Article

Research and Application of Panoramic Visual Perception-Assisted Navigation Technology for Ships

Chiming Wang [1], Xiaocong Cai [1], Yanan Li [1], Runxuan Zhai [1], Rongjiong Wu [1], Shunzhi Zhu [1,*], Liangqing Guan [2], Zhiqiang Luo [2], Shengchao Zhang [3] and Jianfeng Zhang [3]

[1] School of Computer and Information Engineering, Xiamen University of Technology, Xiamen 361024, China
[2] Fujian Fuchuan Marine Engineering Technology Research Institute Co., Ltd., Fuzhou 350501, China
[3] Xiamen Port Shipping Co., Ltd., Xiamen 361012, China
* Correspondence: algorithm_007@163.com

Abstract: In response to challenges such as narrow visibility for ship navigators, limited field of view from a single camera, and complex maritime environments, this study proposes panoramic visual perception-assisted navigation technology. The approach includes introducing a region-of-interest search method based on SSIM and an elliptical weighted fusion method, culminating in the development of the ship panoramic visual stitching algorithm SSIM-EW. Additionally, the YOLOv8s model is improved by increasing the size of the detection head, introducing GhostNet, and replacing the regression loss function with the WIoU loss function, and a perception model yolov8-SGW for sea target detection is proposed. The experimental results demonstrate that the SSIM-EW algorithm achieves the highest PSNR indicator of 25.736, which can effectively reduce the stitching traces and significantly improve the stitching quality of panoramic images. Compared to the baseline model, the YOLOv8-SGW model shows improvements in the P, R, and mAP50 of 1.5%, 4.3%, and 2.3%, respectively, its mAP50 is significantly higher than that of other target detection models, and the detection ability of small targets at sea has been significantly improved. Implementing these algorithms in tugboat operations at ports enhances the fields of view of navigators, allowing for the identification of targets missed by AISs and radar systems, thus ensuring operational safety and advancing the level of vessel intelligence.

Keywords: panoramic vision; object detection; visual perception; intelligent vessels

1. Introduction

The development and demands of the marine transportation industry have promoted the innovation of ship intelligence technology, and the intelligentization of ships has become a new trend in the development of the shipping industry [1,2]. Ship navigation environment perception refers to the use of radar, LiDAR, Automatic Identification Systems (AISs), depth sounders, cameras, and other sensors to collect ship navigation environment data, as well as the building of intelligent algorithms for perception enhancement, data fusion, target classification, decision recommendations, automatic processing and analyzing of perception data, and distinguishing potential dangers and abnormal situations (such as channel obstacles and past ships). Emergency response measures (such as adjusting the route and speed) to improve the navigation safety of the ship have also been determined.

In the field of ship perception, many scholars have devoted themselves to exploring how to use computer vision and image processing technology to enhance maritime environment perception abilities, and some scholars have used target detection technology to achieve the automatic perception of maritime targets visually [3]. Qu et al. [4] proposed an anti-occlusion vessel tracking method to predict the vessel positions, so as to obtain a synchronized AIS and visual data for data fusion. It can overcome the vessel occlusion problem and improve the safety and efficiency of ship traffic. Chen et al. [5] proposed a

new small-ship detection method using a Convolutional Neural Network (CNN) and an improved Generative Adversarial Network (GAN) to help autonomous ships navigate safely. Faggioni et al. [6] proposed a low-computational method for multi-object tracking based on LiDAR point clouds, tested in virtual and real data, demonstrating good performance. Zhu et al. [7] proposed the target detection algorithm YOLOv7-CSAW for the sea. Compared with the YOLOv7 model, the algorithm enhanced the accuracy and robustness of small-target detection in complex scenes and reduced the occurrence of missed detection. Maritime targets mainly include ships, buoys, and reefs. Due to insufficient datasets, current research on the visual perception of maritime targets is mainly focused on ships [8], and small targets such as buoys mainly use LiDAR for perception [9,10]. Few visual perception methods have been developed that can perceive other targets. Moreover, there are cases of missed detection. Further research is needed for practical applications.

A single camera has a small field of view and captures limited information about the environment. Multiple cameras can capture more environmental information, but it is presented from multiple perspectives; the crew needs to perform frequent perspective switching and cannot watch comprehensive environmental information at the same time. Stitching images from different perspectives into panoramic images for display can solve these problems. In this regard, many scholars have conducted research on panoramic vision technology and have achieved certain results in the field of land transportation [11]. Zhang et al. [12] proposed a distortion correction (DC) and perspective transformation (PT) method based on a LookUp Table (LUT) transformation to enhance the processing speed of image stitching algorithms, generating panoramic images to assist in parking. Christian et al. [13] introduced a real-time image stitching method to improve the horizontal field of view for object detection in autonomous driving. Compared to land transportation systems such as intelligent driving perception in automobiles, which can provide all key visual information related to the driving environment [14] and assist in collision avoidance, pedestrian and vehicle detection, and visual parking aid [15], the panoramic perception systems of ships are still in their infancy and are not widely applied in maritime settings [16].

To enhance the perception capabilities of ship navigation environments, this study is based on panoramic image stitching and object detection algorithms. A real-time environmental perception technology for assisting ship navigation is proposed, and a visual assistance perception system for application on tugboats is constructed. The goal is to ensure the safety of ship navigation and promote the intelligent development of ships.

2. Construction of Ship Panoramic Vision Mosaic Algorithm

2.1. Panoramic Image Stitching Method

The panoramic image stitching algorithm refers to the alignment and fusion of multiple images by means of extraction, matching, and registration of image feature points [17]. Existing image stitching techniques typically extract features from the entire image during the feature extraction stage, resulting in many invalid feature points, which use computing resources and influence the accuracy of subsequent registration [18]. In terms of image fusion algorithms, there are mainly average weighting methods and fade-in/fade-out methods. Although these fusion algorithms can weaken the traces produced by image stitching, the effect is general, and there is still room for improvement [19].

This paper presents an image stitching approach: the SSIM, SURF, and Elliptical Weighted Algorithm Stitcher (SSIM-EW). In addition to the image stitching method based on Speeded-Up Robust Features (SURFs) [20], the region-of-interest search method based on the Structural Similarity Index Measure (SSIM) [21] and the elliptical weighted fusion method are added to improve the algorithm stitching performance and reduce issues such as gaps and artifacts that may arise during the stitching process. The specific steps are illustrated in Figure 1.
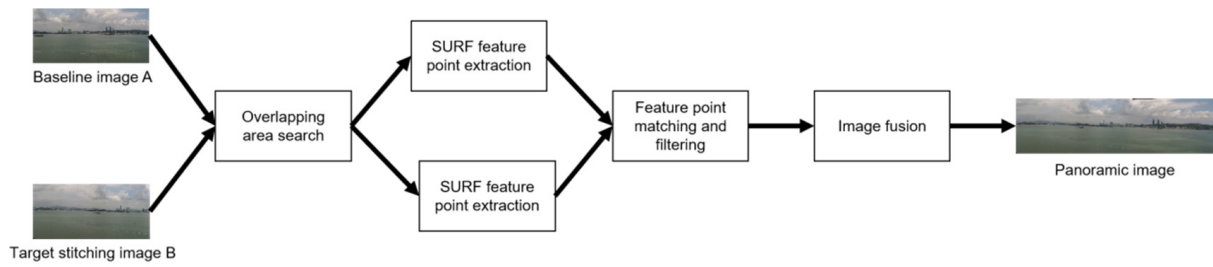
Figure 1. Flowchart of panoramic image stitching.

2.1.1. The Region-of-Interest Search Method Based on SSIM

By comparing the SSIM values of different image regions, the algorithm identifies the region with the highest SSIM value as the region of interest of the stitching algorithm. This approach reduces computational complexity, minimizes interference from invalid feature points, accelerates feature point extraction speed, and enhances image registration accuracy. The specific steps are as follows:

The image is divided into N equal areas along the image splicing direction, as shown in Figure 2. N is a hyperparameter; the larger the value of N is, the slower the speed is and the higher the positioning accuracy of the overlapping area is. In this research, the value of N is 10.
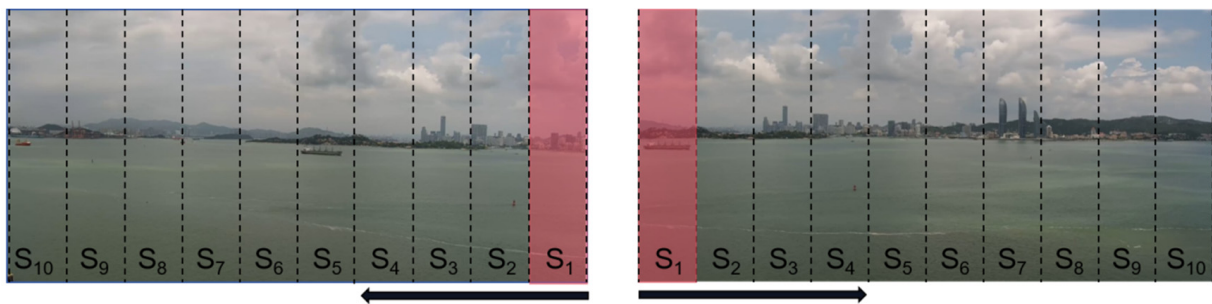

Figure 2. Schematic diagram of image area division.

After the image is divided, the regions are superimposed and combined along the divided direction, and each image forms N sub-images M of different sizes, as follows:

$$M_k = \sum_{i=1}^{k} S_i, \ k \in [1, N]. \tag{1}$$

The *SSIM* values of the corresponding sub-images are compared, and the *SSIM* calculation method is as follows:

$$SSIM(x, y) = \frac{\left(2\mu_x \mu_y + C_1\right)\left(2\sigma_{xy} + C_2\right)}{\left(\mu_x^2 + \mu_y^2 + C_1\right)\left(\sigma_x^2 + \sigma_y^2 + C_2\right)}, \tag{2}$$

where μ is the mean of the image, σ is the standard deviation, σ^2 is the image variance, and σ_{xy} represents the covariance between images x and y, while C_1 and C_2 are constants.

The result of the region-of-interest search is shown in Figure 3, where the sub-image *M5* has the highest *SSIM* value. Therefore, the region of interest is the combination of regions *S1–S5*.
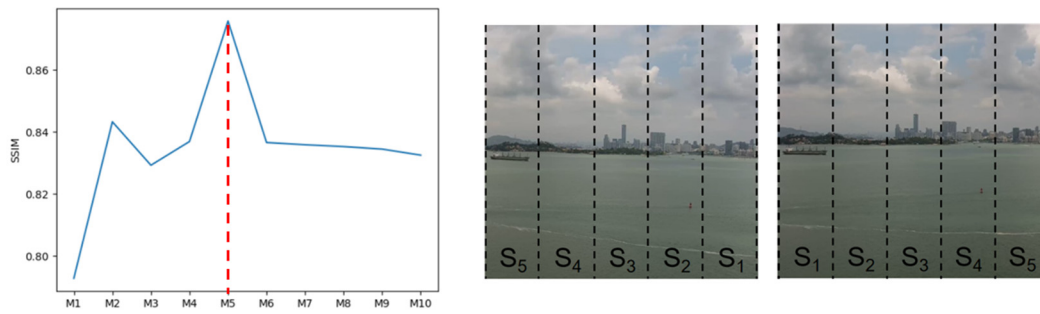
Figure 3. Results of overlapping area search.

2.1.2. Feature Point Extraction and Matching

In the SURF algorithm for feature extraction, box filters are used to process images and construct the scale space and to evaluate whether a pixel is located at an edge by comparing the determinants of the matrix. And in the area around each key point, through the calculation of continuous sub-windows of varying scales, the final generated feature vector is a 64-dimensional vector. The expression for the Hessian matrix is as follows:

$$H(I(x,y)) = \begin{bmatrix} \frac{\delta^2 I}{\delta x^2} & \frac{\delta^2 I}{\delta x \delta y} \\ \frac{\delta^2 I}{\delta x \delta y} & \frac{\delta^2 I}{\delta y^2} \end{bmatrix}, \tag{3}$$

where $I(x,y)$ represents the pixel point.

The process of feature point matching is divided into coarse matching and fine matching. In the coarse matching process, the Euclidean distance between the two groups of feature points is calculated, the matching points are preliminarily filtered according to the Euclidean distance, and the feature point pairs satisfying the following formula are retained.

$$\frac{Nearest\ Euclidean\ distance}{Second - nearest\ Euclidean\ distance} < Ratio. \tag{4}$$

The *Ratio* represents the threshold for the difference in matching pairs, which is used to measure the degree of difference between the nearest Euclidean distance matching pair and the second-nearest Euclidean distance matching pair. A smaller *Ratio* leads to fewer retained matching pairs but higher accuracy. In this research, the *Ratio* is set to 0.35 (0 < *Ratio* < 1).

In the fine matching, the RANSAC algorithm [22] is used to randomly sample the feature point pairs retained in the coarse matching stage, the best homography transformation matrix is continuously and iteratively calculated, and the outliers in the matching points are eliminated through the matrix. The final matching result is shown in Figure 4.
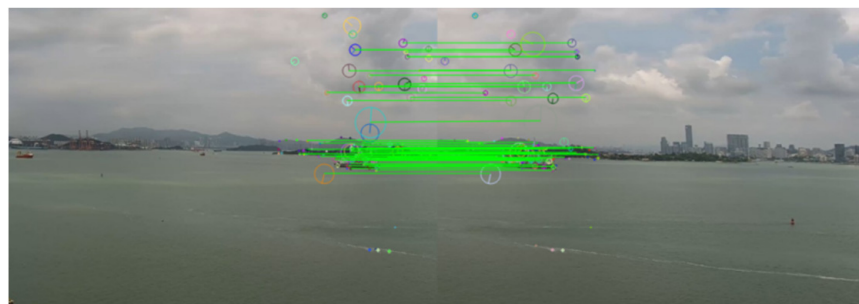

Figure 4. Result of feature point matching.

The homography transformation [23] is expressed as follows:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} m_0 & m_1 & m_2 \\ m_3 & m_4 & m_5 \\ m_6 & m_7 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = H \begin{bmatrix} x \\ y \\ 1 \end{bmatrix},$$ (5)

where H represents the homography transformation matrix, and (x, y) and (x', y') represent the pixel coordinates before and after the transformation, respectively. The pixel points of the image to be registered can be mapped to the pixel coordinate system of the reference image one by one through the best homography transformation matrix.

2.1.3. Image Fusion Based on Elliptic Function Weighting

The elliptical weighted fusion method can solve the issues of noticeable image stitching artifacts and unnatural transitions. As shown in Figure 5, $M_L(i, j)$ is the reference image whose abscissa ranges from 0 to L_2, $M_R(i, j)$ is an image to be stitched whose abscissa ranges from L_1 to L_3, x is the abscissa of any pixel point of the panoramic image, and L_1 and L_2 are the left and right boundaries of the overlapping region, respectively. An ellipse is constructed by taking L_1 as the center of the circle, $L_1 L_2$ as the semi-major axis, and the length of the semi-minor axis as 1. The ellipse curve in the overlapping area is taken as the weight curve of the reference image pixel point.
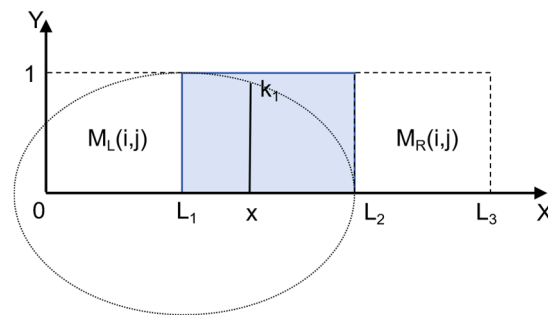

Figure 5. Weighted fusion based on elliptic function.

Let k_1 and k_2 denote the weights of the left and right images, respectively; then, the calculation method is as follows:

$$k_1 = \sqrt[2]{1 - \frac{(x - L_1)^2}{(L_2 - L_1)^2}}, \ x \in (L_1, L_2),$$ (6)

and

$$k_2 = 1 - k_1.$$ (7)

Let $M(i, j)$ represent the pixel point of the fused panoramic image; then, the calculation formula is as follows:

$$M(i, j) = \begin{cases} M_L(i, j), \ i < L_1, \\ k_1(i) \times M_L(i, j) + k_2(i) \times M_R(i, j), \ (i, j) \in M_L \cap M_R, \\ M_R(i, j), \ i > L_2. \end{cases}$$ (8)

2.2. Validation of Ship Panoramic Image Stitching Algorithm

2.2.1. Test Dataset

The image data used are a real image of the Xiamen sea scene taken by a shore-based camera. Due to the lack of panoramic images of real scenes, this paper crops two images with partially overlapping areas from a complete image and then compares the stitched and fused images with the original images for verification. Following the principles mentioned above, a test dataset is established as follows. Left and right images are cropped from

one image. The left image serves as the reference image, while the right image undergoes a slight stretching transformation to simulate differences between different cameras. A portion of the maximum rectangular sub-image extracted from this is selected as the image to be stitched, as shown in Figure 6. The size of the experimental images is constrained within the range of 640×470–680×540 pixels.
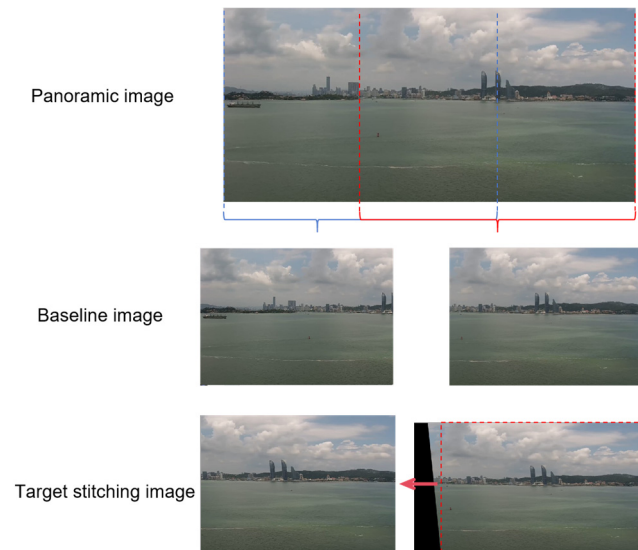

Figure 6. Making test datasets.

2.2.2. Verification Results

Comparison of Theoretical Data

To verify the effectiveness of the SSIM-EW algorithm proposed in this paper, the algorithm is compared with a cross-combination of the SIFT-based image registration method, the SURF-based image registration method, the average weighted fusion (AWF) method, and the fade-in/fade-out fusion method (DFF). The higher the evaluation index of Peak Signal-to-Noise Ratio (PSNR) [24], the higher the quality of the stitching. The results are shown in Table 1.

Table 1. Validation results of stitching algorithm comparison.

Serial Number	Stitching Method	PSNR	Time (s)	AP_Time (s)
①	SIFT + AWF	25.382	0.471	0.017
②	SIFT + DFF	25.386	0.517	0.017
③	SURF + AWF	24.690	0.436	0.021
④	SURF + DFF	24.679	0.485	0.021
⑤	SSIM-EW	25.736	0.435	0.020

In the table, "Time" represents the complete runtime of the stitching algorithm, and "AP_time" indicates the actual application stitching time. In practical applications, low-visibility weather such as rainy and foggy days will seriously affect the accuracy of the algorithm; after fixing the camera parameters, the precomputed homography matrix H and fusion weight matrix can be directly used for panoramic image stitching without needing to run the entire image stitching algorithm, which can reduce the algorithm running time and the adverse effects of bad weather on the stitching effect.

It can be seen from Table 1 that compared with other algorithms, the SSIM-EW algorithm proposed in this paper achieves the best performance in terms of both PSNR and time, and although it is not the best in terms of AP_time, the algorithm also meets the real-time requirements of practical applications. Therefore, the performance of SSIM-EW is the best.

Comparison of Fused Images

The image stitching results of different algorithms are shown in Figure 7. By comparing the locally enlarged images, it can be observed that there are still noticeable stitching seams and severe ghosting in the stitching results of SIFT + AWF and SURF + AWF. In the stitching results of SIFT + DFF and SURF + DFF, there are no obvious stitching seams, but slight ghosting still exists. The SSIM-EW proposed in this paper shows superior stitching results compared to other algorithms, with no apparent stitching seams or ghosting. The image stitching effect is significantly better than that of other algorithms.
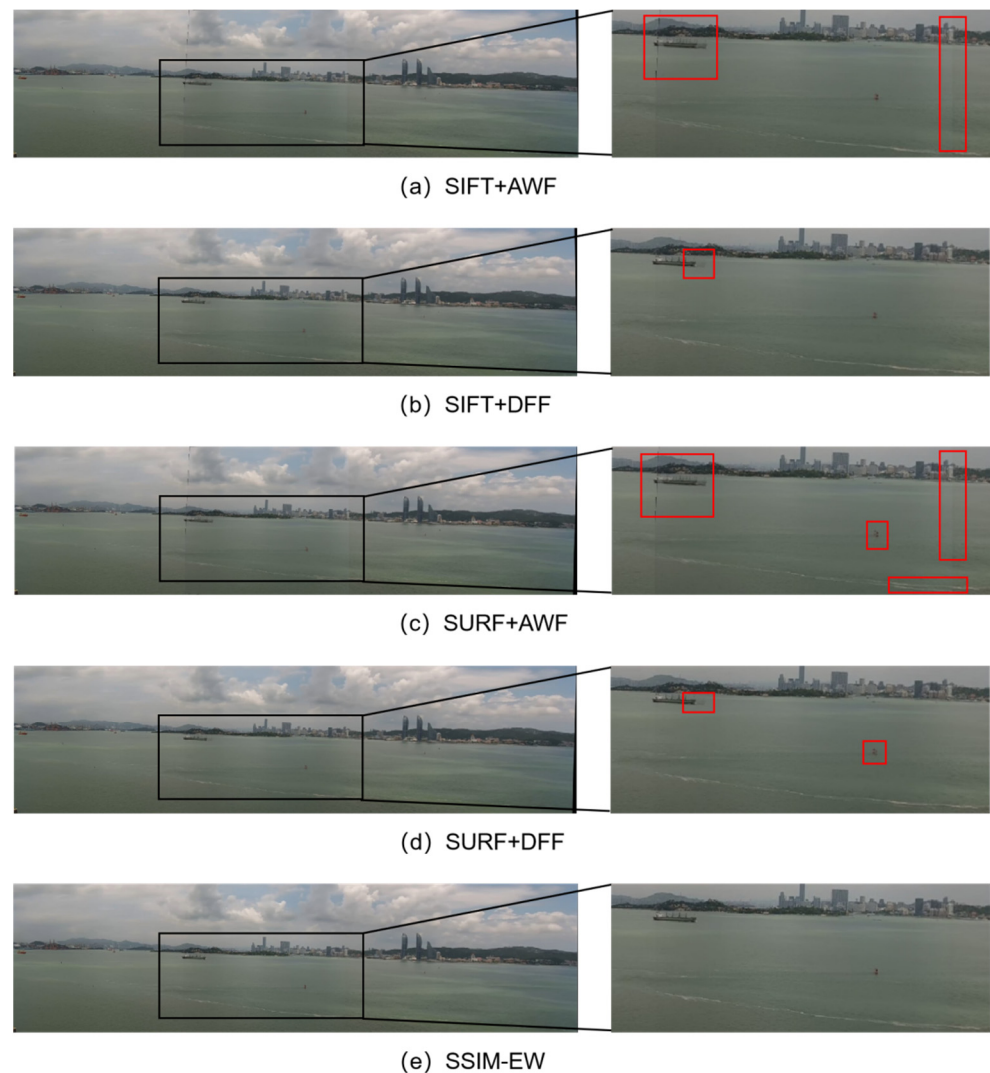

(a) SIFT+AWF

(b) SIFT+DFF

(c) SURF+AWF

(d) SURF+DFF

(e) SSIM-EW

Figure 7. Comparison of the results of different stitching algorithms. Inside this red frame are obvious flaw of picture stitching.

3. Construction of Ship Navigation Target Perception Model Based on YOLOv8

3.1. Model Architecture

As the latest version of the YOLO series of models, the YOLOv8 model further improves the performance, and the detection capability is faster, more accurate, and easier to apply [25]. Based on the YOLOv8s network, this paper optimizes the network structure and regression loss function, proposes the YOLOv8-SGW network, and strengthens the detection capacity of small maritime targets. The optimized YOLOv8-SGW network structure is shown in Figure 8.
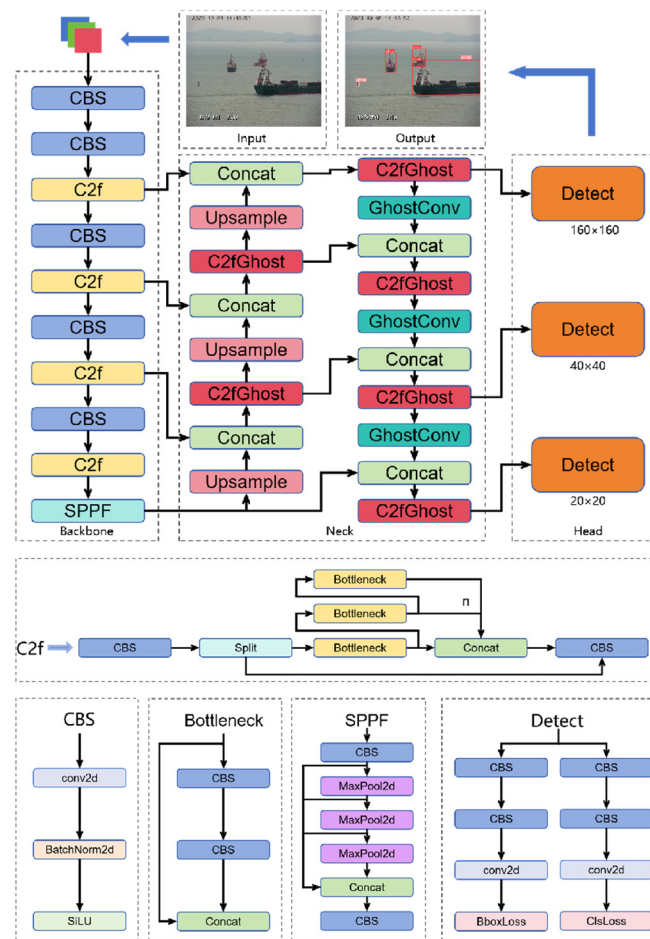
Figure 8. YOLOv8-SGW network architecture for target perception in a ship navigation environment. (1) represents the improvement in increasing the detection head size, while (2) represents the introduction of GhostNet. For more detailed information, please refer to Section 3.2.

3.2. Model Network Optimization

In the ship environment perception scene, in addition to the normal ship target, there are many small targets (such as buoys and reefs). To solve the problems of missed detection and wrong detection in the YOLOv8 model's small-target detection, this research improves the neck and head network of the YOLOv8 model, strengthens the small-target detection ability of the model, introduces lightweight GhostNet [26] in the neck network, strengthens the feature fusion ability, and ensures the lightweight characteristic of the model.

3.2.1. Increased Detection Head Size

The traditional YOLOv8s has three detection heads with sizes of 80×80, 40×40, and 20×20. Larger detection head sizes can extract target features from higher network layers, enhancing the perception capabilities for small targets. The original head layer does not have residual connections with higher network layers, leading to the loss of some target details in the feature maps generated by the head layer. To further improve the model's ability to detect small targets, the feature maps from the neck section are connected as residuals to even higher network layers, and the detection head size is increased from 80×80 to 160×160, as shown in Figure 9.
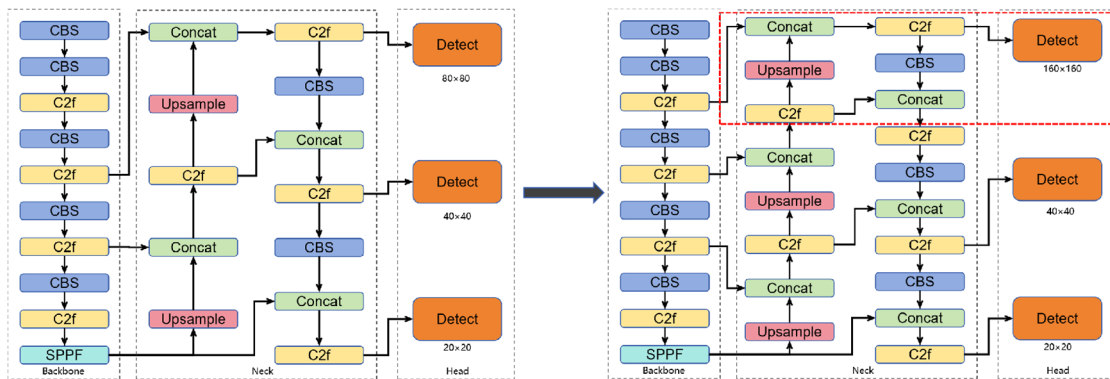
Figure 9. Changes in the model network structure after increasing the size of the detection head. Inside this red frame is the extra added network layers.

3.2.2. Introduction of GhostNet

Increasing the size of the detection head deepens the network layers and enhances the network's feature extraction capabilities. However, it also significantly increases the network parameters and computational complexity. To reduce the network parameters and computational complexity, the GhostConv module is introduced in the neck network. The GhostConv module uses some special linear transformations to generate partial output features. Compared to traditional convolution operations, the computational complexity of the GhostConv module is low while still preserving the feature information for the target objects obtained after traditional convolutions. The comparison between the CBS module and the GhostConv module is shown in Figure 10.
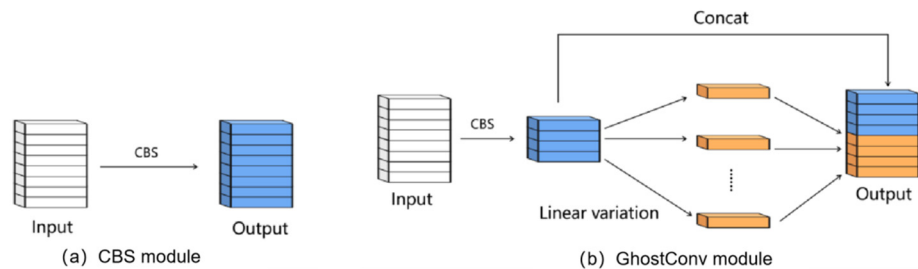

Figure 10. CBS module and GhostConv module.

The change in the model network structure after introducing the GhostNet network is shown in Figure 11. The CBS module in the C2f module is replaced with the GhostConv module, the C2fGhost module is formed, and then, all Conv and C2f modules in the neck network are replaced with GhostConv and C2fGhost modules, and the feature information extraction capacity is retained while reducing the number of network parameters.
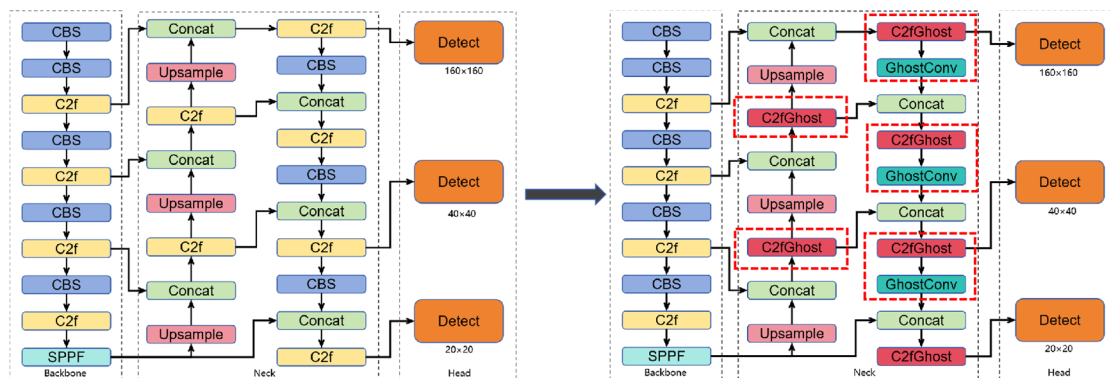

Figure 11. Changes in network structure after introducing GhostNet.

3.3. Improvement of Regression Loss Function

YOLOv8 adopts Complete-IoU (CIoU) loss [27] as the bounding box regression loss function. CIoU loss focuses on enhancing the fitting ability of bounding box losses, but it does not consider the detrimental impact of low-quality data on model performance. Addressing this flaw, in this paper, CIoU loss is replaced with Wise-IoU (WIoU) loss [28]. WIoU loss can adapt to distinguish between annotation box qualities, enabling the model to dynamically focus on different quality anchor boxes, reducing the negative impact of low-quality data on the model and thus improving the overall performance of the model. In Figure 12, the blue rectangle represents the predicted box, and the green rectangle represents the annotated box, where w_g and h_g denote the width and height of the smallest bounding box, respectively.
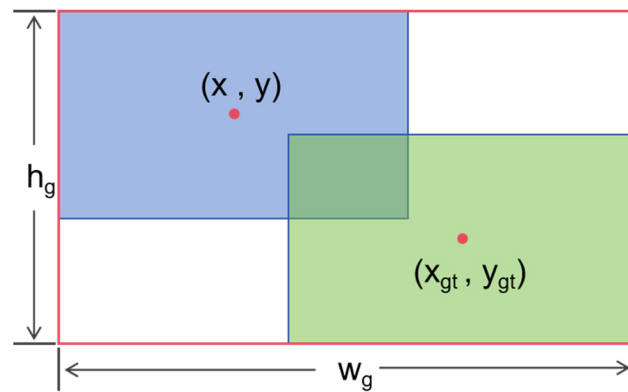

Figure 12. Schematic diagram of prediction box and callout box.

WIoU loss is calculated as follows:

$$\mathcal{L}_{WIoU} = \frac{\beta}{\delta \alpha^{\beta - \delta}} \mathcal{R}_{|WIoU} \mathcal{L}_{IoU}, \tag{9}$$

where δ and α are hyperparameters, which are taken as 1.9 and 3, respectively, in this research, and $\mathcal{R}_{WIoU}$, $\mathcal{L}_{IoU}$, and β are calculated as follows:

$$\mathcal{R}_{WIoU} = exp\left(\frac{(x - x_{gt})^2 + (y - y_{gt})^2}{\left(w_g^2 + h_g^2\right)}\right), \tag{10}$$

$$\mathcal{L}_{IoU} = 1 - IoU, \tag{11}$$

and

$$\beta = \frac{\mathcal{L}_{IoU}^*}{\overline{\mathcal{L}_{IoU}}}, \tag{12}$$

where *IoU* represents the intersection–union ratio between the predicted box and the annotated box. $\overline{\mathcal{L}_{IoU}}$ is the mean dynamic intersection over union loss of the predicted box, $\mathcal{L}_{IoU}^*$ is the current prediction's intersection over the union loss value, and β represents the outlier score of the predicted box. A lower outlier score indicates a higher quality for the predicted box.

3.4. Verification of Perceptual Models

3.4.1. Experimental Dataset

Based on the visual terminal deployed on the Xiamen coast, a database of marine targets is constructed, and 2107 marine target images with annotations are obtained through manual screening and annotation. There are two main types of maritime targets: 'Ship' and 'Other'. The dataset is divided into a training set, a validation set, and a test set in the ratio of 6:2:2. A partial dataset is shown in Figure 13.
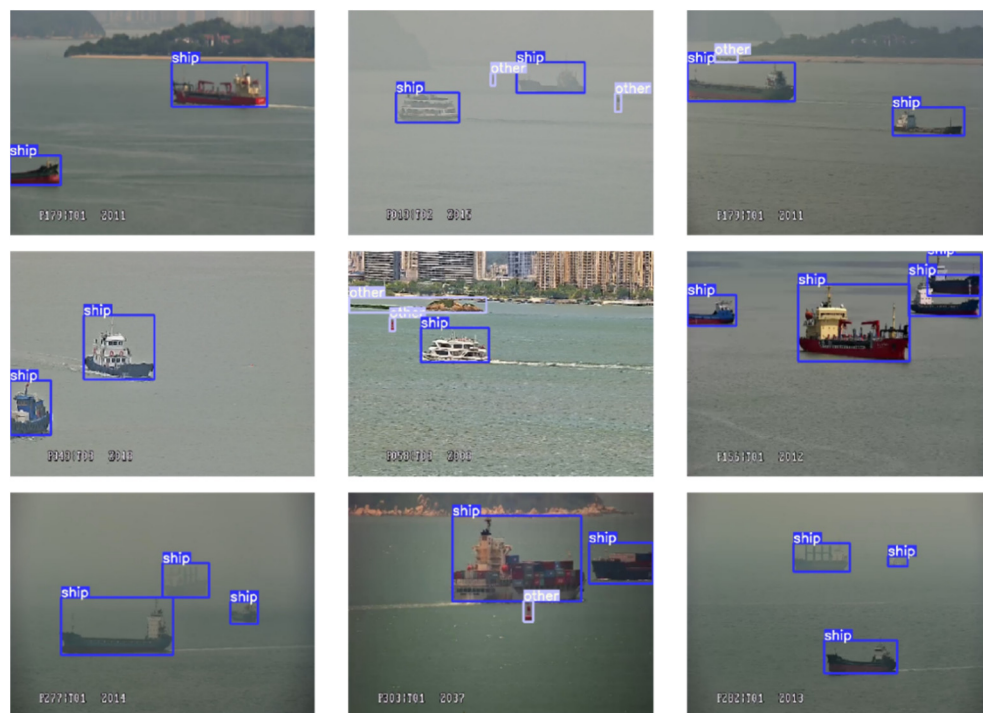
Figure 13. Partial dataset display.

3.4.2. Experiment

To verify the effectiveness of the improved model, ablation experiments are set to verify the impact of the improved method. The Mean Average Precision (mAP) [29] and frames per second (FPS) [30,31] are introduced to evaluate the performance of the model. The higher the mAP is, the stronger the ability of the model to perceive the target is. The experimental results are shown in Table 2.

Table 2. Ablation experiments. In the table, "SGnet" represents the model network optimization improvement, while "WIoU" represents the model regression loss function enhancement.

Serial Number	Model	P (%)	R (%)	mAP50 (%)	FPS
①	Baseline	84.6	75.2	82.9	126
②	Baseline + SGnet	82.7	79.5	84.8	110
③	Baseline + WIoU	86.7	75.3	83.3	126
④	Baseline + SGnet + WIoU	86.1	79.5	85.2	110

The specific conclusions are as follows:

(1) By introducing GhostNet and increasing the size of the detection head to improve the model network and enhance the small-target perception ability, the recall rate and mAP are increased by 4.3% and 1.9%, respectively.

(2) The original regression loss function is replaced by WIoU loss to reduce the harm of low-quality data. The precision rate, recall rate, and mAP are increased by 2.3%, 0.1%, and 0.4%, respectively.

(3) The proposed model ④ in this paper has demonstrated significant improvements in P, R, and mAP compared to the baseline model, with increases of 1.5%, 4.3%, and 2.3%, respectively. Although there is a slight decrease in FPS, achieving a detection speed of 110 FPS is still sufficient to meet the real-time perception requirements for marine applications, considering that the transmission speed of the shipboard camera is only 25 frames per second. Therefore, the improved model proposed in this paper has significantly enhanced detection accuracy without compromising real-time performance.

Some of the target detection results of the improved model in this paper are shown in Figure 14. In the figure, small targets such as ship targets and buoys can be detected without false detection or missed detection.

Figure 14. Partial target detection results.

To demonstrate the superiority of the YOLOv8-SGW network, this paper selects YOLOv5s, YOLOv7-tiny, YOLOv8s, and YOLOv8-SGW for comparison experiments. The experimental results are presented in Table 3. In terms of mAP50, YOLOv8-SGW achieved the highest accuracy, outperforming YOLOv5s, YOLOv7-tiny, and YOLOv8s by 1.6%, 3.9%, and 2.3%, respectively. In terms of the AP for "ship", YOLOv8-SGW falls short of the best performer, YOLOv5s, by 0.2%. However, when it comes to the AP for "Other", YOLOv8-SGW significantly outperforms the other models, achieving improvements of 3.4%, 6.7%, and 4.7%, respectively. Overall, the improved YOLOv8-SGW model proposed in this paper demonstrates the best performance on this dataset, effectively enhancing the detection capabilities for small objects such as buoys and reefs.

Table 3. Comparison of AP among different networks on the dataset.

	Ship's AP (%)	Other's AP (%)	mAP50 (%)
YOLOv5s	92.3	75.0	83.6
YOLOv7-tiny	90.8	71.7	81.3
YOLOv8s	92.0	73.7	82.9
YOLOv8-SGW	92.1	78.4	85.2

4. Application Verification of Tugboat Panoramic Visual Perception-Assisted Navigation

4.1. Construction of Application Platform

Taking the tugboat navigation and operation environment as the verification scene, to solve the limitations of current perception methods such as AIS, radar, and crew lookout, the tugboat panoramic visual perception assistant navigation system is constructed, and the image mosaic algorithm SSIM-EW is used to remove the visual redundant information and generate the tugboat panoramic visual perception scene. The improved YOLO perception model is used to achieve the recognition of small targets in the navigation environment such as distant ships and buoys, and the data for the electronic chart, AIS, and radar are fused to achieve the multi-dimensional perception of the navigation environment in the sea area to provide more comprehensive and accurate information support for ship navigation and improve navigation safety.

In practical applications, the images of multiple cameras are processed by perceived algorithms and image stitching algorithms at the same time, instead of performing panoramic image stitching first and then obstacle recognition. The perception algorithm's performance on the source image will be better; the perception results can be displayed on the source image or panoramic image according to the needs of the crew.

The camera perception data are transmitted through the IP network, the radar and Electronic Chart Display and Information System (ECDIS) data are transmitted through the NEMA0183 interface, and the ship AIS information is transmitted through the AIS base station. All the above information is transmitted to the data collection terminal at the ship

end through the gateway for preliminary data cleaning work. The 5G network is used as the ship-end network communication method to ensure the safe transmission of data in the ship-end local area network and the cloud server and realize the remote access, backup, etc., of the data. The architecture of the above tug panoramic visual perception-assisted navigation system is shown in Figure 15.
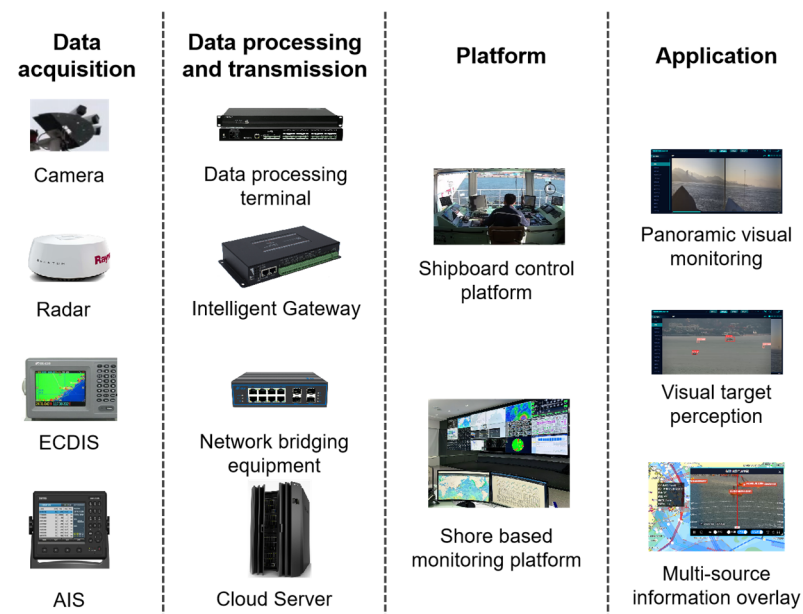

Figure 15. System architecture diagram.

4.2. Deployment of Relevant Equipment

The visual perception equipment is installed on a 3720 kW new energy hybrid tugboat. This tugboat is mainly used to assist with the docking and departure of large ships entering the port, and it can complete towing, pushing, and other operations in the port. The operational sea area is 24.498311° N–24.501598° N and 118.069349° E–118.075789° E (Figure 16).

Figure 16. Tugboat operation water area.

The navigation and navigation aids on the tugboat mainly include radar, an ECDIS, a magnetic compass, a GPS, a depth sounder, an AIS, and an anemometer. For specific ship parameters, please refer to Table 4.

Table 4. Tugboat parameters.

Ship's Name	XIA GANG TUO 30
MMSI	413545920
LOA	38.35 m (excluding bow and stern fenders)
Breadth	10.60 m
Depth	4.90 m
Designed Draft	3.70 m
Speed	>13 kn
Bollard Pull	Forward > 60 tons; astern > 55 tons
Crew Capacity	8 people

The installation distribution of visual perception equipment according to the perception requirements of the ship environment is shown in Figure 17. The forward-looking camera group is installed directly in front of the compass deck, approximately 1.5 m away from the compass deck, and the rear-looking camera group is installed on the mast on the compass deck, about 2.4 m above the compass deck. The vertical field of view of the panoramic camera group is approximately 30°, which can carry out panoramic intelligent monitoring for 360° around the hull. The structure of the hull itself should not enter the field of view as far as possible to avoid blocking the camera's field of view.
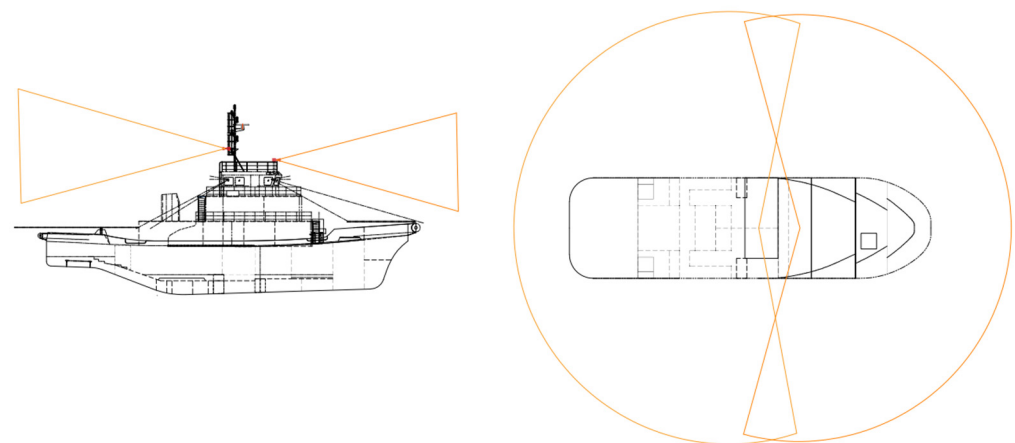

Figure 17. Installation design of sensing device.

4.3. Conversion of Different Data Coordinate Systems

Radar, AIS, electronic charts, and video data have different spatial information benchmarks. AIS data and electronic chart data are typically based on the WGS84 coordinate system. Radar data, on the other hand, often use a polar coordinate system centered around the radar base station. Video data are represented in pixel coordinates. The global coordinate system is established as a unified geographic reference, and the coordinate conversion is adopted to unify the coordinate space benchmark to realize the superposition and display of data from different dimensions.

4.3.1. Global Coordinate System and Inertial Coordinate System

As shown in Figure 18, taking a certain point (X_{lon0}, X_{lat0}) as the original center, the y axis points due north, and the right-hand Cartesian coordinate system is established as the global coordinate system as a unified geographic reference system. The corresponding inertial coordinate system is established with the tug as the original center. The transformation of the inertial coordinate system and the global coordinate system requires only a simple translation.
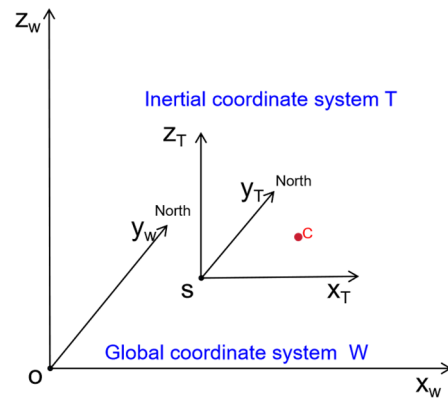
Figure 18. Global coordinate system and inertial coordinate system.

The conversion relationship between the inertial coordinate system and the global coordinate system is shown in Formula (13). (x_S, y_S, z_S) are the coordinates of point S in the global coordinate system.

$$\begin{bmatrix} x_W \\ y_W \\ z_W \end{bmatrix} = \begin{bmatrix} x_S \\ y_S \\ z_S \end{bmatrix} + \begin{bmatrix} x_T \\ y_T \\ z_T \end{bmatrix}. \tag{13}$$

4.3.2. Conversion of the Global Coordinate System and the Pixel Coordinate System

(1) The inertial coordinate system turns into the camera coordinate system

In practical application scenarios, a change in ship heading, rolling, and pitching will make the relationship between the camera coordinate system and inertial coordinate system change [32]. The pitching angle θ, rolling angle β, and heading angle φ are used to transform the inertial coordinate system into the camera coordinate system. The rotation of the inertial coordinate system around the x, y, and z axes can realize the transformation of the camera coordinate system (as shown in Figure 19), and the relationship is shown in Formula (14).

$$\begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = \begin{bmatrix} \cos\varphi & -\sin\varphi & 0 \\ \sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x_T \\ y_T \\ z_T \end{bmatrix}. \tag{14}$$
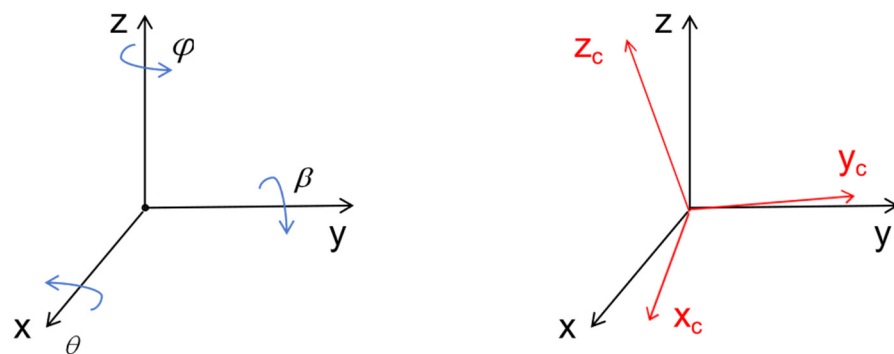

Figure 19. Rotation of inertial coordinates.

(2) Camera coordinate system turns into the image coordinate system

The camera coordinate system and image coordinate system are conversions between 3D and 2D, belonging to the perspective projection relationship. At this time, the unit of projection point is still the distance unit mm, not the pixel. The image coordinates

are calculated using Formula (15); z_c represents the distance between the point and the optical axis.

$$z_c \begin{bmatrix} x_I \\ y_I \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix}. \tag{15}$$

(3) Image coordinate system turns into pixel coordinate system

The pixel coordinate system and image coordinate system are both on the imaging plane, but their respective origins and measurement units are different (as shown in Figure 20). The unit of the image coordinate system is mm, which belongs to the physical unit, while the unit of the pixel coordinate system is pixel. According to Formula (16), the pixel coordinates corresponding to the image coordinates are calculated, where dx and dy represent how many millimeters each column and row represents, i.e., 1 pixel = dx mm.

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{dx} & 0 & u_0 \\ 0 & \frac{1}{dy} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_I \\ y_I \\ 1 \end{bmatrix}. \tag{16}$$
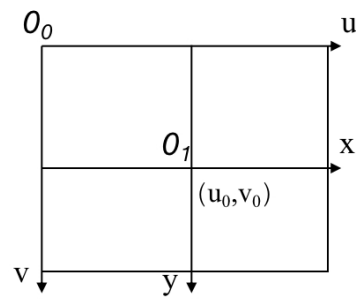

Figure 20. Pixel coordinate system.

In summary, by using the conversion method for the pixel coordinate system, camera coordinate system, inertial coordinate system, and global coordinate system, the mutual mapping relationship between the pixel coordinate system and global coordinate system is determined, as shown in Formula (17), where F and Q represent the camera's inner and outer parameter matrices, respectively. F, Q, and z_c can be obtained through camera calibration.

$$z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = FQ \left(\begin{bmatrix} x_W \\ y_W \\ z_W \end{bmatrix} - \begin{bmatrix} x_S \\ y_S \\ z_S \end{bmatrix} \right), \tag{17}$$

where

$$F = \begin{bmatrix} \frac{1}{dx} & 0 & u_0 \\ 0 & \frac{1}{dy} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f_z & 0 & 0 \\ 0 & f_z & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{18}$$

and

$$Q = \begin{bmatrix} \cos\varphi & -\sin\varphi & 0 \\ \sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix}. \tag{19}$$

By utilizing the mutual mapping relationship between different coordinate systems, AISs, radar, and visual data can be overlaid on charts and videos.

4.4. Application of Visual Assistance for Perception

4.4.1. Panoramic Vision-Aided Operation

Using the SSIM-EW algorithm proposed in this research, the visual redundancy between images is eliminated, the different perspective images taken by the camera are

combined in real time, and the continuous and seamless panoramic perception image of the sea area around the tugboat is output, which provides a larger field of vision, reduces the frequency of view switching when the crew looks outward, effectively avoids collisions between the tugboat and other ships, obstacles, and dangerous areas, and improves the tugboat's operation efficiency. Combined with other perception modules, the panoramic vision technology can intuitively monitor the changes in the surrounding environment of the tugboat to ensure the safety and efficiency of the operation (as shown in Figure 21).
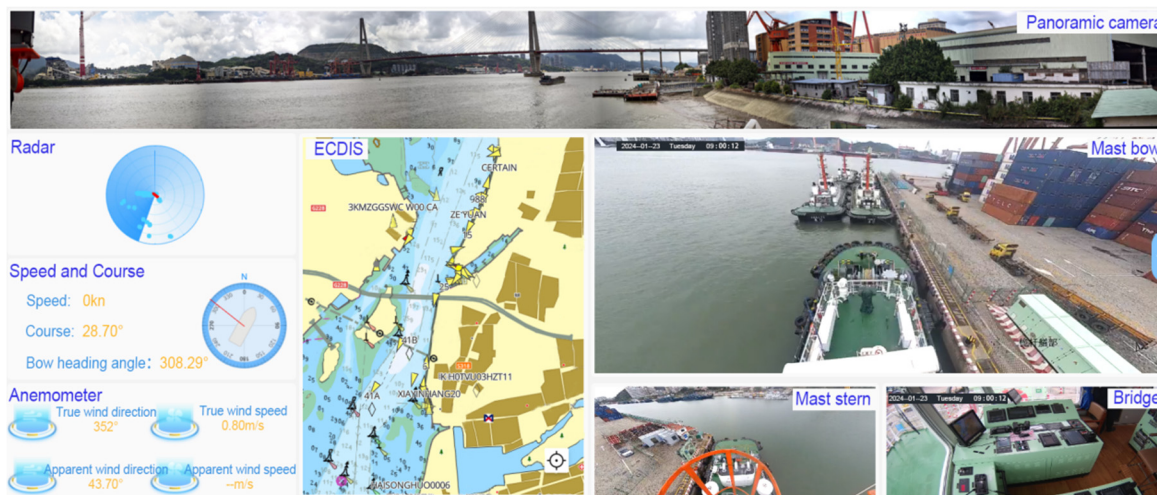

Figure 21. Panoramic vision auxiliary operation. Due to the vignetting of the camera lenses, the overlapping areas of panoramic images may exhibit local dark phenomena.

4.4.2. Visual Assistance for Perception

According to the relevant regulations of Xiamen Maritime Bureau, when the visibility distance is less than 1000 m, ships should stop in time. Therefore, this paper mainly focuses on the situation where the visibility distance is greater than 1000 m. When the visual perception system of the ship is affected by the visibility, in addition to using defogging algorithms such as dark channel enhancement and Retinex to enhance the visual perception ability of the ship, radar and AISs can be used to assist in the perception of the ship.

AISs, radar, and LiDAR are common target recognition and monitoring systems, but their perception range and information provided have certain limitations. AISs can provide basic information such as the position, course, and speed of a ship but only if these ships carry and turn on the AIS device; otherwise, this information cannot be obtained, and such a perception mode is too passive. And the update frequency of AISs is too slow to meet the real-time performance. Radar can detect the surrounding objects, but it only provides point targets and cannot provide the appearance, color, and other details of a ship [33,34]. In addition, the detection ability of radar for small objects, low-energy reflective objects, or especially stationary objects is poor [35–37]. LiDAR can provide high ranging accuracy and angle resolution in real time at a short and medium distance [38] and has strong anti-magnetic interference ability, but it has high cost, a small detection range, and is easily affected by adverse climate.

Supplementing visual perception methods on a ship can help to make up for the lack of AIS and radar perception and improve the navigation safety of the ship. The following are some of the visually aided perception application scenarios:

(1) The AIS, radar target, and visible light images are superimposed on ECDIS, which can provide more intuitive information for the crew [39]. The crew can intuitively understand the position, appearance, and surrounding environment of the target by observing the superimposed image to better judge the intention and possible action of the target, as shown in Figure 22.
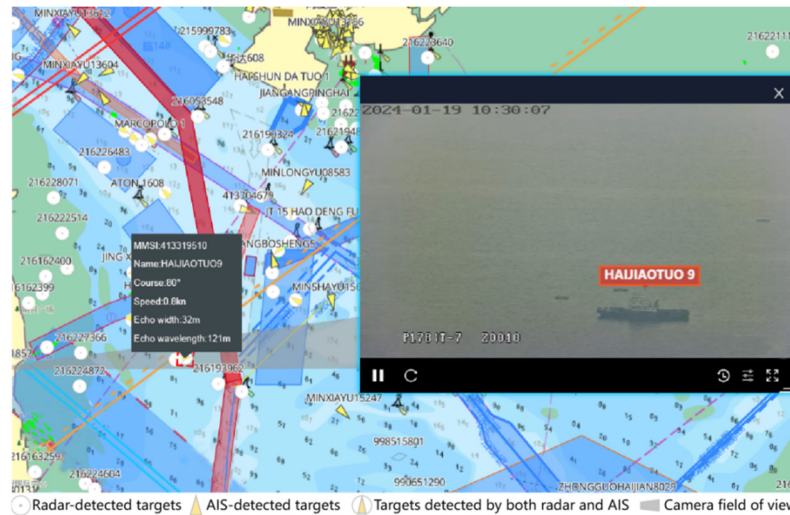
Radar-detected targets ▲ AIS-detected targets ⬭ Targets detected by both radar and AIS ▬ Camera field of view

Figure 22. Multi-source information overlay display.

(2) As shown in Figure 23, the visual perception model identifies buoys without AIS information. The AIS information for a ship with an MMSI of 312446303 seriously lags behind, and the position displayed on the ECDIS is wrong. The visual perception model helps to perceive the correct position of the ship.

Figure 23. AIS information exception.

(3) The visual perception model identifies two stationary boats next to the ship *HAI-JIAOTUO 9* (Figure 24) that are not sensed by the radar and are not equipped with (turned on) AIS equipment.
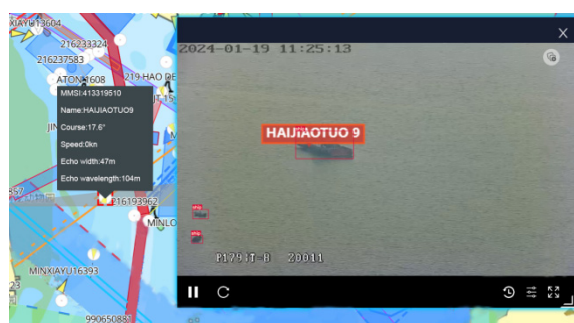

Figure 24. Radar information missing.

(4) As shown in Figure 25, the visual perception model and the radar detect the reef in front of the ship, and the visual assistant ranging function is turned on to determine the distance between the obstacle and the ship.
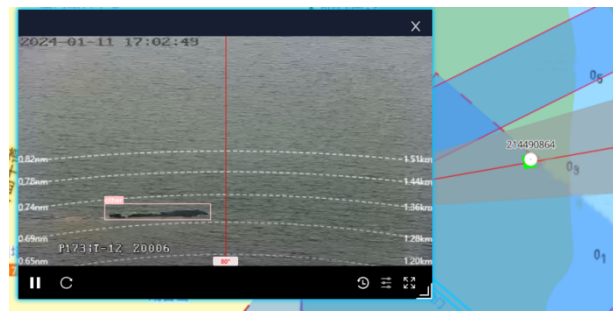
Figure 25. Obstacle distance location.

At present, this visual perception model can only identify targets as "Ship" and "Other". In future research, these categories should be further refined into specific types of maritime targets, such as passenger ship, tugboat, cargo ship, buoy, reef, and so on. Additionally, the current approach of the system for handling different data is limited to a simple overlay display on the same layer, requiring sailors to manually determine whether the AIS data, radar data, and visual perception data belong to the same target. This is an area that requires further research and refinement. One potential solution is to utilize a data matching algorithm and a data fusion algorithm to achieve the integration of different data sources for the same target.

5. Conclusions

This paper proposes a ship environment auxiliary perception technology based on panoramic vision that breaks through the limitations of traditional ship perception technology and promotes the development of ship intelligence. The specific research contents are as follows:

(1) To solve the problems of the instability of the ship panoramic image registration and the unnatural stitching effect, the traditional image stitching method is improved by the region-of-interest search method based on SSIM and the elliptical weighted fusion method. The panoramic image stitched by the improved SSIM-EW algorithm has a significant improvement in the quality evaluation index, and the transition is natural without stitching traces such as ghosting and dislocation. The experimental results show that the splicing effect of the SSMI-EW algorithm is significantly better than that of other algorithms.

(2) To solve the problem of poor perception accuracy of small targets at sea, based on the YOLOv8 model, this study increases the size of the detection head, introduces GhostNet, uses the WIoU function instead of the original loss function, and proposes the YOLOv8-SGW model. The mAP of the YOLOv8-SGW model is increased by 2.3%, its detection accuracy is significantly higher than that of YOLOv5s, YOLOv7-tiny, and YOLOv8, and the detection ability of small targets is greatly improved.

(3) The above technology is applied to the tugboat, and the panoramic vision-assisted tugboat operation is successfully realized, which expands the perception field of view. Moreover, the visual perception model can detect the targets that AIS and radar cannot, which enhances the perception ability of ships and has great significance and practical application value for the safe navigation of ships and the development of ship intelligence.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: If code and datasets are needed, please contact the corresponding author; they will be available for free.

Conflicts of Interest: Liangqing Guan and Zhiqiang Luo were employed by the Fujian Fuchuan Marine Engineering Technology Research Institute Co., Ltd. Shengchao Zhang and Jianfeng Zhang were employed by the Xiamen Port Shipping Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Sèbe, M.; Scemama, P.; Choquet, A.; Jung, J.L.; Chircop, A.; Razouk, P.M.A.; Michel, S.; Stiger-Pouvreau, V.; Recuero-Virto, L. Maritime transportation: Let's slow down a bit. *Sci. Total Environ.* **2022**, *811*, 152262. [CrossRef] [PubMed]
2. Li, H. Research on Digital, Networked and Intelligent Manufacturing of Modern Ship. *J. Phys. Conf. Ser.* **2020**, *1634*, 012052. [CrossRef]
3. Yang, D.; Solihin, M.I.; Zhao, Y.; Yao, B.; Chen, C.; Cai, B.; Machmudah, A. A review of intelligent ship marine object detection based on RGB camera. *IET Image Proc.* **2024**, *18*, 281–297. [CrossRef]
4. Qu, J.; Liu, R.W.; Guo, Y.; Lu, Y.; Su, J.; Li, P. Improving maritime traffic surveillance in inland waterways using the robust fusion of AIS and visual data. *Ocean Eng.* **2023**, *275*, 114198. [CrossRef]
5. Chen, Z.; Chen, D.; Zhang, Y.; Cheng, X.; Zhang, M.; Wu, C. Deep learning for autonomous ship-oriented small ship detection. *Saf. Sci.* **2020**, *130*, 104812. [CrossRef]
6. Faggioni, N.; Ponzini, F.; Martelli, M. Multi-obstacle detection and tracking algorithms for the marine environment based on unsupervised learning. *Ocean Eng.* **2022**, *266*, 113034. [CrossRef]
7. Zhu, Q.; Ma, K.; Wang, Z.; Shi, P.B. YOLOv7-CSAW for maritime target detection. *Front. Neurorob.* **2023**, *17*, 1210470. [CrossRef] [PubMed]
8. Cheng, S.; Zhu, Y.; Wu, S. Deep learning based efficient ship detection from drone-captured images for maritime surveillance. *Ocean Eng.* **2023**, *285*, 115440. [CrossRef]
9. Adolphi, C.; Parry, D.D.; Li, Y.; Sosonkina, M.; Saglam, A.; Papelis, Y.E. LiDAR Buoy Detection for Autonomous Marine Vessel Using Pointnet Classification. In Proceedings of the Modeling, Simulation and Visualization Student Capstone Conference, Suffolk, VA, USA, 20 April 2023.
10. Hagen, I.B.; Brekke, E. In Kayak Tracking using a Direct Lidar Model. In Proceedings of the Global Oceans 2020: Singapore-US Gulf Coast, Biloxi, MS, USA, 5–30 October 2020; pp. 1–7.
11. Abbadi, N.K.E.L.; Al Hassani, S.A.; Abdulkhaleq, A.H. A Review Over Panoramic Image Stitching Techniques. *J. Phys. Conf. Ser.* **2021**, *1999*, 012115. [CrossRef]
12. Zhang, J.; Liu, T.; Yin, X.; Wang, X.; Zhang, K.; Xu, J.; Wang, D. An improved parking space recognition algorithm based on panoramic vision. *Multimed. Tools Appl.* **2021**, *80*, 18181–18209. [CrossRef]
13. Kinzig, C.; Cortés, I.; Fernández, C.; Lauer, M. Real-time seamless image stitching in autonomous driving. In Proceedings of the 2022 25th International Conference on Information Fusion (FUSION), Linköping, Sweden, 4–7 July 2022; pp. 1–8.
14. Zhu, H.; Yuen, K.V.; Mihaylova, L.; Leung, H. Overview of Environment Perception for Intelligent Vehicles. *IEEE Trans. Intell. Transp. Syst.* **2017**, *18*, 2584–2601. [CrossRef]
15. Taha, A.E.; AbuAli, N. Route Planning Considerations for Autonomous Vehicles. *IEEE Commun. Mag.* **2018**, *56*, 78–84. [CrossRef]
16. Martelli, M.; Virdis, A.; Gotta, A.; Cassarà, P.; Summa, M.D. An Outlook on the Future Marine Traffic Management System for Autonomous Ships. *IEEE Access* **2021**, *9*, 157316–157328. [CrossRef]
17. Wang, Z.; Yang, Z. Review on image-stitching techniques. *Multimed. Syst.* **2020**, *26*, 413–430. [CrossRef]
18. Wei, X.; Yan, W.; Zheng, Q.; Gu, M.; Su, K.; Yue, G.; Liu, Y. Image Redundancy Filtering for Panorama Stitching. *IEEE Access* **2020**, *8*, 209113–209126. [CrossRef]
19. Chang, C.-H.; Sato, Y.; Chuang, Y.-Y. Shape-preserving half-projective warps for image stitching. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 3254–3261.
20. Wang, G.; Liu, L.; Zhang, Y. Research on Scalable Real-Time Image Mosaic Technology Based on Improved SURF. *J. Phys. Conf. Ser.* **2018**, *1069*, 012162. [CrossRef]
21. Bakurov, I.; Buzzelli, M.; Schettini, R.; Castelli, M.; Vanneschi, L. Structural similarity index (SSIM) revisited: A data-driven approach. *Expert Syst. Appl.* **2022**, *189*, 116087. [CrossRef]

22. Fischler, M.A.; Bolles, R.C. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM* **1981**, *24*, 381–395. [CrossRef]
23. Lowe, D.G. Distinctive Image Features from Scale-Invariant Keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [CrossRef]
24. Hore, A.; Ziou, D. Image quality metrics: PSNR vs. SSIM. In Proceedings of the 2010 20th International Conference on Pattern Recognition, Istanbul, Turkey, 23–26 August 2010; pp. 2366–2369.
25. Terven, J.; Córdova-Esparza, D.-M.; Romero-González, J.-A. A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS. *Mach. Learn. Knowl. Extr.* **2023**, *5*, 1680–1716. [CrossRef]
26. Han, K.; Wang, Y.; Tian, Q.; Guo, J.; Xu, C.; Xu, C. Ghostnet: More features from cheap operations. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 1580–1589.
27. Zheng, Z.; Wang, P.; Ren, D.; Liu, W.; Ye, R.; Hu, Q.; Zuo, W. Enhancing Geometric Factors in Model Learning and Inference for Object Detection and Instance Segmentation. *IEEE Trans. Cybern.* **2022**, *52*, 8574–8586. [CrossRef] [PubMed]
28. Tong, Z.; Chen, Y.; Xu, Z.; Yu, R. Wise-IoU: Bounding Box Regression Loss with Dynamic Focusing Mechanism. *arXiv* **2023**, arXiv:2301.10051.
29. Bucak, S.S.; Jin, R.; Jain, A.K. Multiple Kernel Learning for Visual Object Recognition: A Review. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *36*, 1354–1369. [PubMed]
30. Amir, S.; Siddiqui, A.A.; Ahmed, N.; Chowdhry, B.S. Implementation of line tracking algorithm using Raspberry pi in marine environment. In Proceedings of the 2014 IEEE International Conference on Industrial Engineering and Engineering Management, Selangor, Malaysia, 9–12 December 2014; pp. 1337–1341.
31. Jaszewski, M.; Parameswaran, S.; Hallenborg, E.; Bagnall, B. Evaluation of maritime object detection methods for full motion video applications using the pascal voc challenge framework. In Proceedings of the Video Surveillance and Transportation Imaging Applications, San Francisco, CA, USA, 8–12 February 2015; pp. 298–304.
32. Bäumker, M.; Heimes, F. New calibration and computing method for direct georeferencing of image and scanner data using the position and angular data of an hybrid inertial navigation system. In Proceedings of the OEEPE Workshop, Integrated Sensor Orientation, Hanover, Germany, 17–18 September 2001; pp. 1–17.
33. Goudossis, A.; Katsikas, S.K. Towards a secure automatic identification system (AIS). *J. Mar. Sci. Technol.* **2019**, *24*, 410–423. [CrossRef]
34. Karataş, G.B.; Karagoz, P.; Ayran, O. Trajectory pattern extraction and anomaly detection for maritime vessels. *Internet Things* **2021**, *16*, 100436. [CrossRef]
35. Lazarowska, A. Review of Collision Avoidance and Path Planning Methods for Ships Utilizing Radar Remote Sensing. *Remote Sens.* **2021**, *13*, 3265. [CrossRef]
36. Cheng, Y.; Xu, H.; Liu, Y. Robust small object detection on the water surface through fusion of camera and millimeter wave radar. In Proceedings of the IEEE International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021; pp. 15263–15272.
37. Clunie, T.; DeFilippo, M.; Sacarny, M.; Robinette, P. Development of a perception system for an autonomous surface vehicle using monocular camera, lidar, and marine radar. In Proceedings of the 2021 IEEE International Conference on Robotics and Automation (ICRA), Xi'an, China, 30 May–5 June 2021; pp. 14112–14119.
38. Paoletti, S.; Rumes, B.; Pierantonio, N.; Panigada, S.; Jan, R.; Folegot, T.; Schilling, A.; Riviere, N.; Carrier, V.; Dumoulin, A.J.R.I. SEADETECT: Developing an automated detection system to reduce whale-vessel collision risk. *Res. Ideas Outcomes* **2023**, *9*, e113968. [CrossRef]
39. Wu, Y.; Chu, X.; Deng, L.; Lei, J.; He, W.; Królczyk, G.; Li, Z. A new multi-sensor fusion approach for integrated ship motion perception in inland waterways. *Measurement* **2022**, *200*, 111630. [CrossRef]