisms and is characterized by its rapid learning and application capabilities. The KB-VMS represents a novel solution to the challenge of providing comprehensive remote and nearshore monitoring.

3. KB-VMS System

This paper introduces a novel bi-directional vessel monitoring system, namely the knowledge-based vessel monitoring system (KB-VMS), which integrates knowledge transfer to support both remote and nearshore vessel detection. The system comprises two key modules: the remote satellite monitoring module and the nearshore monitoring module. Firstly, the raw images captured by satellite or inshore cameras undergo preprocessing and are then fed into their respective modules. Subsequently, the image data undergo automated processing and extraction of high-level features. Finally, the remote and nearshore monitoring modules analyze the monitoring information and output the number or type of vessels. The proposed KB-VMS system is capable of performing joint remote and nearshore monitoring with high accuracy, resulting in a practical and reliable maritime monitoring solution.

3.1. Problem Statement

We defined this vessel monitoring task as a hybrid task that can be divided into the remote satellite monitoring sub-task and nearshore camera monitoring sub-task.

The remote satellite monitoring sub-task was formulated as follows: given a synthetic-aperture radar (SAR) image, the model judged if there are any vessels in the maritime space of that SAR image, if any, given the number of the vessels. We defined a SAR training dataset as $I_s^i, \forall i \in \{1, \cdots, \delta\}$, where $\delta$ is the total number of image in the SAR dataset, and $I_s^i$ is the given SAR image. The output of this sub-task was formulated as $f^s : I_s^i \to \{y_s\}$, where $y_s$ shows the number of vessels of the $I_s^i$.

The nearshore camera monitoring sub-task was formulated as follows: given an outboard profile image, the model determined what type of vessel that given image involves. The outboard profile image dataset was formulated as $I_o^i, \forall i \in \{1, \cdots, \lambda\}$, where $I_o^i$ is the $i$th image in the dataset. $\lambda$ denotes the total number of image in the outboard profile image dataset. The output of this sub-task was defined as $f^o : I_o^i \to \{y_o\}$, where $y_o$ is the prediction label of the type of vessel.

3.2. Remote Satellite Monitoring Module

The remote satellite monitoring module is a two-step pipeline system consisting of a preprocessing and augmentation unit and a remote residual network block unit. The architecture of the remote satellite monitoring unit is illustrated in Figure 2.

![Figure 2. The architecture of remote satellite monitoring module.](image-url)
3.2.1. Preprocessing and Augmentation Unit

Preprocessing and augmentation units arm to raise the image’s quality make the model more effective, decrease model training time, and increase model inference speed.

We preprocessed the remote image set in three steps, grayscale conversion, mean normalization, and data standardization.

1) Grayscale Conversion

Grayscale compression reduces an image to its bare minimum number of pixels. It helps in simplifying algorithms and also eliminates the complexities related to computational requirements. It makes room for easier learning for remote satellite monitoring module image processing. We use the weighted mean method to realize the grayscale conversion. Grayscale compression reduces an image to its minimum number of pixels. This simplifies algorithms and eliminates computational complexities, making it easier for the remote satellite monitoring module to process images. We use the weighted mean method [32] to achieve grayscale conversion.

2) Zero-Centering Normalization

Zero-centering of the data is performed in this task prior to data standardization. Each pixel value is subtracted from the average value of the pixels’ sub-sample, resulting in zero-centered data, where the average of the pixels is zero. Using zero-centered image data is advantageous when employing activation functions as it helps prevent gradient saturation. This step effectively reduces repeated shocks during training of the remote satellite monitoring module.

3) Data Standardization

Data standardization involves modifying the data of each channel/tensor so that the mean is zero and the standard deviation is one. This process ensures that standardized data falls within the same range as activation functions, specifically between 0 and 1. As a result, there are fewer non-zero gradients during remote satellite monitoring module training, which enables the neurons in our network to learn more quickly.

Upon completion of the preprocessing step, the resultant remote image data is purified and rendered more trainable, thereby enabling its feeding into a data augmentation component. The latter enhances the accuracy of remote satellite monitoring performance through a series of transformations. In view of the characteristics of SAR image, we have technically adopted three kinds of data enhancement techniques to make full use of the limited training data.

The remote detection module is mainly applied to sea monitoring, which is far away from the port area. Vessels in the sea travel freely, making the angle of vessels entering the monitoring area variable. Hence, we adopt the random rotation [33], translation, flipping and scaling technologies to effectively simulate this feature. The algorithm of the preprocessing and augmentation unit is described in Algorithm 1.

3.2.2. Remote Residual Networks Block Unit

The output from the preprocessing and augmentation unit, $I_s^1$, was subsequently fed into the residual network block unit for feature extraction. To achieve effective monitoring of offshore vessel traffic, we utilized the basic block of the residual network as a backbone.

In the residual networks for remote unit, the SAR image information was fed to a convolution layer to obtain higher level representations. The convolution layer is the core layer of a convolutional neural network (CNN), which extracts advanced information by scanning information through kernels. Then, those representations were adopted by a maxpooling layer to obtain the most representative features. The formula for the convolution layer is shown below:

$$s_{ij} = \sum_m \sum_n x_{i+m} \times w_{m,n}$$ (1)
where $s$ is the feature extracted, $x$ is the input of the convolution layer, the $w$ is the weight of the convolution kernel, $i, j$ are the dimension of the extracted information, and $m, n$ are the dimension of the convolution kernel.

**Algorithm 1** Proprocessing and Augmentation Algorithm for Remote Detection

**Require:** synthetic-aperture radar image set $I_i^s, \forall \in \{1, \cdots, \delta\}$; hyper-parameters $u, v, z$; Red channel function, Red($\cdot$); Green channel function, Green($\cdot$); Blue channel function, Blue($\cdot$); Blurring function, AddingNoise function, BrightnessRegulation function.

1: # Grayscale conversion
2: for $i$ in $1 \cdots \delta$ do
3:      $I_i^s = u \times \text{Red}(I_i^s) + v \times \text{Green}(I_i^s) + z \times \text{Blue}(I_i^s)$
4: endfor
5: # Mean normalization
6: $I_{\text{mean}} \leftarrow \sum_{i=1}^{\delta} I_i^s$
7: for $i$ in $1 \cdots \delta$ do
8:      $I_i^s = I_i^s - I_{\text{mean}}$
9: endfor
10: # Data standardization
11: $I_i^s = \frac{I_i^s}{\delta \times \delta}$, $\delta$ is the standard deviation
12: # Augmentation
13: $I_i^s \leftarrow \text{Translation}(I_i^s)$
14: $I_i^s \leftarrow \text{Scaling}(I_i^s)$
15: $I_i^s \leftarrow \text{Flipping}(I_i^s)$
16: $I_i^s \leftarrow \text{RandomRotation}(I_i^s)$
17: $I_i^s = I_i^s \cup I_i^s \cup I_i^s \cup I_i^s$

Ensure: preprocessed synthetic-aperture radar image set $I_i^s$

Next, these representations were passed through several residual building blocks [34] to achieve deeper feature extraction. The structure of the residual building block is illustrated in Figure 3, and it comprises a convolution layer, a batch normalization layer, and an activation layer. Each residual building block has two paths, the Residual($x$) path and the Residual($x$)+$x$ path, the latter of which can be realized using feed-forward neural networks with “shortcut connections”. These shortcut connections aid in information transmission by allowing it to skip one or more layers, thus preventing the model from overfitting.

![Figure 3. The structure of the residual building block.](image-url)
After processing the residual building block information, it is sent to the next layer, which is the avgpooling layer, for dimensionality reduction. Then, the information is passed into a fully connected layer component for the final prediction. The formula is as follows:

$$y_s \leftarrow \text{softmax}(W_s \cdot I^s + b_s)$$  \hspace{1cm} (2)

where $W$ and $b$ are parameters in a neural network.

Details of the residual networks block unit for the remote task are presented in Algorithm 2.

**Algorithm 2 Residual Networks for Remote Algorithm**

**Require:** Preprocessed SAR image set $I^s_i, \forall \in \{1, \cdots, \delta\}$; basic block of residual networks, Residual(); convolution layer, Conv(); maxpooling function, Maxpool(); avgpooling function, Avgpool(); Fully Connected Layer, FC().

1. $I^s_i \leftarrow \text{Conv}(I^s_i)$
2. $I^s_i \leftarrow \text{Maxpool}(I^s_i)$
3. for $i$ in [3,4,6,3] do # [3,4,6,3] is the number of the residual basic block
4.   for $j$ in 1 to $i$ do
5.     $I^s_i \leftarrow \text{Residual}(I^s_i)$
6.   endfor
7. endfor
8. $I^s_i \leftarrow \text{Avgpool}(I^s_i)$
9. $I^s_i \leftarrow \text{FC}(I^s_i)$ # Dimensionality reduction
10. # Remote identification:
11. $y_s \leftarrow \text{softmax}(W_s \cdot I^s_s + b_s)$

**Ensure:** Remote detection labels $y_s$

### 3.3. Nearshore Monitoring Module

The nearshore monitoring module is similar to the remote satellite monitoring module in that it has a two-step pipeline system. This system consists of a preprocessing and augmentation unit as well as a residual network block for the nearshore unit. The architecture of the nearshore monitoring module is shown in Figure 4.

![Figure 4. The architecture of the nearshore camera monitoring module.](image)

In the preprocessing and augmentation unit, we used the same methods for data preprocessing. However, we abandoned the data enhancement method for remote monitoring because they were not suitable for the vessel nearshore monitoring task. Instead, we added some new data enhancement methods that are tailored to the characteristics
of nearshore monitoring. These methods help improve the detection performance of our system. The nearshore surveillance cameras are remote monitoring equipment that works all day and in all weather, and the images captured by the cameras show a large exposure difference. Consequently, we deliberately selected blurring, brightness regulation, and adding noise technologies to simulate various photo scenarios, facilitating the model’s better recognition, as shown in Algorithm 3.

**Algorithm 3 Preprocessing and Augmentation Algorithm for Nearshore Detection**

**Require:** Nearshore image set $I_o$, $\forall \in \{1, \cdots, \delta\}$, hyper-parameters $u, v, z$; Red channel function, Red(); Green channel function, Green(); Blue channel function, Blue(); Random Rotation function, Translation function, Scaling function, Flipping function.

1: # Grayscale conversion
2: for $i$ in $1 \cdots \delta$ do
3: $I_o^i = u \times \text{Red}(I_o^i) + v \times \text{Green}(I_o^i) + z \times \text{Blue}(I_o^i)$
4: endfor
5: # Mean normalization
6: $I_{mean}^o \leftarrow \sum_{i=1}^{\delta} I_o^i$
7: for $i$ in $1 \cdots \lambda$ do
8: $I_o^i = I_o^i - I_{mean}^o$
9: endfor
10: # Data standardization
11: $\bar{I}_o^i = I_o^i / \delta$, $\delta$ is the standard deviation
12: # Augmentation
13: $\hat{I}_o^i \leftarrow \text{Blurring}(I_o^i)$
14: $\tilde{I}_o^i \leftarrow \text{AddingNoise}(I_o^i)$
15: $\hat{I}_o^i \leftarrow \text{BrightnessRegulation}(I_o^i)$
16: $I_o^i = I_o^i \cup \hat{I}_o^i \cup \tilde{I}_o^i$

**Ensure:** Preprocessed nearshore image set $I_o^i$

The residual networks block for the nearshore unit has the same structure as that for the remote unit, as shown in the details above. However, it is important to note that although the residual network blocks for nearshore and remote scenarios have the same architecture; they are two separate modules with different input requirements.

The information extracted from the residual building blocks was connected to the avgpooling layer for dimensionality reduction. Then, the information was passed into a fully connected layer component for classifying the type of vessel. The formula is as follows:

$$y_o \leftarrow \text{softmax}(W_o \cdot I_o^i + b_o)$$

(3)

where $W$ and $b_o$ are parameters in a neural network.

### 3.4. Two-Phase Training Mode

We adopt a two-phase training mode to train the system to further enhance its detection accuracy; the training step is shown in Figure 5. The traditional training method is isolated and occurs purely based on specific tasks, datasets, and training separate isolated models on them. However, obtaining satisfactory model performance using only a limited amount of data for training is tricky.

The core idea of transfer learning is reusing a pretrained model as the starting point for a model on a new task; that is, a model trained on one task is repurposed on a second, related task as an optimization that allows rapid progress when modeling the second task.
Inspired by the inductive transfer learning, in the first training phase, a residual network was trained by a large-scale hierarchical image database (the number of types: 1000). The goal of this training phase is to obtain a powerful and universal classification residual network, which contains abundant available knowledge [35]. To transfer knowledge effectively, we carefully selected the pretrained model. After evaluating five pretrained models, including ResNet18, ResNet34, ResNet50, ResNet101, and ResNet152, we found that utilizing the ResNet34 model resulted in the most accurate performance for our model.

In the second training phase, the nearshore and remote data sets were used respectively for the retraining of the well-trained residual network module. In this scenario, the task of the first training phase and the task of the second training phase domains are the same, yet the first training phase and the second training phase tasks are different from each other. The algorithms try to utilize the inductive biases of the first training phase domain to help improve the second training phase task, obtaining bilateral friendly information representations.

4. Experiments and Results

4.1. Dataset

Two extensive vessel detection datasets (i.e., nearshore and remote) are chosen to evaluate the performance of the KB-VMS system. The sample of these two datasets are shown in Figures 6 and 7. The nearshore dataset includes outboard profile images of vessels from five distinct categories, namely, Cargo, Carrier, Cruise, Military, and Tankers. The number of images in each category ranges from 832 to 2120. The remote dataset used in this study consists of images obtained from synthetic-aperture radar (SAR) satellites. The dataset was categorized based on the number of vessels present in the image, which was discretized into five categories: 1, 2, 3, 4, and greater than 4. The mathematical statistics of both datasets are presented in Table 2.

In the nearshore dataset, the number of cargo vessel images is the highest, with a total of 2120 images, accounting for approximately 34% of the dataset, while the other four types of vessels each account for less than 20%. In the remote dataset, the phenomenon of imbalanced data distribution is more pronounced. Among the images, there are 439 images with only 1 vessel, 98 images with 2 vessels, 26 images with 3 vessels, and only 15 images with 4 vessels. There are 43 images with more than 4 vessels. It can be seen that both datasets suffer from imbalanced data distribution issues. To address the issue of model training bias arising from imbalanced data, we employed data augmentation techniques to balance the dataset, achieving a relatively uniform distribution of data among each class. The distribution of the balanced dataset is presented in Table 3.
<table>
<thead>
<tr>
<th>Dataset Category</th>
<th>Vessel #1</th>
<th>Vessel #2</th>
<th>Vessel #3</th>
<th>Vessel #4</th>
<th>Vessel # &gt; 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Remote Dataset</td>
<td>352</td>
<td>79</td>
<td>21</td>
<td>12</td>
<td>35</td>
</tr>
<tr>
<td>Nearshore Dataset</td>
<td>1696</td>
<td>733</td>
<td>666</td>
<td>934</td>
<td>974</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Category</th>
<th>Original Data</th>
<th>Balanced Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cargo</td>
<td>2120</td>
<td>2120</td>
</tr>
<tr>
<td>Carrier</td>
<td>916</td>
<td>2170</td>
</tr>
<tr>
<td>Cruise</td>
<td>832</td>
<td>2121</td>
</tr>
<tr>
<td>Military</td>
<td>1167</td>
<td>2334</td>
</tr>
<tr>
<td>Tankers</td>
<td>1217</td>
<td>2434</td>
</tr>
</tbody>
</table>

4.2. Evaluation Metric

The KB-VMS system detection performance is evaluated quantitatively in terms of accuracy, precision, recall, and F1, which are as follows:

\[
\begin{align*}
\text{Accuracy} & = \frac{TP + TN}{TP + FP + FN + TN}, \\
\text{Recall} & = \frac{TP}{TP + FN}, \\
\text{Precision} & = \frac{TP}{TP + FP}, \\
F1 & = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}},
\end{align*}
\]
where TP represents a positive sample being predicted as a positive sample. TN represents a negative sample being predicted as a negative sample. FP represents a negative sample being predicted as a positive sample. and FN represents a positive sample being predicted as a negative sample.

4.3. Experiment Setup

(1) Parameters Setting

We adopted a pretrained ResNet34 model on the ImageNet dataset as the backbone of the KB-VMS monitoring module. The whole model is optimized with the proposed loss function that integrates the probabilistic classification loss with the multi-class cross-entropy loss. The adopted optimizer is stochastic gradient descent (SGD) with momentum. we empirically set the batch size as 32 and the learning rate is 0.001. Experimental data is partitioned randomly into training and testing sets at an 8:2 ratio.

(2) Experimental Environment.

The experiments were performed on a computer system comprising a 64-bit Windows 10 operating system, a 12th Gen Intel Core i7-12700 processor, 32 GB of memory, and an NVIDIA GeForce RTX 3060 graphics card. The PyTorch 11.7 deep learning framework was employed, with PyCharm serving as the primary software tool and Python 3.11 as the programming language.

4.4. Experimental Results

To provide a quantitative assessment of our proposed method, we compared the KB-VMS system performance with that of several modern convolutional neural network (CNN)–based methods, including commonly used classification deep learning models such as AlexNet, GoogLeNet, VGG-16, VGG-19, ResNet-34, ResNet-50, ResNet-101, ResNet50-32x4d, and state-of-the-art (STOA) models including CNN-MLP, Cascade CNN, DCMSNN, DDNET, GBCNN, and KB-VMS. The selection of models was based on their established reputation in the field and their potential for achieving high accuracy in detection tasks. The evaluation of the models was conducted using established performance metrics and a rigorous experimental setup to ensure the validity and reliability of the results.

The main experimental results are shown in Table 4. The KB-VMS model exhibits remarkable enhancements over the baseline models in both nearshore and remote detection tasks, with particularly notable increases in detection accuracy for remote sensing. The accuracy of the baseline models for nearshore detection ranges from 70.22% to 79.43%, while the accuracy for remote monitoring is lower due to the smaller training dataset, with the highest monitoring accuracy only reaching 57.53%. In comparison to the baseline models, our model shows a significant increase in detection accuracy, with nearshore detection accuracy reaching 97.18% and remote detection accuracy reaching 94.43%. Our model builds upon the ResNet-34 architecture, incorporating a supplementary data processing and augmentation module as well as transfer learning strategies during model training, resulting in substantial improvements in vessel detection performance. The baseline ResNet-34 model exhibited nearshore and remote detection accuracies of 78.94% and 55.38%, respectively. In contrast to the original ResNet-34 model, our model demonstrated an 18% increase in nearshore detection accuracy and a 38% increase in remote detection accuracy. The models CNN-MLP, Cascade CNN, DCMSNN, DDNET, GBCNN, and KB-VMS were not capable of conducting nearshore detection. Nonetheless, these models showed good remote detection performance compared to baseline models. Based on F1 scores, all of these models, except GBCNN, achieved detection performance of around 90%. The CNN-MLP model demonstrated the highest remote detection performance, with an F1 score of 91.91. Our model exhibited a 4% improvement in F1 score compared to CNN-MLP.

We further evaluated the performance for nearshore and remote vessel detection on the baseline models and our model using the ROC curve. We calculated the AUC-ROC value and plotted the ROC curve, which showed that our model achieved a higher true
The ROC curve experimental results are shown in Figures 8 and 9.

<table>
<thead>
<tr>
<th>Method</th>
<th>Nearshore Detection Task (%)</th>
<th>Remote Detection Task (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Accuracy</td>
<td>Recall</td>
</tr>
<tr>
<td>AlexNet [36]</td>
<td>70.22</td>
<td>70.52</td>
</tr>
<tr>
<td>GoogLeNet [37]</td>
<td>73.67</td>
<td>74.01</td>
</tr>
<tr>
<td>VGG-16 [36]</td>
<td>75.23</td>
<td>75.48</td>
</tr>
<tr>
<td>VGG-19 [38]</td>
<td>73.67</td>
<td>73.82</td>
</tr>
<tr>
<td>ResNet-34 [34]</td>
<td>78.94</td>
<td>79.13</td>
</tr>
<tr>
<td>ResNet-50 [34]</td>
<td>79.43</td>
<td>79.53</td>
</tr>
<tr>
<td>ResNet-101 [34]</td>
<td>79.34</td>
<td>79.47</td>
</tr>
<tr>
<td>ResNext50-32x4d [39]</td>
<td>78.54</td>
<td>78.67</td>
</tr>
<tr>
<td>CNN-MLP [22]</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Cascade CNN [26]</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>DCMSNN [13]</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>DDNET [24]</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>GBCNN [29]</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>KB-VMS</td>
<td>97.18</td>
<td>97.12</td>
</tr>
</tbody>
</table>

Table 4. Comparison of quantitative evaluation indices with other methods.

Figure 8. ROC curves for different methods in nearshore detection. (a) AlexNet; (b) GoogLeNet; (c) VGG-16; (d) VGG-19; (e) ResNet-34; (f) ResNet-50; (g) ResNet-101; (h) ResNext50-32x4d; (i) KB-VMS.
Figure 9. ROC curves for different methods in remote detection. (a) AlexNet; (b) GoogLeNet; (c) VGG-16; (d) VGG-19; (e) ResNet-34; (f) ResNet-50; (g) ResNet-101; (h) ResNext50-32x4d; (i) KB-VMS.

Based on the results obtained from the ROC curves in the nearshore monitoring task, it is apparent that baseline deep learning models display a good recognition performance for vessels classified as Carrier, as evidenced by their ROC value of approximately 0.96. Conversely, the recognition performance for vessels classified as Cargo or Tankers is less satisfactory, with ROC values hovering around 0.90. Nevertheless, the KB-VMS demonstrates a distinct detection behavior, achieving an average ROC value of 0.954, with ROC values for Cargo vessels and Tankers also reaching as high as 0.995.

Based on the ROC curves obtained in the remote monitoring task, it is apparent that the baseline models’ overall ROC curve performance is sub-optimal. Specifically, when the monitored area in the satellite cloud images contains only two to three vessels, the baseline models’ average ROC value is around 0.65, indicating a high likelihood of detection errors when the number of vessels in the satellite cloud images is low. In contrast, our proposed model displays exceptional monitoring performance in remote detection tasks, achieving a macro-average ROC value of 0.993, as well as ROC values of 0.98 and 0.99 when the satellite cloud images contain two or three vessels, respectively. These observations serve to highlight our model’s capacity for delivering stable performance in remote detection tasks.

4.5. Preprocessing and Augmentation Unit Impact Study

To investigate the impact of various data processing techniques on the model, we conducted an extensive exploration of each approach employed in the preprocessing and augmentation unit. Transfer learning was not utilized in training the model in this
experiment to remove any extraneous variables. The study employed ResNet-34 as the fundamental test model, and the findings are presented in Table 5. The experimental outcomes reveal that the chosen data processing techniques enhance the model’s capability to achieve more accurate recognition.

Table 5. Preprocessing and augmentation method performance.

<table>
<thead>
<tr>
<th>Preprocessing Method</th>
<th>Accuracy (%)</th>
<th>Recall (%)</th>
<th>Precision (%)</th>
<th>F1 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nearshore</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Original</td>
<td>73.52</td>
<td>72.57</td>
<td>73.09</td>
<td>73.53</td>
</tr>
<tr>
<td>Blurring</td>
<td>74.06</td>
<td>74.13</td>
<td>74.19</td>
<td>73.97</td>
</tr>
<tr>
<td>Adding Noise</td>
<td>76.61</td>
<td>76.86</td>
<td>76.91</td>
<td>76.85</td>
</tr>
<tr>
<td>Brightness Regulation (reduction)</td>
<td>77.91</td>
<td>78.15</td>
<td>78.34</td>
<td>78.21</td>
</tr>
<tr>
<td>Brightness Regulation (augmentation)</td>
<td>78.24</td>
<td>78.52</td>
<td>78.75</td>
<td>78.58</td>
</tr>
<tr>
<td>Remote</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Original</td>
<td>42.09</td>
<td>41.05</td>
<td>38.86</td>
<td>37.69</td>
</tr>
<tr>
<td>Translation</td>
<td>44.38</td>
<td>42.65</td>
<td>39.17</td>
<td>38.92</td>
</tr>
<tr>
<td>Scaling</td>
<td>45.41</td>
<td>43.52</td>
<td>41.48</td>
<td>41.03</td>
</tr>
<tr>
<td>Flipping</td>
<td>46.68</td>
<td>45.39</td>
<td>45.62</td>
<td>43.59</td>
</tr>
<tr>
<td>Random Rotation</td>
<td>47.44</td>
<td>46.73</td>
<td>47.03</td>
<td>45.26</td>
</tr>
</tbody>
</table>

Furthermore, we conducted an in-depth investigation of the impact of different data processing and augmentation techniques on the model’s ability to recognize various types of vessels. Specifically, we evaluated the effects of these techniques on nearshore monitoring tasks and present our findings in Figure 10. Our experimental results demonstrate that adjusting the brightness of the training dataset, through either brightness reduction or augmentation, can significantly improve the model’s ability to detect Military vessels. However, this adjustment can also increase the risk of the model misclassifying Cargo vessels as Tankers, leading to reduced prediction accuracy for Cargo vessels. Moreover, our analysis shows that methods such as blurring, adding noise, and brightness regulation can effectively enhance the model’s prediction accuracy for Carriers, Cruise vessels, and Tankers.

Figure 10. Data processing and augmentation performance analysis for nearshore detection. (a) Original; (b) blurring; (c) adding noise; (d) brightness reduction; (e) brightness augmentation.
A comprehensive examination of data processing and augmentation techniques for remote sensing tasks was carried out. The outcomes of experiments are presented in Figure 11. The utilization of data augmentation methods in remote sensing tasks has contributed to enhancing the detection accuracy of the model, particularly in scenarios where monitoring regions contain only one vessel. We observed that in the absence of data processing, the model tends to misclassify images featuring three vessels in the monitoring region as having more than four vessels. This issue can be effectively resolved by applying flipping and random rotation techniques. Furthermore, our findings demonstrate that data processing or augmentation can potentially have a slight adverse effect on the model’s detection performance when the monitoring area encompasses only one vessel.

Figure 11. Data processing and augmentation performance analysis for remote detection. (a) Original; (b) translation; (c) scaling; (d) flipping; (e) random rotation.

4.6. Transfer Learning Impact Study

We investigate the impact of transfer learning on model performance by comparing the loss values and accuracy during the training process. The experimental results, as depicted in Figure 12, indicate that transfer learning provides the model with generalizable recognition capabilities for both remote sensing and coastal monitoring tasks. This ability facilitates rapid learning and improves the detection accuracy of the model. Notably, the benefits of transfer learning are particularly evident in remote sensing tasks, where the introduction of this technique leads to a significant reduction in performance fluctuations and results in faster convergence and higher accuracy, as evidenced by both the loss values and accuracy metrics.

4.7. Error Analysis

We perform a comprehensive error analysis of the baseline deep learning model and our proposed model using a confusion matrix. The experimental results, shown in Figures 13 and 14, reveal that the overall misclassification rate of our model is still significantly lower than that of the baseline models.
Figure 12. Transfer learning impact results. (a) Nearshore monitoring training results (left: without knowledge transfer technique. Right: with knowledge transfer technique); (b) Remote monitoring training results (left: without knowledge transfer technique. Right: with knowledge transfer technique).

For the nearshore detection error analysis, the results of Figure 13 indicate that discriminating between Cargo and Tanker vessels presents a challenge for both models in the context of nearshore monitoring. Specifically, the baseline models often misclassify these two types of vessels and exhibit a bias towards labeling Tankers as Cargo vessels. In contrast, our proposed model performs well in recognizing Cargo vessels and only misclassifies seven out of the total samples as Tankers. However, we also observe that our model tends to predict Tankers as Cargo vessels and misclassifies 18 Tankers. The experimental results also demonstrate that our model significantly reduces the misclassification of Cruise vessels. Baseline models tend to misclassify Cruise vessels as Military, Carrier, or Cargo vessels, but our model rarely misclassifies Cruise vessels as Carriers, Cargo vessels, or Tankers, and only occasionally misclassifies them as Military vessels.
We present an analysis of the performance of remote vessel detection with a focus on the impact of data augmentation on detection accuracy. Our findings indicate that the baseline models are most effective in detecting images with over four vessels in the detection area, while images with two or three vessels are more challenging to detect accurately. Specifically, our analysis shows that when there are only two vessels in the detection area, the baseline model tends to misclassify it as only one vessel, and in the detection area with only three vessels, the baseline model tends to overestimate the number of vessels, identifying three or more vessels. The KB-VMS outperforms the baseline models, particularly in the detection area with four vessels, where our model achieves high detection precision. While there is a slight probability that our model may misclassify the detection area with only two vessels as having only one vessel, this tendency is still a significant improvement compared to that of the baseline models. Our results demonstrate the effectiveness of data augmentation and transfer learning for further improvements in remote vessel detection through the use of advanced deep learning techniques.
5. Conclusions

In this article, we introduce a novel bi-directional vessel monitoring system (KB-VMS) that leverages knowledge transfer to enable realistic maritime monitoring scenarios. Our approach entails the integration of remote satellite equipment and nearshore detection equipment, thus enabling a comprehensive and efficient monitoring framework. A meticulous two-stage training mode based on transfer learning mechanism was designed for the KB-VMS system, resulting in improved vessel identification performance in both nearshore and remote detection areas compared to state-of-the-art models and baselines on two real-world vessel monitoring datasets.

In addition, our experiments have also provided insights into the impact of different data augmentation techniques and transfer learning mechanisms on ship detection tasks. Furthermore, our supplementary experiments results revealed several characteristic behaviors of deep-learning-based ship detection models in vessel monitoring task. For instance, in nearshore identification task, distinguishing between Tankers and Cargo...
vessels presents a greater challenge for these models. Moreover, in remote detection tasks, the deep-learning-based ship detection models exhibit less accurate judgments when detecting two or three ships in the detection area but accurate detection when the number of ships exceeds four in the detection area. These experimental findings can provide valuable clues for future in-depth investigations into the vessel monitoring field. Further analysis of the nearshore detection errors reveals that our model greatly reduces the misclassification of other types of vessels as military vessels compared to the base model. This is certainly a reduction in unnecessary interference with non-military vessels for the military domain, as mistakenly identifying commercial vessels as military vessels may lead to warnings, interceptions, or even attacks on these vessels. In addition, properly distinguishing between military and non-military vessels can improve the accuracy of intelligence analysis and optimize the use of military resources. Meanwhile, vessel monitoring system can help maritime enterprises to realize remote monitoring and management. Correctly determining the number of vessels in the water can help shipping companies to make better transportation planning and improve logistics efficiency, and safety. In addition, analyzing the number of vessels can help shipping companies to understand the competitive environment and predict future market trends so as to make more informed business decisions.

In summary, the KB-VMS system can not only improve the combat effectiveness and survivability of ships but can also improve the competitiveness and social benefits of the maritime transportation industry.

The field of ship monitoring presents numerous avenues for further exploration. Among these, the integration of sonar analysis with ROI analysis holds promise for providing added insight into ship movements and behaviors, thereby enhancing the capabilities of vessel detection and classification systems. Moreover, the incorporation of attention mechanisms can prove effective in improving model performance. By gaining a better understanding of the decision-making process, potential biases or errors can be identified and addressed. Visualization of the model’s focus on particular image regions further aids in the identification of potential biases or errors, ultimately leading to improved system performance. Further research into the integration of these techniques holds the potential to advance the field and contribute to improved maritime safety and security.

**Author Contributions:** Conceptualization, Methodology, Resources, Software, Validation, Supervision, and Writing—original draft preparation, J.L.; Methodology, Conceptualization, Writing—review and editing, Y.Y.; Data curation, Validation, X.L.; Validation, Software, J.S.; Funding acquisition, Methodology, Supervision, R.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Young Innovative Talents Grants Program of Guangdong Province (Grant No. 2022KQNCX024), the Ocean Young Talent Innovation Program of Zhanjiang City (Grant No. 2022E05002), the National Natural Science Foundation of China (Grant No. 52171346), the Natural Science Foundation of Guangdong Province (Grant No. 2021A1515012618), the special projects of key fields (Artificial Intelligence) of Universities in Guangdong Province (Grant No. 2019KZDZX1035), the program for scientific research start-up funds of Guangdong Ocean University, and the College Student Innovation Team of Guangdong Ocean University (Grant No. CXTD2021013).

**Conflicts of Interest:** The authors declare no conflict of interest. This submitted manuscript is approved by all authors for publication. We would like to declare that the work described is original research that has not been published previously, and is not under consideration for publication elsewhere, in whole or in part. The authors listed have approved the manuscript that is enclosed.

**References**


31. Połap, D.; Włodarczyk-Sielicka, M.; Wawrzyniak, N. Automatic ship classification for a riverside monitoring system using a cascade of artificial intelligence techniques including penalties and rewards. ISA Trans. 2022, 121, 232–239. [CrossRef]

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