

Article

# Ship Detection under Low-Visibility Weather Interference via an Ensemble Generative Adversarial Network

Xinqiang Chen <sup>1</sup> , Chenxin Wei <sup>1</sup>, Zhengang Xin <sup>1</sup> , Jiansen Zhao <sup>2,\*</sup>  and Jiangfeng Xian <sup>1</sup> 

<sup>1</sup> Institute of Logistics Science and Engineering, Shanghai Maritime University, Shanghai 201306, China; chenxinqiang@stu.shmtu.edu.cn (X.C.); 202130510052@stu.shmtu.edu.cn (C.W.); 202330510005@stu.shmtu.edu.cn (Z.X.); jfxian@shmtu.edu.cn (J.X.)

<sup>2</sup> Merchant Marine College, Shanghai Maritime University, Shanghai 201306, China

\* Correspondence: jszhao@shmtu.edu.cn

**Abstract:** Maritime ship detection plays a crucial role in smart ships and intelligent transportation systems. However, adverse maritime weather conditions, such as rain streak and fog, can significantly impair the performance of visual systems for maritime traffic. These factors constrain the performance of traffic monitoring systems and ship-detection algorithms for autonomous ship navigation, affecting maritime safety. The paper proposes an approach to resolve the problem by visually removing rain streaks and fog from images, achieving an integrated framework for accurate ship detection. Firstly, the paper employs an attention generation network within an adversarial neural network to focus on the distorted regions of the degraded images. The paper also utilizes a contextual encoder to infer contextual information within the distorted regions, enhancing the credibility of image restoration. Secondly, a weighted bidirectional feature pyramid network (BiFPN) is introduced to achieve rapid multi-scale feature fusion, enhancing the accuracy of maritime ship detection. The proposed GYB framework was validated using the SeaShip dataset. The experimental results show that the proposed framework achieves an average accuracy of 96.3%, a recall of 95.35%, and a harmonic mean of 95.85% in detecting maritime traffic ships under rain-streak and foggy-weather conditions. Moreover, the framework outperforms state-of-the-art ship detection methods in such challenging weather scenarios.



**Citation:** Chen, X.; Wei, C.; Xin, Z.; Zhao, J.; Xian, J. Ship Detection under Low-Visibility Weather Interference via an Ensemble Generative Adversarial Network. *J. Mar. Sci. Eng.* **2023**, *11*, 2065. <https://doi.org/10.3390/jmse11112065>

Academic Editor: Sergei Chernyi

Received: 6 October 2023

Revised: 24 October 2023

Accepted: 26 October 2023

Published: 29 October 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** ship detection; adverse weather; image restoration; improved YOLOv5; intelligent maritime transportation

## 1. Introduction

With the rapid advancement of artificial intelligence and computer vision technologies, the traditional navigation methods of maritime ships are undergoing a process of transformation and elevation. Intelligent maritime traffic monitoring systems and automated ship navigation are gradually becoming tangible realities (Liu et al., Cheng et al., Volden et al. [1–3]). Therefore, it is imperative to accurately detect maritime traffic entities (such as ships, buoys, etc.) based on ship vision navigation or port surveillance videos in order to make precise navigational control decisions. This plays a pivotal role in enhancing the safety of automated ship navigation and maritime ship passage within waterways (Forti et al. [4]). To detect maritime navigating ships, various types of sensors, such as cameras and radar, are commonly employed in automated ship navigation. However, the unique maritime weather conditions and locations often expose these systems to adverse weather, such as fog and rain streaks, resulting in the deterioration of ship-monitoring video data (Bahnsen and Moeslund, and Li et al. [5,6]). The presence of rain streaks and fog in the atmosphere severely impacts the visibility of monitoring scenes. Low visibility is inefficient for accurately detecting maritime ships and increases the risk of maritime traffic accidents. Consequently, the development of effective image restoration techniques

becomes crucial to achieve improved visual appearance or distinctive features. Providing clear maritime images to detection systems can significantly enhance the detection performance of maritime ships at sea (Fu et al., Lu et al. [7,8]).

Computer vision technology has become a crucial method for autonomous ship navigation (autonomous driving) and intelligent transportation applications. It can detect and recognize target objects in various scenarios with high precision, while also providing data support for intelligent control decision-making in the transportation field (Yu et al., Yao et al. [9,10]). Previous research has primarily focused on capturing high-quality maritime traffic video data. Wang et al. proposed a rapid and accurate ship detection algorithm based on YOLOv4, which incorporates K-means clustering, model structure refinement, and the Mixupfan method (Wang et al. [11]). Li et al. utilized a background filtering network for rapid filtering of background areas and employed a fine-grained ship classification network for the detection and classification of ship targets (Li et al. [12]). Fence et al. proposed a fast ship detection method based on multi-scale gradient features and a multi-branch support vector machine (Feng et al. [13]). Similar studies can be found in (Shao et al., Lv et al. and Chen et al. [14–16]). For some scenarios where video data cannot be directly obtained, previous research relied on limited exploration using radar data and multi-source data fusion. Chen et al. proposed a study similar to Radar-YOLONet that uses radar images for object detection (Chen et al. [17]). Wang et al. proposed a deep radar object-detection method called RODNet based on cross-fusion supervision of radar–camera data (Wang et al. [18]). Xu et al. used a multiple linear rescaling scheme to quantize the original satellite images into 8-b images, and proposed an adaptive weighting scheme to detect the loss between ships (Xu et al. [19]). Similar studies can be found in (Guo et al., Bai et al. [20,21]). The acquisition of multi-source data usually depends on special physical sensors, which are highly susceptible to the water environment and have high maintenance costs (Shang et al., Lin et al. [22,23]). With the development of deep learning, feature enhancement has been used to strengthen the perception of low-feature targets in low-visibility scenes. This addresses the issues of low accuracy and efficiency in traditional object-detection algorithms. Wang et al. constructed a new feature enhancement module (FEM) and utilized an attention mechanism to achieve real-time accurate detection of multiple targets in foggy conditions (Wang et al. [24]). M. Hassaballah et al. utilized an image enhancement scheme to achieve robust detection and tracking of vehicles (Hassaballah et al. [25]). While these methods can effectively detect target objects, they may not fully address the unique characteristics of maritime traffic environments, such as tides, water currents, and channel divisions. Therefore, they may not guarantee the safety of maritime traffic. It is important to consider these factors when developing and implementing object-detection algorithms for maritime traffic environments.

To address these problems, the paper presents an integrated framework for maritime ship detection under adverse weather conditions using computer vision techniques. This framework leverages adversarial neural networks to generate attention maps that focus on distorted regions within the images. These attention maps guide the contextual auto-encoder in performing local feature inference, achieving a rational and effective restoration of distorted areas in low-visibility images. Moreover, the restored images are concurrently fed into the discriminative network to facilitate the evaluation of the restored regions in the generated images. This process serves as feedback to guide the generative network in achieving optimal results for the enhancement of low-visibility images. Next, a weighted bidirectional feature pyramid network (BiFPN) is introduced to achieve rapid multi-scale feature fusion, enhancing the accuracy of maritime ship detection (Tan et al. [26]). This involves iteratively applying top-down and bottom-up multi-scale feature fusion to enhance the accuracy of ship detection in repaired low-visibility images. Our proposed framework is evaluated on the synthetic SeaShip dataset, which includes challenges related to low-visibility conditions such as rain streaks and fog, as well as small-target detection. Experimental results show that the model framework we proposed exhibits effectiveness and superiority over existing algorithms. The main contributions of this work are summarized as follows:

- The paper has proposed a novel integrated framework for detecting and recognizing ships navigating in low-visibility maritime environments.
- The paper has proposed the use of a weighted BiFPN in the YOLOv5 detector, achieving top-down and bottom-up multi-scale feature fusion to improve the accuracy of ship detection in low-energy image restoration.
- The paper's proposed framework achieves an average accuracy of 96.3%, a recall of 95.35%, and a harmonic mean of 95.85% in detecting maritime traffic ships under rain streak and foggy weather conditions.

## 2. Materials and Methods

The proposed framework for ship detection in low-visibility maritime images in this paper consists of two main logical steps: image restoration for maritime traffic and ship detection in maritime traffic, as shown in Figure 1. Firstly, an attention map is generated in the recurrent network within the generative network to identify low-visibility areas in the image that are disturbed by rain streaks and foggy weather. Meanwhile, the context autoencoder within the generative network performs local inference and restoration on the rain streaks and fog areas, enabling them to generate more realistic local images. More specifically, firstly, an attention map is generated in the recurrent network within the generative network to identify low-visibility areas in the image that are disturbed by rain streaks and foggy weather. Meanwhile, the context autoencoder within the generative network performs local inference and restoration on the rain streaks and fog areas, enabling them to generate more realistic local images. More specifically, firstly, images of rainy and foggy weather are input into the model framework. The images, after passing through the generative attention map network (Residual Block and LSTM + Convs modules), generate an attention map for the rain streaks and fog (low-visibility) areas of the two-dimensional image. This enhances the perceptibility of the distorted areas and provides guidance for subsequent image restoration. Secondly, the generated attention map and the original image are passed into the generative contextual autoencoder (Convs + ReLu, Dilated Convs + ReLu, and Deconv + avgpool + ReLu modules). This allows for the extraction of surrounding structure and feature information from the distorted areas. By combining these extracted features, contextual information is inferred and restored, resulting in the generation of relatively intact images. Meanwhile, the restored images are input into the discriminator for image quality assessment. Finally, the restored images are input into the detection model. Since some areas may have lower restoration quality during the image restoration process, a multi-scale fusion method is used to achieve detection and recognition of low-resolution ships. The discriminative network evaluates whether the images generated by the generative network are realistic, and provides feedback to the generative network. Next, the detection model incorporates BiFPN into YOLOv5 for multi-scale feature fusion, further enhancing the accuracy of ship detection in the low-visibility image restoration regions.

### 2.1. Rain-Streak and Fog Imaging Modeling

Rain streaks and fog in images can affect the detection performance of ships in both human and computer vision. Therefore, removing rain streaks and fog, which means restoring blurry images to clean images, is an important problem in computer vision. To better remove rain streaks and fog, we first mathematically model the rain streaks and foggy scenes. The widely used rain streaks model scene (Quan et al., Luo et al. and Li et al. [27–29]) modeling formula is shown in Equation (1):

$$C = F_{img} - S \quad (1)$$

where  $F_{img}$  represents the rain streak image,  $C$  represents the background of clean water transportation and  $S$  represents the rain streaks. Therefore, we need to remove the rain streaks  $S$ .

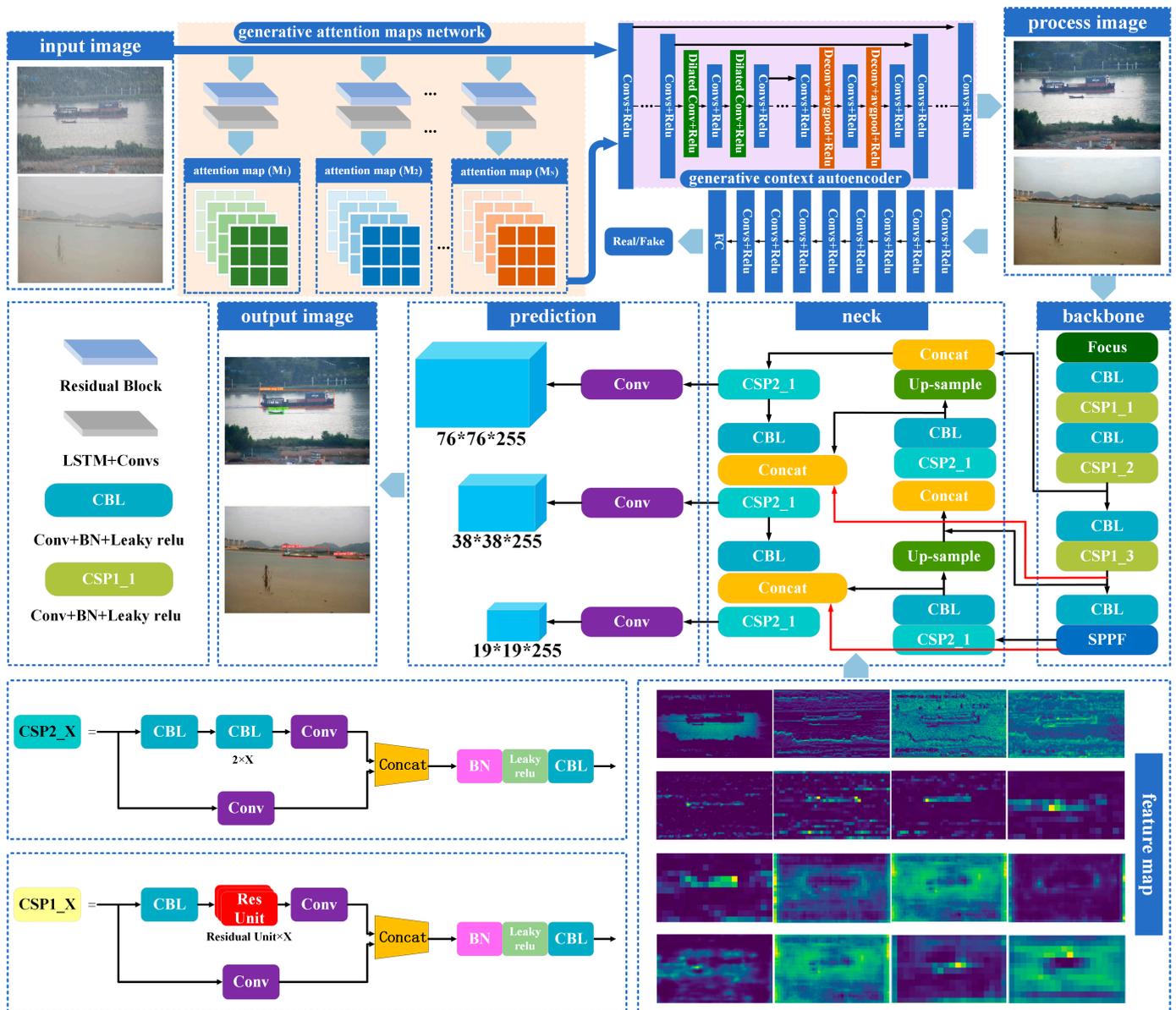


Figure 1. Schematic overview for ship detection framework for distorted image restoration under adverse weather.

For the simulation of foggy conditions, we have found that the most realistic methods predominantly utilize depth-based techniques to synthesize their own datasets. The widely used foggy model scene (Nayar and Narasimhan, Narasimhan and Nayar [30,31]) modeling formula is shown in Equation (2):

$$N(y) = \frac{H(y) - L}{e^{-\varphi d(y)}} + L \tag{2}$$

where  $N(x)$  represents the clean image,  $H(y)$  represents the fog image, and  $L$  represents the global atmospheric light.  $e^{-\varphi d(y)}$  is the transmission map.  $\varphi$  is the attenuation coefficient and  $d(y)$  refers to the image scene depth.

### 2.2. Generative Adversarial Network

Generative adversarial networks (GAN) (Goodfellow et al. [32]) have gained widespread application in the field of image restoration in recent years, and have yielded significant

results (Qian et al. [33]). A GAN network consists of two components: a generator and a discriminator. The generator takes random noise as input and produces a feature vector representing the target as output. The discriminator is a classifier that takes a vector as input and outputs a judgment on whether that vector is real or fake. More specifically, the generator takes low-visibility ship images as input and generates the restored image after passing through the attention map and context autoencoder within the generator. Furthermore, the restored image is used as input to the discriminator to distinguish between the generated images and real images, thereby guiding the generator to produce more realistic images. Finally, the sigmoid function outputs 1 or 0, indicating real or fake for the restored image. To make the model more efficient, the generator and discriminator evolve in a minimax game, where they mutually constrain and encourage each other to achieve more realistic image outputs. The optimization objective function of the GAN model is shown in Equation (3):

$$\min_G \max_D V(D, G) = E_{T \sim R_{imgnoise}} [\log(D(T))] + E_{B \sim P_{imgdrop}} [\log(1 - D(G(B)))] \quad (3)$$

where  $G$  represents the generator,  $D$  represents the discriminator, and  $B$  represents the low-visibility image input to the generator adversarial network.  $T$  represents the clear image sample corresponding to the low-visibility image of  $B$ .  $E$  represents the expected value.  $D(T)$  is the output of the discriminator for the real clear image  $T$ , which is a probability value.  $D(G(B))$  represents the output of the discriminator for the restored images generated by the generator  $G(B)$ .

### 2.3. Generative Attention Map Network

Rain streaks and fog in the atmosphere can significantly reduce the visibility of the maritime background in monitoring equipment, and cause image distortion. The distorted regions are perceptible to the human eye, but not explicitly delineated in computer-generated images. To address this problem, the paper proposes a method to identify the distorted regions by generating attention maps using adversarial neural networks. The generation of attention maps in this context is inspired by the principles of recurrent neural networks (RNNs) and residual networks. More specifically, as shown in Figure 1, the recurrent network consists of five layers of ResNet network layers and LSTM + Convs. The ResNet network is responsible for pre-extracting global features from the image (He et al. [34]). After passing through the LSTM + Convs module, both global and local features of the distorted image are fed back into the convolutional layers to generate the attention map.

The generated attention map is then concatenated with the original image and fed into the next identical module. In this process, the attention map from the previous layer guides the subsequent layer of the same network to focus more on the distorted regions. This iterative process is repeated in a loop. The attention map is essentially a two-dimensional array of the same size as the original image, where each element's value ranges from 0 to 1. The attention map is a non-binary mapping, signifying that attention gradually increases from non-raindrop regions to raindrop regions, with values varying even within the raindrop regions. This gradual increase in attention is meaningful because the areas around raindrops also need attention, and the transparency within the raindrop regions varies in reality. Therefore, a higher value in a specific region of the array indicates more attention from the attention map to that area, enabling focused restoration of the distorted regions in the image.

### 2.4. Generative Context Autoencoder

The framework obtains the distorted regions of the image through attention maps, but the original image information collected is missing (distorted) under the interference of rain streaks and fog. Therefore, the network uses a context autoencoder to help the generator produce a clear and complete image guided by the attention map, which is equivalent to

restoring and repairing low-visibility images. Figure 1 shows the architecture of the context autoencoder. The network introduces a dilated convolution network to increase the size and perception ability of the receptive field, so as to better capture the global features of the input attention map and the context information of the distorted regions. At the same time, the generative network ensures the restoration of high-resolution in the distorted regions by introducing the Deconv + avgpool module. Since we need to extract image feature information from different network layers to infer more context information, we set up two loss functions in the context encoder: multi-scale loss and perceptual loss. The multi-scale loss can effectively extract image features to obtain context information on different scales and form outputs of different sizes to capture the details and structure information in the image. The use of multi-scale loss is effective in extracting image features to obtain context information on different scales and generating outputs of varying sizes to capture fine-grained details and structural information in the image. The objective function for the multi-scale loss is shown in Equation (4):

$$\varepsilon_M(\{F\}, \{R\}) = \sum_{i=1}^M W_i \varepsilon_{MSE}(F_i, R_i) \tag{4}$$

where  $F_i$  represents the  $i$ -th output extracted from the context autoencoder, and  $R_i$  represents the  $i$ -th output ground-truth image information, which has the same scale as the  $F_i$ .  $\varepsilon_{MSE}$  represents the mean squared error between the output at different scales and the corresponding ground-truth image, and  $\{W\}_{i=1}^M$  represents the weight magnitudes for different scales.

To generate more realistic images, the generative network needs to pay attention to the high-level structure and content of the image, rather than pixel-level noise. This approach prioritizes the perceptual features of the image over subtle pixel variations. Besides the pixel-based multi-scale loss in our image-based approach, the generative network also uses perceptual loss to ensure visual consistency and feature fidelity between the generated image and the target image. The objective function for the perceptual loss is shown in Equation (5):

$$\varepsilon_P(O, T) = \varepsilon_{MSE}(VGG(O), VGG(T)) \tag{5}$$

where  $O$  represents the output image of the generative network, which is the image after the restoration process.  $O$  is obtained by the generator  $G$  using the input image  $I$  and the attention map.  $VGG(O)$  represents the image features extracted from  $O$  using a pre-trained VGG-16 network, while  $VGG(T)$  represents the image features extracted from  $T$ .  $\varepsilon_{MSE}$  represents the mean-squared-error loss function, which calculates the difference between the features of the reconstructed image and the ground-truth image after the restoration process. The VGG-16 mentioned in this paper refers to a pretrained convolutional neural network (CNN) that is solely used for feature extraction from images. To summarize, the loss function for low-visibility image restoration is shown in Equation (6):

$$\varepsilon_G = \varepsilon_M(\{F\}, \{R\}) + 10^{-2} \varepsilon_P(O, T) + \varepsilon_{GAN}(O) + \varepsilon_{A \sim M} \tag{6}$$

where  $\varepsilon_{A \sim M}$  represents the loss value of the attention map with the distortion region mask and  $\varepsilon_{GAN}(O) = 1 - D(G(O))$ . To verify the authenticity of the repaired distorted images, the generative network utilizes attention maps to guide the discriminator's focus on the restored regions, evaluating the quality of the generated images based on both global- and local-image content. Additionally, the generative network employs fully connected layers to determine the authenticity of the restored low-visibility images.

### 2.5. Detection Method

Considering the navigation control decisions of maritime ships, the ability to detect other ships at sea in real time and accurately handle emergency situations (such as collision avoidance and locating missing vessels) is crucial. The YOLO model, as a one-stage

detection model, has certain advantages in this regard. While two-stage object-detection models may offer superior accuracy, they do not stand out in terms of real-time performance. Furthermore, the equipment and monitoring devices on intelligent unmanned ships typically lack the computational capacity to support higher-precision target-detection algorithms. Moreover, this model not only detects the positions of ships but also classifies different types of ships. Therefore, choosing YOLOv5 for improvement can be effectively deployed in the ship's driving system to enhance the efficiency of maritime traffic management. YOLOv8 is the latest model-detection framework in the YOLO series. Although YOLOv8 has better accuracy and speed on GPU devices than YOLOv5, making it a better choice for real-time object detection, it is important to consider the device limitations of the ship perception system and the lack of GPU support. YOLOv5, with its smaller model parameters and ease of training, becomes a more suitable solution for such problems while maintaining a certain level of accuracy.

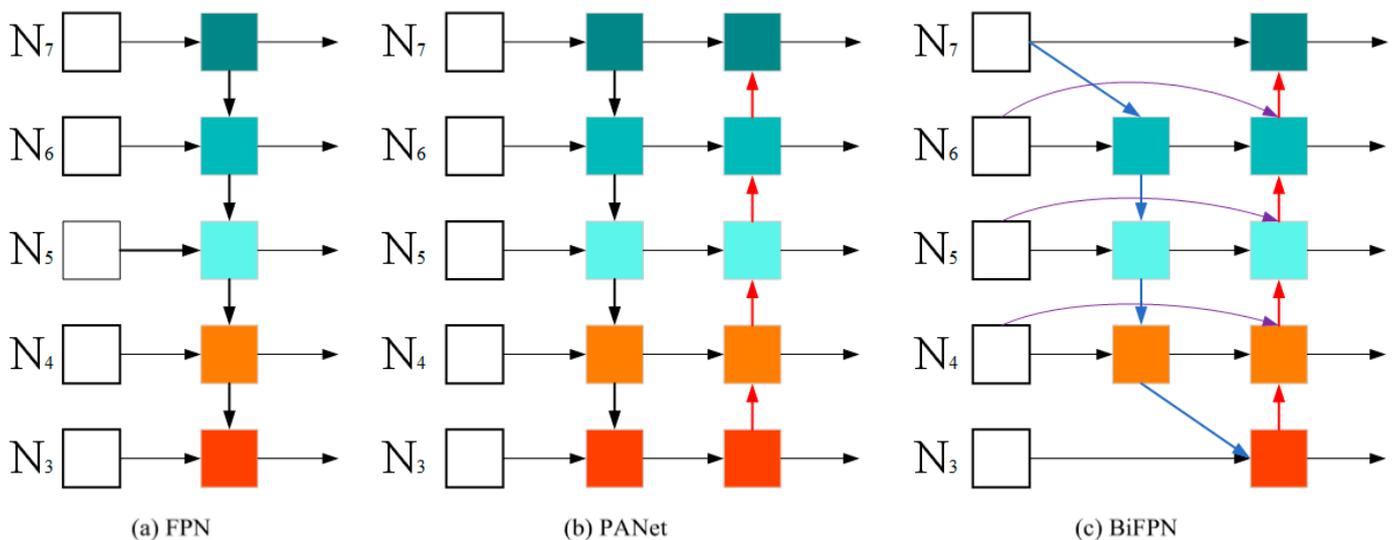
In this section, the detailed specifics of the YOLOv5 detector will be introduced. The network structure of YOLOv5 consists of three main components: backbone, neck and head. As shown in Figure 1, first, the restored low-visibility ship image is preprocessed (scaling the input ship image to a uniform size), and then sent to the backbone network, which transforms the original input image into multi-layer feature maps. The backbone network of YOLOv5 consists of CBL, CSP1\_X and SPPF modules. The CBL module is composed of a convolutional layer, a batch normalization layer and an activation function. This module is mainly used to extract the local spatial information of the ship features and normalize the feature information. The CSP1\_X convolutional module splits the input feature map into a backbone convolutional layer and a branch convolutional layer. The backbone convolutional layer uses  $1 \times 1$  convolution to reduce the channel number and the parameter amount. The branch convolutional layer further extracts feature extraction on the feature map. This design effectively reduces the computational load by reducing the parameter count while enhancing the feature extraction capability. Additionally, this module adopts a residual approach, which further enhances the model's expressive power. The detection model employs spatial pyramid pooling (SPPF) to pool features from input feature maps of different scales, enabling the model to capture maritime ship objects at various scales. Then, the detection model introduces the BiFPN feature pyramid structure in the neck network to handle the ships of different scales and sizes that are distorted by image restoration.

In order to recognize different types of ships, the key is to collect image features of different ships (including color, shape, etc.) through supervised learning for further processing. Therefore, by inputting the annotated dataset into the model training, anchors of different sizes and aspect ratios are preset. The setting of these anchors can effectively divide the prediction box space into several subspaces, thereby reducing the difficulty of recognizing different types of ships. Through down-sampling by factors of 32, 16, and 8, different sizes of feature maps ( $20 \times 20$ ,  $40 \times 40$ , and  $80 \times 80$ ) are produced. And these feature maps are input into the neck, where deep semantic features and low-level semantic features are fully fused. Each feature region is then input into the prediction head. Finally, the obtained feature maps are input into the head for feature regression and classification, fitting the best bounding boxes and positions for different types of ships.

### 2.6. Introduce the BiFPN Structure

Due to the attention adversarial network in the restoration process of low-visibility images, the distortion in the restoration of target ships in the images and the uneven distribution of small ships cause the image features of such target ships to be insignificant. Due to issues such as image distortion and occlusion, during the later stages of model training, the features extracted from the restored normal regions are significantly more prominent than those from the un-restored normal regions. The features of the restored normal area are found to be more significant than those of the un-restored normal area during the subsequent feature extraction, leading to a selective focus on the significant

features. Therefore, to solve such problems, the YOLO series previously adopted the FPN (feature pyramid network) structure. As shown in Figure 2a, this network structure adopts a top-down approach to aggregate multi-scale features, allowing the high-level feature map to be transmitted to the low-level feature map (different square colors represent different feature maps, the arrow direction represents the direction of feature transmission.) and aggregating features on different scales. However, this information transmission can only be unidirectional, and cannot be reversed. More specifically, as the number of down-sampling or convolution operations increases, the receptive field of the high-level feature map gradually increases, and the overlap area between the receptive fields also continues to increase. At this time, the information represented by the pixel points is the information for a region, which has stronger semantic information and is more conducive to the classification of different ships. The low-level feature map can utilize more fine-grained feature information, ensuring that the network can capture more details. It has stronger positional information, which is more conducive to the positioning of ships. Then, the process of transmitting high-level features can lead to the loss and degradation of feature information. Therefore, a unidirectional FPN cannot effectively solve such problems.



**Figure 2.** Comparative schematic diagram of different feature-fusion network structures: (a) FPN, (b) PANet, and (c) BiFPN.

To better solve the FPN problem, it is necessary to create a new path from bottom to top, transmitting the positional information to the predicted feature map as well, so that the predicted feature map simultaneously possesses higher semantic information and positional information (which is beneficial for object detection). As shown in Figure 2b, PANet proposes a bidirectional feature network from top to bottom and from bottom to top, and generates a new feature map from bottom to top. This is followed by adaptive feature pooling in the later stages. This network structure enhances the feature expression capability of the backbone network, allowing different target ships to choose different feature maps. This avoids the one-to-one matching between ship size and network depth. However, the ROI of this network structure can only rely on a single layer of features, leading to the problem of information loss from other feature layers. Then, the dual transmission paths can also lead to insufficient information transmission.

Therefore, in order to better solve the above problems, this study introduces the BiFPN weighted bidirectional (top-down + bottom-up)-feature pyramid network structure as a multi-scale feature-fusion method, and combines the idea of multi-level feature fusion. This is an effective bidirectional cross-scale weighted feature-fusion method, which enables the fusion and transfer of features from high-resolution ship images and low-resolution ship images. This further avoids the problems of erroneous ship detection and recognition

caused by occlusions and image restoration distortions between ships of different sizes, and better balances the feature information on different scales under different circumstances. As shown in Figure 2c, the feature pyramid network structure is shown in the diagram, where the left part represents the input section consisting of feature maps from the backbone network. These feature maps have different levels and scales of information. Typically, lower-level feature maps have higher resolution but relatively less semantic information, while higher-level feature maps have lower resolution but contain more semantic information. The input feature maps are fused and propagated through various paths. The top-down path starts from higher-level feature maps and gradually increases the resolution of the feature maps through upsampling or interpolation operations to obtain higher-resolution feature maps. The bottom-up path starts from lower-level feature maps and gradually decreases the resolution of the feature maps through pooling or convolution operations to capture broader receptive fields and more detailed information. The lateral connections are used to fuse the feature maps from the top-down and bottom-up paths. By using  $1 \times 1$  convolution operations, the channel dimensions of the feature maps from the bottom-up path are matched to be added to or concatenated with the feature maps from the top-down path. Multiple iterations are performed to enrich and diversify the levels of the feature pyramid. Finally, the fused and propagated feature map is output on the right side of the image. More specifically, the blue lines are the top-down pathways, which convey the semantic information of the high-level features; the red lines are the bottom-up pathways, which convey the location information of the low-level features; the purple lines are the newly added edges between the input nodes and the output nodes at the same level (N4, N5, N6), which fuse more image features without adding too much cost. Meanwhile, In the BiFPN network, nodes with only a single input edge are eliminated. This is because a node with just one input edge that does not perform feature fusion contributes minimally to the feature network that integrates different features. Therefore, removing such a node has a negligible impact on our network, while it simplifies the bidirectional network. This is applicable to the first node on the right of N7. If the original input node and the output node are at the same level, the network will add an extra edge between the original input node and the output node. This allows for the fusion of more features without significantly increasing costs, thereby improving the efficiency of ship detection and recognition.

### 3. Experimental Design

#### 3.1. Data Description

To validate the proposed maritime ship detection framework for low-visibility scenarios at sea, it is essential to consider all influencing factors in the maritime environment to ensure the reliability and authenticity of the model validation. The SeaShip dataset is acquired by the monitoring cameras in a deployed coastline video-surveillance system. This dataset includes labels for various types of ships and high-precision bounding boxes, and covers all possible imaging variations, such as different scales, parts of the hull, lighting, viewpoints, backgrounds, and occlusions. This dataset comprises 7000 images, which are divided into two subsets: one simulates rain-streak conditions, and the other simulates foggy weather. Both datasets are split into training, testing, and validation sets, in a 2:1:1 ratio. The dataset encompasses six types of ships, namely ore carriers, bulk cargo carriers, container ships, general cargo ships, fishing boats, and passenger ships (Shao et al. [35]).

This dataset has collected maritime ship navigation data under forty-five different background conditions. Ship detection accuracy is often affected by background changes which pose challenges for separating foreground target ships from complex background environments. Note that the datasets also involved image distortion, occlusions, hull parts and small-target detection (i.e., small ship imaging size) interferences. Details for the datasets can be found in Table 1. The proposed method was implemented with PyTorch 1.7.1 framework and Python 3.7. The operating system is Ubuntu 20.04 OS, and the CPU is Intel(R) Xeon(R) Gold 6230R CPU @ 2.10 GHz. The GPU used for the experimental platform is Quadro RTX 5000.

**Table 1.** Information from Marine ship data.

Ship Category	Resolution	Image Distortion	Small Target Detection	Ship Obstruction	Hull Parts
Ore carrier	1920 × 1080	✓	✓	✓	✓
Bulk cargo carries	1920 × 1080	✓	✓	✓	✓
Container ship	1920 × 1080	✓	/	✓	✓
General cargo ship	1920 × 1080	✓	✓	✓	✓
Fishing boat	1920 × 1080	✓	✓	✓	✓
Passenger ship	1920 × 1080	✓	/	✓	✓

(Symbol ✓ indicates the situation that exists in the dataset.)

### 3.2. Evaluation Indicators

To validate the performance of the GYB framework proposed in this study, our approach involved five metrics: recall (R), F1-score, precision (P), and the average precision  $AP_{0.5}$  and  $AP_{0.5:0.95}$  to quantitatively evaluate the performance of the framework. the performance of the framework. Firstly, we need to introduce some common variables. TP (true positive) represents the positions and labels of the ships detected by different algorithms consistent with the ground truth. TN (true negative) represents that both the ground truth and detected ship labels are negative (which correctly predict the negative samples). FN (false negative) means that the different detection algorithms recognize the correct ship positions and labels as wrong (this sample is a positive sample). FP (false positives) means the different detection algorithms predict the wrong ship positions and labels as correct (this sample is a negative sample). Precision is the ratio of correctly predicted ship positions and labels among all predicted true ship labels within the range of [0, 1]. Recall represents the percentage of correctly predicted true ship labels among the total actual true ship labels, and lies within the range of [0, 1].

To better evaluate the performance of different algorithms, the F1-score, which is the harmonic mean, precision (P) and recall (R) are introduced. The F1-score reaches its optimum only when both precision and recall tend toward their maximum values. The average precision (AP) can be obtained by calculating the area under the precision–recall (P-R) curve, which is bounded by the horizontal and vertical axes.  $AP_{0.5}$  and  $AP_{0.5:0.95}$  are commonly utilized metrics, respectively signifying average precision values at an IOU (intersection-over-union) threshold of 50% and the mean values across IOU thresholds, ranging from 50% to 95%. In this study, the detection performance of the proposed method was assessed using  $AP_{0.5}$  and  $AP_{0.5:0.95}$ . In accordance with Equations (7)–(10), the positions of the detected ships are closer to the ground truth when the values of P, R,  $F_1$ ,  $AP_{0.5}$  and  $AP_{0.5:0.95}$  are larger. The frames-per-second (fps) is introduced as a performance criterion to assess the real-time performance of this framework, for which the calculation formula is shown in Equation (11):

$$P = \frac{TP}{TP + FP} \tag{7}$$

$$R = \frac{TP}{TP + FN} \tag{8}$$

$$F_1 = 2 \frac{P \times R}{P + R} \tag{9}$$

$$AP = \int_0^1 P(r) dr \tag{10}$$

$$fps = \frac{1}{COT} \tag{11}$$

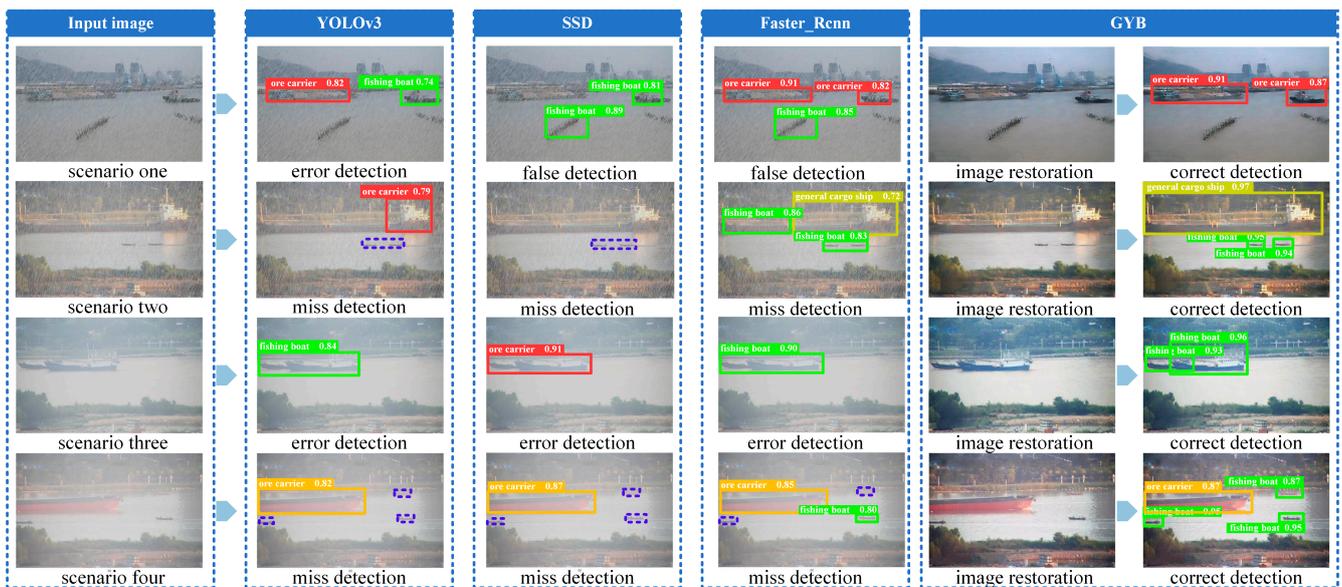
where COT represents the average time consumed per frame in the ship validation dataset.

### 4. Discussion and Result

#### 4.1. Discussion

To illustrate the entire workflow, we provide a descriptive output of each step within the proposed framework in this paper. In the context of maritime surveillance video imaging, the interference caused by rain streaks and fog leads to reduced visibility in monitoring video data. Furthermore, the complexity and diversity of maritime ship navigation environments exacerbate this issue. Existing ship-detection algorithms tend to exhibit abnormal detection behavior in such scenarios, often resulting in missed detections, where they fail to accurately fit bounding boxes around navigating ships.

As shown in scenario one in Figure 3, under the interference of rain streaks and the complex water traffic environment, SSD and Faster\_Rcnn failed to correctly distinguish the features of water obstacles and ships, resulting in false and missed detections of ships (purple dashed box). As shown in scenario two in Figure 3, the interference of rain streaks reduced the visibility, and the sailing ships were far from the monitoring equipment, resulting in YOLOv3 and SSD missing the detection of small-ship targets. Meanwhile, the YOLOv3 detector failed to generate the correct ore-carrier bounding box. The significant difference between the edge features of the general cargo ship and the edge features of the cargo on board caused Faster\_Rcnn to generate multiple incorrect candidate regions (proposal regions), resulting in fishing boats and general cargo ships being generated (i.e., a single ship corresponds to multiple bounding boxes).

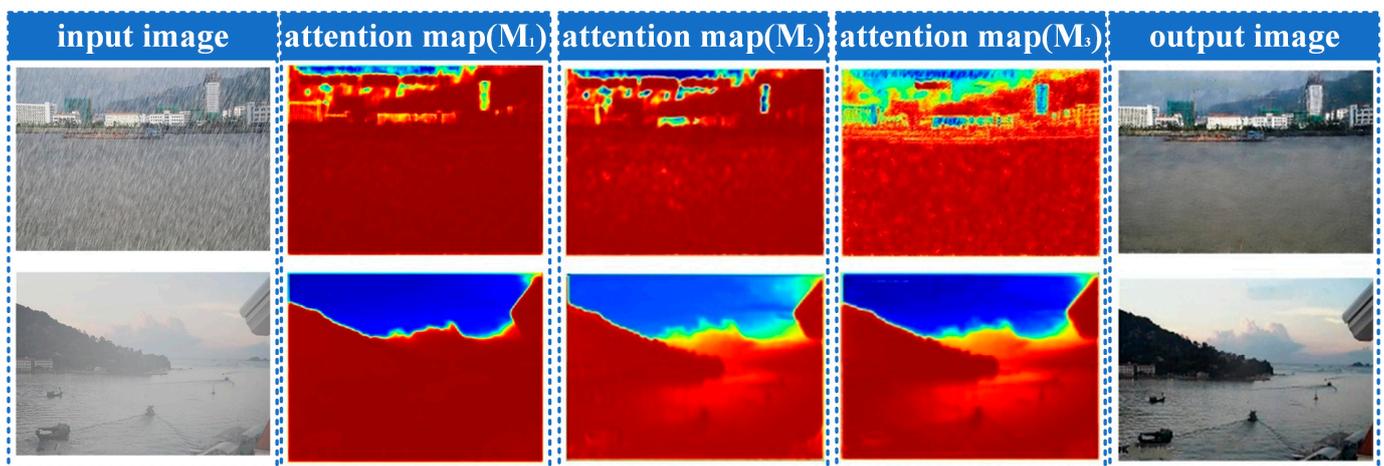


**Figure 3.** Comparative diagram of the effects of YOLOv3, SSD, Faster\_Rcnn, and GYB on the detection of ships in typical water transportation under different weather disturbances.

As shown in scenario three in Figure 3, the visibility of the video image data collected under the interference of fog is significantly reduced, and the ships are occluded by each other; the features of small ships are similar to those of large ships, resulting in SSD, YOLOv3 and Faster\_Rcnn being unable to distinguish fine-grained ship features, causing ship misdetection (i.e., multiple ships correspond to only one bounding box). As shown in scenario four in Figure 3, the small-target fishing boats in this scene are confused with the surrounding environmental features, and SSD, Faster\_Rcnn and YOLOv3 cannot identify effective regions, resulting in some small ships being missed (purple dashed box).

By restoring the low-visibility images and obtaining a clear image as the input of the enhanced YOLOv5 detector, our proposed algorithm framework can effectively address the challenges of low-visibility detection in the aforementioned scenarios and achieve multi-scale fusion for water ship detection. Then, the restoration of the distorted area of the image becomes more important. As shown in Figure 4, the visualization of the low-

visibility region attention-map learning process in the proposed framework is presented. This process focuses on the low-visibility regions by using the attention maps generated by the adversarial neural network, while understanding the structure and edge features around the low-visibility regions. And by guiding the context encoding encoder with the attention map, the global and local features and relations of the low-visibility regions are captured, generating high-quality image regions. In this way, we can ensure that the detector can extract more realistic and effective ship image features. In order to better quantify the effects of image dehazing and deraining, the PSNR (peak signal-to-noise Ratio) and SSIM (structural similarity) are introduced as performance criteria to assess the results after image restoration. PSNR (peak signal-to-noise ratio) is a reference value for evaluating image quality, while SSIM (structural similarity) is an indicator for measuring the similarity between the restored image and the real clear image. The evaluation indicators are summarized in Table 2. It can be observed that the PSNR metrics of the framework for image deraining and defogging are 30.32 and 32.68, respectively, and the SSIM metrics are 0.9289 and 0.9360, respectively. Overall, it has achieved a good quality of image restoration, providing an important foundation for subsequent ship detection and recognition.



**Figure 4.** Schematic diagram of the process of image restoration by the context encoder guided by the attention map.

**Table 2.** Quantitative evaluation results of image deraining and dehazing.

Datasets	Evaluation Indicators	
	PSNR	SSIM
Rain streaks	30.32	0.9289
Fog	32.68	0.9360

#### 4.2. Results

By calculating the data difference between the real position and the detected position of each type of ship in low-visibility maritime surveillance videos, we further quantified the experimental data regarding the framework’s performance. The evaluation results are summarized in Table 3. For rain streak scene, the evaluation metrics (P,R,F1) of our proposed GYB framework model are 95.2%, 94.3%, 94.8%.  $AP_{0.5}$  and  $AP_{0.5:0.95}$  are 0.970 and 0.701, respectively, which are more than 10% higher than the traditional algorithms YOLOv3, SSD and Faster\_Rcnn models. Similarly, for the fog scene, the evaluation indicators of the GYB model are more than 20% higher than traditional algorithms, while also meeting the real-time requirements. Simultaneously, we find that the framework proposed in this paper achieves a fps of 28.67 frames and 29.06 frames in the rain streak and fog scenes by calculating the fps of different models. Therefore, this framework ensures the accuracy of

ship detection and classification while meeting the real-time requirements of ship systems. At the same time, in order to further verify the accuracy and reliability of the framework for recognizing different types of ships, we conducted a separate indicator evaluation for the detection and recognition of individual ships, and summarize the evaluation indicators in Tables 4 and 5. It can be seen that, whether in rainy or foggy weather, the P, R, and F1 of single-ship detection and recognition are all above 92%, especially due to the obstruction of ship structures of different sizes and limited visibility in rainy and foggy weather. In summary, the experimental results show that the proposed framework can effectively solve the problem of ship detection and recognition problems, even in low-visibility conditions.

**Table 3.** Performance statistics of ship detection for waterborne navigation in different weather conditions.

Data	Model	Evaluation Indicators					
		P	R	F1	AP <sub>0.5</sub>	AP <sub>0.5:0.95</sub>	fps
rain streaks	GYB	95.2%	94.3%	94.8%	0.970	0.701	28.67
	YOLOv3	71.7%	48.4%	57.8%	0.578	0.256	11.98
	SSD	83.5%	79.4%	81.4%	0.822	0.434	6.47
	Faster_Rcnn	60.6%	59.7%	61.0%	0.614	0.294	3.26
fog	GYB	97.4%	96.4%	96.9%	0.984	0.742	29.06
	YOLOv3	69.9%	58.9%	60.6%	0.867	0.268	12.68
	SSD	89.8%	84.9%	87.3%	0.867	0.408	7.03
	Faster_Rcnn	61.2%	63.5%	62.3%	0.655	0.317	3.78

**Table 4.** Performance statistics of detection and recognition of different types of ships under rain streak weather conditions.

Rain Streaks	Evaluation Indicators				
	P	R	F1	AP <sub>0.5</sub>	AP <sub>0.5:0.95</sub>
ore carrier	94.8%	92.6%	93.7%	0.971	0.660
passenger ship	94.4%	92.9%	93.6%	0.971	0.656
container ship	99.9%	100%	99.9%	0.995	0.783
bulk cargo carrier	92.8%	92.1%	92.4%	0.949	0.696
general cargo ship	95.9%	95.1%	95.5%	0.971	0.737
fishing boat	93.5%	93.0%	93.2%	0.963	0.662

**Table 5.** Performance statistics of detection and recognition of different types of ships under foggy weather conditions.

Fog	Evaluation Indicators				
	P	R	F1	AP <sub>0.5</sub>	AP <sub>0.5:0.95</sub>
ore carrier	98.4%	90.7%	94.4%	0.985	0.714
passenger ship	98.3%	96.5%	97.4%	0.966	0.715
container ship	96.7%	99.0%	97.8%	0.995	0.786
bulk cargo carrier	97.2%	97.7%	97.4%	0.99	0.766
general cargo ship	99.2%	97.7%	98.4%	0.992	0.771
fishing boat	94.7%	96.7%	95.7%	0.974	0.703

In order to gain a deeper understanding of the roles and importance of each part in the model framework, we conducted a series of ablation experiments. In these experiments,

we sequentially added the image restoration module and the BiFPN module, observing their impact on the performance of the model framework. The image restoration method has been renamed as GY and compared with the GYB framework proposed in Table 6 for the aforementioned experiments. The evaluation metrics of the control experiment results are summarized in Table 6. We found that after adding image restoration and the BiFPN module, the detection and recognition metrics of different ships have improved to some extent. Under the conditions of rain streak weather, the YOLOv5 model with the added BiFPN module has seen improvements in its P, R, and F1 score by 9.1%, 7.5%, and 8.6%, respectively, compared to the original model. The GY model framework has seen improvements in its P, R, and F1 scores by 15.3%, 51.2%, and 38%, respectively, compared to the original model. Meanwhile, both  $AP_{0.5}$  and  $AP_{0.5:0.95}$  have improved to varying degrees. In combination with the model framework proposed in this paper, all evaluation metrics of the model have seen a significant improvement. This ensures that the accuracy of the model is improved without reducing the speed of model inference. Similarly, under the conditions of foggy weather, after adding different modules to the model all evaluation metrics of the model have seen a significant improvement.

**Table 6.** Comparison of experimental results of different modules.

Data	Model	Evaluation Indicators					
		P	R	F1	$AP_{0.5}$	$AP_{0.5:0.95}$	fps
Rain streaks	YOLOv5	65.0%	34.1%	44.7%	0.405	0.216	33.37
	YOLOv5 + BiFPN	74.1%	41.6%	53.3%	0.503	0.288	31.75
	GY	80.3%	85.3%	82.7%	0.887	0.516	30.45
	GYB	95.2%	94.3%	94.8%	0.970	0.701	28.67
Fog	YOLOv5	70.8%	55.5%	62.2%	0.622	0.368	35.37
	YOLOv5 + BiFPN	75.9%	49.1%	48.1%	0.580	0.351	30.85
	GY	83.6%	70.0%	76.2%	0.786	0.509	29.78
	GYB	97.4%	96.4%	96.9%	0.984	0.742	29.06

The experimental results show that the proposed framework can effectively solve the problem of ship detection and recognition problems, even in low-visibility conditions. It can be seen that, under the conditions of low-visibility small-target ship detection, large-area distortion of the image, and the challenge of area occlusion between ships, the new ship-detection framework (GYB) proposed in this paper has more robust performance than the traditional algorithms in the real complex low-visibility sea environment.

## 5. Conclusions

The detection of ships in maritime traffic navigation is of paramount practical significance for safeguarding navigation safety and facilitating intelligent control decision-making. In this paper, we propose an integrated framework for the detection and recognition of maritime ships under low-visibility conditions. This framework achieves accurate detection and recognition of ships in distorted maritime video data. The proposed framework utilizes adversarial neural networks to generate attention maps, which enable the identification of rain streaks and fog areas within the images. These attention maps guide the contextual autoencoder to selectively restore the low-visibility regions of the images, based on the surrounding information. Then, the restored images are input into the YOLOv5 detector, which incorporates multi-scale feature fusion, to achieve accurate detection and classification of navigating ships in the video images. We validated the performance of the proposed GYB framework using the SeaShip dataset. Experimental results showed that the proposed framework can achieve satisfactory performance for maritime ship detection under rainy and foggy weather conditions, with an average precision of 96.3%, an average recall of

95.4%, an average harmonic mean of 95.9% and an average fps of 28.87 frame. The average precision  $AP_{0.5}$  and  $AP_{0.5:0.95}$  are 0.977 and 0.722. The experimental results demonstrate that the proposed framework significantly improves the precision of ship detection and classification under adverse weather conditions.

The following directions can be expanded to further enhance the model applicability in the future. First, the SeaShip dataset exhibits a limited diversity in ship categories. It is advisable to augment the dataset with additional ship types, including small fishing boats, to enhance the model's robustness. Second, the density of maritime ships collected is not particularly high, and it is worthwhile for us to further investigate the validation of maritime ship detection in high-density scenarios. Last, but not least, we can also add the detection of ships with different rotation angles to further verify the performance of the model under different water transportation scenarios.

**Author Contributions:** Conceptualization, X.C. and J.Z.; methodology, X.C., C.W., Z.X. and J.Z.; writing—original draft preparation, X.C., C.W. and J.Z.; writing—review and editing, Z.X., C.W., J.Z. and J.X.; funding acquisition, X.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was jointly supported by National Natural Science Foundation of China (52331012, 52102397, 52071200, 72101157, 71942003).

**Data Availability Statement:** Readers interested in the work can email the corresponding author to request data sharing.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Liu, R.W.; Nie, J.; Garg, S.; Xiong, Z.; Zhang, Y.; Hossain, M.S. Data-Driven Trajectory Quality Improvement for Promoting Intelligent Vessel Traffic Services in 6G-Enabled Maritime IoT Systems. *IEEE Internet Things J.* **2021**, *8*, 5374–5385. [\[CrossRef\]](#)
2. Cheng, S.; Zhu, Y.; Wu, S. Deep learning based efficient ship detection from drone-captured images for maritime surveillance. *Ocean Eng.* **2023**, *285*, 115440. [\[CrossRef\]](#)
3. Volden, Ø.; Cabecinhas, D.; Pascoal, A.; Fossen, T.I. Development and experimental evaluation of visual-acoustic navigation for safe maneuvering of unmanned surface vehicles in harbor and waterway areas. *Ocean Eng.* **2023**, *280*, 114675. [\[CrossRef\]](#)
4. Forti, N.; d’Afflisio, E.; Braca, P.; Millefiori, L.M.; Carniel, S.; Willett, P. Next-Gen Intelligent Situational Awareness Systems for Maritime Surveillance and Autonomous Navigation [Point of View]. *Proc. IEEE* **2022**, *110*, 1532–1537. [\[CrossRef\]](#)
5. Bahnsen, C.H.; Moeslund, T.B. Rain Removal in Traffic Surveillance: Does it Matter? *IEEE Trans. Intell. Transp. Syst.* **2019**, *20*, 2802–2819. [\[CrossRef\]](#)
6. Li, M.; Cao, X.; Zhao, Q.; Zhang, L.; Meng, D. Online Rain/Snow Removal from Surveillance Videos. *IEEE Trans. Image Process.* **2021**, *30*, 2029–2044. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Fu, K.; Li, Y.; Sun, H.; Yang, X.; Xu, G.; Li, Y.; Sun, X. A Ship Rotation Detection Model in Remote Sensing Images Based on Feature Fusion Pyramid Network and Deep Reinforcement Learning. *Remote Sens.* **2018**, *10*, 1922. [\[CrossRef\]](#)
8. Lu, H.; Li, Y.; Lang, L.; Wu, L.; Xu, L.; Yan, S.; Fan, Q.; Zheng, W.; Jin, R.; Lv, R.; et al. An Improved Ship Detection Algorithm for an Airborne Passive Interferometric Microwave Sensor (PIMS) Based on Ship Wakes. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5302012. [\[CrossRef\]](#)
9. Yu, C.; Cai, J.; Chen, Q. Multi-resolution visual fiducial and assistant navigation system for unmanned aerial vehicle landing. *Aerosp. Sci. Technol.* **2017**, *67*, 249–256. [\[CrossRef\]](#)
10. Yao, P.; Sui, X.; Liu, Y.; Zhao, Z. Vision-based environment perception and autonomous obstacle avoidance for unmanned underwater vehicle. *Appl. Ocean Res.* **2023**, *134*, 103510. [\[CrossRef\]](#)
11. Wang, B.; Han, B.; Yang, L. Accurate Real-time Ship Target detection Using Yolov4. In Proceedings of the 2021 6th International Conference on Transportation Information and Safety (ICTIS), Wuhan, China, 22–24 October 2021; pp. 222–227.
12. Li, J.; Tian, J.; Gao, P.; Li, L. Ship Detection and Fine-Grained Recognition in Large-Format Remote Sensing Images Based on Convolutional Neural Network. In Proceedings of the IGARSS 2020–2020 IEEE International Geoscience and Remote Sensing Symposium, Waikoloa, HI, USA, 26 September–2 October 2020; pp. 2859–2862.
13. Feng, J.; Li, B.; Tian, L.; Dong, C. Rapid Ship Detection Method on Movable Platform Based on Discriminative Multi-Size Gradient Features and Multi-Branch Support Vector Machine. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 1357–1367. [\[CrossRef\]](#)
14. Shao, Z.; Wang, L.; Wang, Z.; Du, W.; Wu, W. Saliency-Aware Convolution Neural Network for Ship Detection in Surveillance Video. *IEEE Trans. Circuits Syst. Video Technol.* **2020**, *30*, 781–794. [\[CrossRef\]](#)
15. Lv, Y.; Li, M.; He, Y. An Effective Instance-Level Contrastive Training Strategy for Ship Detection in SAR Images. *IEEE Geosci. Remote Sens. Lett.* **2023**, *20*, 4007505. [\[CrossRef\]](#)

16. Chen, X.; Wang, Z.; Hua, Q.; Shang, W.L.; Luo, Q.; Yu, K. AI-Empowered Speed Extraction via Port-Like Videos for Vehicular Trajectory Analysis. *IEEE Trans. Intell. Transp. Syst.* **2023**, *24*, 4541–4552. [[CrossRef](#)]
17. Chen, X.; Guan, J.; Wang, Z.; Zhang, H.; Wang, G. Marine Targets Detection for Scanning Radar Images Based on Radar- YOLONet. In Proceedings of the 2021 CIE International Conference on Radar (Radar), Haikou, China, 15–19 December 2021; pp. 1256–1260.
18. Wang, Y.; Jiang, Z.; Li, Y.; Hwang, J.N.; Xing, G.; Liu, H. RODNet: A Real-Time Radar Object Detection Network Cross-Supervised by Camera-Radar Fused Object 3D Localization. *IEEE J. Sel. Top. Signal Process.* **2021**, *15*, 954–967. [[CrossRef](#)]
19. Xu, Q.; Li, Y.; Shi, Z. LMO-YOLO: A Ship Detection Model for Low-Resolution Optical Satellite Imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 4117–4131. [[CrossRef](#)]
20. Guo, Y.; Liu, R.W.; Qu, J.; Lu, Y.; Zhu, F.; Lv, Y. Asynchronous Trajectory Matching-Based Multimodal Maritime Data Fusion for Vessel Traffic Surveillance in Inland Waterways. *IEEE Trans. Intell. Transp. Syst.* **2023**, 1–14. [[CrossRef](#)]
21. Bai, J.; Li, S.; Huang, L.; Chen, H. Robust Detection and Tracking Method for Moving Object Based on Radar and Camera Data Fusion. *IEEE Sens. J.* **2021**, *21*, 10761–10774. [[CrossRef](#)]
22. Shang, W.L.; Gao, Z.; Daina, N.; Zhang, H.; Long, Y.; Guo, Z.; Ochieng, W.Y. Benchmark Analysis for Robustness of Multi-Scale Urban Road Networks Under Global Disruptions. *IEEE Trans. Intell. Transp. Syst.* **2022**, 1–11. [[CrossRef](#)]
23. Lin, J.C.W.; Srivastava, G.; Zhang, Y.; Djenouri, Y.; Aloqaily, M. Privacy-Preserving Multiobjective Sanitization Model in 6G IoT Environments. *IEEE Internet Things J.* **2021**, *8*, 5340–5349. [[CrossRef](#)]
24. Wang, H.; Xu, Y.; He, Y.; Cai, Y.; Chen, L.; Li, Y.; Sotelo, M.A.; Li, Z. YOLOv5-Fog: A Multiobjective Visual Detection Algorithm for Fog Driving Scenes Based on Improved YOLOv5. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 2515612. [[CrossRef](#)]
25. Hassaballah, M.; Kenk, M.A.; Muhammad, K.; Minaee, S. Vehicle Detection and Tracking in Adverse Weather Using a Deep Learning Framework. *IEEE Trans. Intell. Transp. Syst.* **2021**, *22*, 4230–4242. [[CrossRef](#)]
26. Tan, M.; Pang, R.; Le, Q.V. EfficientDet: Scalable and Efficient Object Detection. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 10778–10787.
27. Quan, R.; Yu, X.; Liang, Y.; Yang, Y. Removing Raindrops and Rain Streaks in One Go. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 9143–9152.
28. Luo, Y.; Xu, Y.; Ji, H. Removing Rain from a Single Image via Discriminative Sparse Coding. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; pp. 3397–3405.
29. Li, Y.; Tan, R.T.; Guo, X.; Lu, J.; Brown, M.S. Rain Streak Removal Using Layer Priors. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 2736–2744.
30. Nayar, S.K.; Narasimhan, S.G. Vision in bad weather. In Proceedings of the Seventh IEEE International Conference on Computer Vision, Kerkyra, Greece, 20–27 September 1999; Volume 822, pp. 820–827.
31. Narasimhan, S.G.; Nayar, S.K. Contrast restoration of weather degraded images. *IEEE Trans. Pattern Anal. Mach. Intell.* **2003**, *25*, 713–724. [[CrossRef](#)]
32. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. *Adv. Neural Inf. Process. Syst.* **2014**, 27.
33. Qian, R.; Tan, R.T.; Yang, W.; Su, J.; Liu, J. Attentive Generative Adversarial Network for Raindrop Removal from A Single Image. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 2482–2491.
34. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
35. Shao, Z.; Wu, W.; Wang, Z.; Du, W.; Li, C. SeaShips: A Large-Scale Precisely Annotated Dataset for Ship Detection. *IEEE Trans. Multimed.* **2018**, *20*, 2593–2604. [[CrossRef](#)]

**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.