IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 26 July 2022, accepted 10 August 2022, date of publication 17 August 2022, date of current version 23 August 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3199352

RESEARCH ARTICLE

Light-SDNet: A Lightweight CNN Architecture for Ship Detection

MENGYAO ZHANG, XIANWEI RONG, AND XIAOYAN YU, (Member, IEEE)
School of Physics and Electronic Engineering, Harbin Normal University, Harbin 150025, China

Corresponding author: Xiaoyan Yu (yuxiaoyan@hrbnu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61401127, and in part by the Provincial Natural Science Foundation and Cultivation Project of National Natural Science Foundation of Harbin Normal University under Grant XPPY202208.

ABSTRACT Ship detection plays a vital role in monitoring and managing maritime safety. Most recently proposed learning-based object detection methods have achieved marked progress in detection accuracy, but the size of these models is too large to be applied to mobile devices with limited resources. Although some compact models have been presented in the previous study, they achieve unsatisfactory results in ship detection, especially under extreme weather conditions. To address these challenges, this article presents a lightweight convolutional neural network (CNN) called Light-SDNet to perform an end-to-end ship detection under different weather conditions. In the proposed model, we introduce the improved CA-Ghost, C3Ghost, and DepthWise Convolution (DWConv) into the You Only Look Once version 5 (YOLOv5) to reduce the number of model parameters, while remaining its powerful feature expression ability. We use parallel attention to highlight the features that contribute to the ship detection in the marine surveillance. To enhance the adaptability of the proposed model, a hybrid training strategy with generating synthetically-degraded images is proposed to augment the volume and diversity of the original datasets. The proposed strategy enables Light-SDNet to improve the ship detection results under severe weather conditions such as haze, rain, and low illumination. We compare Light-SDNet with other competitive approaches on a large-scaled ship dataset called SeaShips. We show that Light-SDNet achieves a better balance between the detection accuracy and the model complexity. The ship detection results on degraded marine images have proven the superior performance of the proposed model in terms of detection accuracy, robustness and efficiency.

INDEX TERMS Ship detection, convolutional neural network, lightweight structure, attention mechanism.

I. INTRODUCTION

It is increasingly important to enhance maritime traffic safety with the development of offshore economic activities and the exploration of marine resources. In particular, ship collision accidents occur frequently under extreme weather conditions. The Automatic Identification System (AIS) [1] has achieved remarkable results in maritime surveillance. However, the Class A AIS that can send self-ship information is only mandatory on ships that can load more than 300 tons, so it may lead to omissions in the detection of small and medium-sized ships. Meanwhile, some illegal ships deliberately turned off related equipment in an attempt to evade detection and surveillance. Thus, the video surveillance system is essential to further improve maritime supervision and security. The maritime supervisors can obtain intuitive visual information by observing surveillance video images, while their visual fatigue caused by the long-term observation may result in the neglect of important information. With the rapid development of deep-learning technology, many advanced ship detection methods have been proposed, which provides a strong support for building an intelligent maritime video surveillance system.

Unlike traditional methods that require the hand-crafted features and suffer from unpleasant detection results, the progressive learning-based methods achieve an end-to-end object detection and better performance by extracting useful features automatically [2], [3]. Currently, learning-based methods can be typically divided into two categories. One is a single-stage framework such as SSD [4] and YOLO series [5],

The associate editor coordinating the review of this manuscript and approving it for publication was Kegen Yu.

[6], [7], [8], [9], and the other is a two-stage framework such as R-CNN [10], Fast R-CNN [11], and Faster R-CNN [12]. The former enables faster detection with lower computational burden, while the latter tends to achieve more accurate results in exchange of slower detection speed [13], [14]. The aforementioned methods are not lightweight enough so that they are unsuitable to be applied in the maritime video surveillance systems with limited memory and computation power.

To address this problem, many efforts have been paid to develop compact and efficient CNN models. EfficientNet [15] adopts a compound scaling method, which markedly reduces the model parameters and improves the classification speed via scaling up any dimension of the network (width, depth or resolution). MobileNetV3 [16] is a very lightweight and low-latency model obtained by neural architecture search, whose modules used internally are inherited from the depthwise separable convolution [17] and the inverted residual structure with a linear bottleneck [18]. In addition, it uses a SENet [19] attention module after the pointwise convolution to enhance important features. Considering the detection accuracy and processing speed, ShuffleNetV1/V2 [20], [21] manages the exchange of information between the groups via the channel shuffle operations. GhostNet [22] uses the Ghost modules to extract more features from cheap operations. The aforementioned infrastructures are capable of the extraction of effective features, but in the expense of decreased detection accuracy.

In reality, the visual quality of images captured from maritime surveillance systems is generally affected by poor imaging conditions, such as rain, haze, and low illumination [23]. Image deterioration adversely affects vessel traffic safety and security, thus the accurate ship detection in the degraded maritime images becomes intractable. To improve the accuracy of ship detection under bad weather conditions (e.g., rain, fog, low illumination), the degraded maritime images are typically recovered before ship detection using image restoration algorithms [24], [25], [26]. However, using these restored images tends to a decline of the accuracy and robustness of ship detection due to the loss of detailed features.

To make ship detection more robust and accurate under different weather conditions, a hybrid data training strategy is introduced to enlarge the diversity and volume of the original dataset. In addition, we propose a compact and efficient network based upon improved YOLOv5 for the ship detection on the mobile or embedded devices. By combining the proposed model and the hybrid data training strategy, there is a great potential for the proposed method to obtain a reliable ship detection with higher accuracy, efficiency, and robustness. The contributions of this study can be summarized as follows:

(1) We propose a lightweight network for ship real-time detection named Light-SDNet. To assign greater weights to more valuable information, both the coordinate and parallel attention mechanisms are introduced into the proposed lightweight network. Specifically, the attention-guided CA-Ghost and the C3Ghost module extract features and fuse features at the Backbone and Neck, respectively.

(2) Extensive experiments on the large ship dataset called SeaShips show that the proposed Light-SDNet can achieve higher detection accuracy with comparative model parameters and computation burden, which is suitable for mobile terminals or embedded systems with the limited computation power and memory capacity.

(3) A hybrid data training strategy is proposed to solve ship detection in adverse weather conditions. Experimental results on degraded ocean images demonstrate the superior performance of our proposed model in terms of detection accuracy, robustness, and efficiency.

The remainder of the paper is organized as follows. Section II briefly reviews existing ship detection methods. Section III presents the proposed YOLOv5-enhanced ship detection framework. Section IV describes the proposed hybrid training strategy and exhibits extensive experimental results on the SeaShips dataset. We finally summarize the main contributions of this study in Section V.

II. RELATED WORKS

A. THE YOLO SERIES

In comparison with the region-based object detection methods, the end-to-end YOLO series [5], [6], [7], [8], [9] is faster due to one time of input processing. To improve the detection accuracy, YOLOv2 [6] adopts the K-means clustering technique for the model training. Moreover, DarkNet-19 was presented based upon the idea of 'Network in Network' [27]. Compared with ResNet [28], YOLOv3 [7] achieves competitive detection accuracy with fewer model parameters via using DarkNet-53 as the Backbone. In addition, YOLOv3 manages the object detection in the multi-scale feature maps via up-sampling and fusion method similar to feature pyramid networks (FPNs) [29]. YOLOv4 [8] achieves the state-of-the-art detection results via combination of multiple optimization strategies such as data augmentation (e.g., mixup [30], mosaic.), network modules (Focus, SPP [31] and improved PANet [32], CSPNet [33]), activation function (mish [34] and swish [35]), and loss function (CIOU) [36].

YOLOv5 algorithm [9] adopts various enhancement techniques at the input, such as mosaic, adaptive image scaling, and adaptive anchors. The main purpose of the first convolution in the Backbone is to reduce model parameters, floating point operations (FLOPs), and memory overhead, so that the forward and backward speed is increased with marginal effects on detection accuracy. By applying gradient change to the feature map, the C3 module is capable of tackling the issue of repeating gradients in the Backbone of a large-scale neural network. Additionally, the network incorporates Spatial Pyramid Pooling-Fast (SPPF) and Path Aggregation Network (PANet) to enhance its feature fusion capabilities.

B. ATTENTION MECHANISM

The attention mechanism is a resource allocator that adaptively assigns weights to features via channel or spatial modeling. SENet [19] captures cross-channel information

with global average pooling, while extracting all channel features may be inefficient and unnecessary. To improve the cost-effective performance of network models, ECA-Net [37] adopts a local cross-channel interaction, i.e., a one-dimensional convolution is used to screen the strong inter-channel dependencies. In this way, ECA-Net markedly improves the network performance yet effectively reduces the number of model parameters. However, they merely focus on the relationship between channels, ignoring the significance of spatial features. In contrast, BAM [38] and CBAM [39] can extract the inter-channel relationship of features and the intra-spatial relationship of features, so they obtain richer high-level features for the vision tasks. In addition, coordinate attention (CA) [40] helps to localize objects of interest more accurately via embedding location information into channel attention. The CA can also be flexibly inserted into the mobile network without any computational overhead.

C. SHIP DETECTION IN MARITIME SURVEILLANCE SYSTEM

Ship detection plays an important role in maritime traffic safety, so extensive efforts have been made in the field of automatic detection of moving ships. For example, Zhang *et al.* proposed a ship detection method based upon discrete cosine transform (DCT) [41]. The detection method primarily includes three stages, i.e., background modelling, background subtraction and horizon detection, which can achieve robust detection results under complex sea conditions with surface waves. According to the visual attention model, Shi *et al.* obtain the saliency maps for the ship detection via fusing directional features, color features, and motion features [42]. Chen *et al.* proposed a real-time ship detection and tracking system based on mean shift, which is able to achieve good automatic tracking performance [43]. However, conventional ship detectors typically endure unsatisfactory detection accuracy under severe marine imaging conditions (e.g., haze, rain, and low-luminance), due to it being highly dependent on hand-crafted features.

CNNs provide a new avenue for accurate and efficient detection of the moving ships owing to its powerful feature extraction capability. A number of CNN-based methods have been recently developed for the ship detection in different maritime environments. For example, Cui *et al.* improved CenterNet with a spatial shuffling attention module to achieve a large-scale ship detection in the synthetic aperture radar (SAR) images [44]. To address the issue of difficult deployment of the existing models on the edge devices with limited memory resources, Ma *et al.* proposed a lightweight object detector via compressing YOLOv4 [45]. To effectively identify ships with various scales in high-resolution optical remote sensing images, Li *et al.* generated candidate ships from the feature maps using a region-proposal network [46]. So far, most detection tasks are implemented in the SAR and optical remote sensing images, while SAR and optical remote sensing images typically endure low signal to noise ratio (SNR), which results in the difficulty in detecting small objects.

To address this problem, great attempts have been undertaken to develop efficient CNN-based models for ship detection in the natural images. For example, Shao et al proposed a saliency-aware CNN framework to achieve an accurate and real-time ship detection in the surveillance video images [47]. The framework includes coastline priors, deep features, and saliency maps. In addition, coarse-to-fine cascaded CNNs for ship detection and tracking have received extensive attention, leading to autonomous maritime surveillance [48]. To develop a robust ship detector under severe weather conditions, an enhanced YOLOv3 is proposed with data augmentation training, whose results demonstrate its effectiveness for ship detection [49]. The existing CNN-based ship detection methods have achieved marked progress, while they may be typically unsuitable for use on the embedded devices and mobile terminals with limited computation power and storage capacity because of their highly computational complexity and large model size.

To achieve a better balance between the model complexity and detection accuracy, we aim to develop a lightweight CNN architecture for ship detection in the maritime video surveillance via improving YOLOv5 and introducing a hybrid training strategy. We also provide an ablation study to show the functions of critical components of light-SDNet, and describe extensive results to verify its good performance in the ship detection under different maritime environments, especially under extreme weather conditions.

III. THE PROPOSED SHIP DETECTION FRAMEWORK

To solve the problems of low detection accuracy and difficult deployment of redundant networks in maritime surveillance, we propose a lightweight ship detection network (Light-SDNet) based on YOLOv5s. In this section, we describe the proposed method's exploration trajectory and overall framework.

A. EXPLORATION TRAJECTORY

We fine-tuned the YOLOv5s network in this part. On the premise of ensuring detection accuracy, the network parameters are compressed to reduce the computation burden. Figure 1 shows the exploration trajectory from YOLO5s to Light-SDNet.

To achieve a lightweight and powerful network, the following changes have been made to the original YOLOv5s:

(1) For extract better location features of the shallow network, the Ghost module [50] with CA replaces the common convolution module of the Backbone to perform 2× downsampling.

(2) DWConv replaces the convolution module used in the Neck network, reducing computational bottleneck and memory overhead.

(3) The C3Ghost replaces the C3 module as the main feature fusion module of the Neck network, guaranteeing

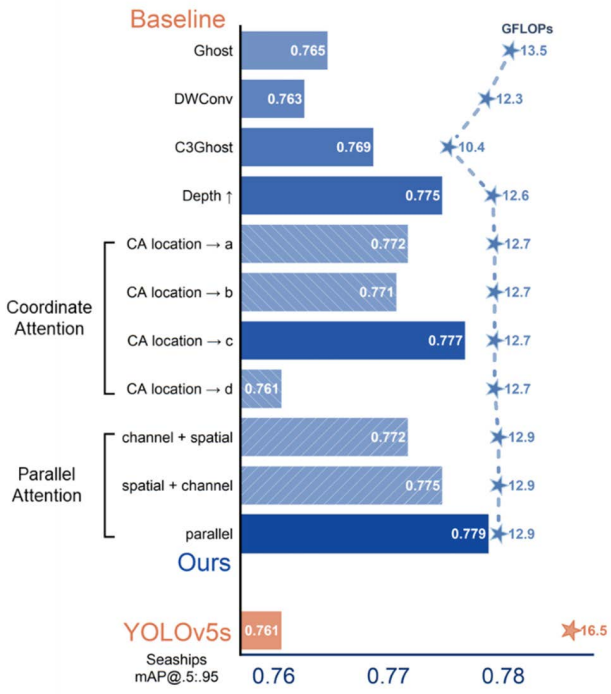

FIGURE 1. We modernize YOLOv5s towards the design of the Light-SDNet. The bars are model accuracies in the YOLOv5s FLOP regime. There is no adoption of the modification if the bar is hatched. In the end, our model can outperform YOLOv5s.

the lightweight nature and detection accuracy of the target network.

(4) Parallel attention (PA) is additionally introduced to further enhance the ability of Neck feature fusion.

In addition, the depth multiplier of the target network is increased from 0.33 to 0.50 to enhance its learning ability. The final architecture of the proposed model is shown in Table 1 and Figure 2.

B. BACKBONE

In the Backbone network, the ship image first goes through a convolution with a size of 6 × 6, a stride of 2, and a padding of 2 to perform downsampling. Then it goes through four stages, all of which contain the Ghost module with CA and C3 module. The process is summarized as follows: the CA-Ghost module performs 2× downsampling of the input from the previous stage, and the C3 module performs feature extraction to obtain a total of four feature maps with different scales.

In the feeding process of feature maps in the Backbone, scale compression and channel expansion lead to the gradual transmission of spatial information to the channel. To compensate for the loss of shallow features, the CA [40] is introduced into the Ghost module to construct the CA-Ghost module. We focus on the width and height of feature maps to improve model performance at a low cost.

Figure 3(a) provides the four types of CA-based Ghosts we designed based on the location of the embedded CA, namely Ghost-a, Ghost-b, Ghost-c, and Ghost-d. Experimental results show that we obtain the largest mAP when integrating CA after DWConv in Ghost. Thus, we use Ghost-c as CA-Ghost. Figure 3(b) depicts the structure of the CA block. The input of the CA-Ghost module goes through two branches: the left branch passes through DWConv to reduce the size of feature maps, and then the convolution module is used to double the channel; the right branch first performs the convolution operation by the ghost module, and DWConv is used under the guidance of CA for downsampling, and then the ghost module doubles the channel again. Finally, the two branches are directly added together as the output. As shown in Table 1, the parameters of CA-Ghost module are reduced by more than two times over those of the ordinary convolution, which reduces computation cost and reinforces positional features.

C. NECK

When the feature map goes the Neck, the channel dimension reaches the maximum, while the resolution of network reaches the minimum. The SPPF module then focuses on spatial information to solve the problem of excessive changes of object scales. As shown in Table 1, we also replaced the common convolution of the Neck with DWConv. Unlike traditional convolution, DWConv is a convolution kernel responsible for one channel, convolving channel by channel, which can markedly reduce the model size. However, this is at the expense of some network fusion capabilities.

The specific structures of three types of C3 modules are shown in Figure 4. As shown in Figure 4(c) and (d), the ghost unit is used to replace the bottleneck (False), resulting in a new C3Ghost module. The C3Ghost module is the main feature fusion module of Neck, containing three convolutions and multiple ghost units, where n represents the number of embedded ghost units. The structure splits the gradient flow into different network paths and integrates all changes into the feature map. The ghost unit reduces computation cost and compresses the model sizes by replacing the original bottleneck (False)'s 3 × 3 standard convolution. In this way, it ensures the fusion and extraction of features and optimizes the accuracy and parameters.

To focus more on the features of the ships in the image, we add a parallel attention mechanism after the C3Ghost module. As shown in Figure 5, the parallel attention module is implemented by combining BAM [38] with ECA-Net [37]. That is, the spatial attention comes from BAM, and the channel attention comes from ECA-Net. By combining the channel attention Mc(F) and the spatial attention Ms(F) from the two attention branches, we can generate the 3D attention map M(F) by taking the sigmoid function. To obtain a refined feature map, this 3D attention map is element-wise multiplied by the input feature map F and then added to the original input feature map. Furthermore, the ablation study in Section IV shows that the parallel structure of ECANet and BAM is more

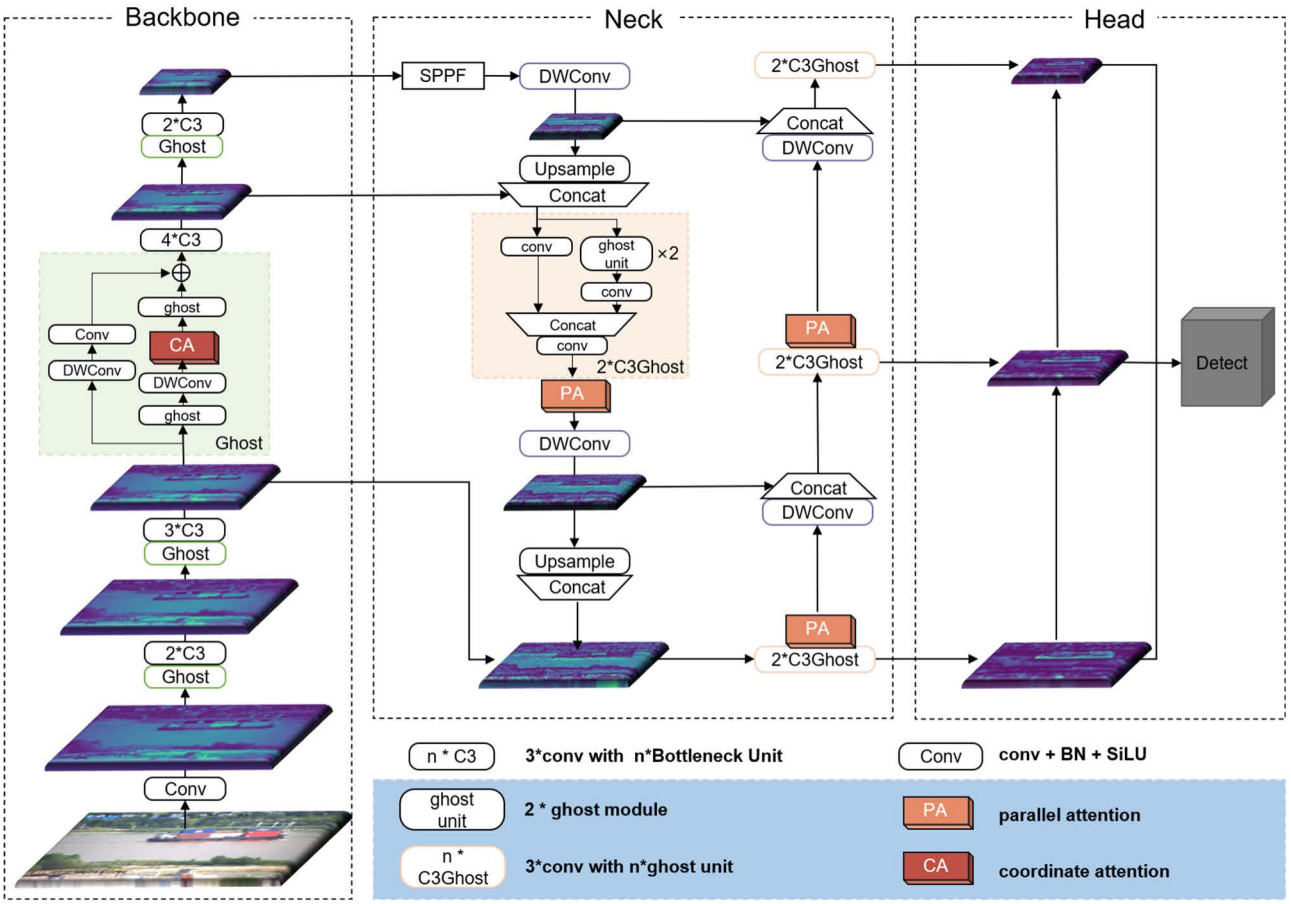

FIGURE 2. Architecture of the proposed Light-SDNet. It mainly includes three parts: the Backbone network, the Neck network, and the Head. The input ship image is extracted through the Backbone network. Multiscale features are further fused in the Neck. Finally, multi-scale object detection is performed on the head. The area covered in blue is our main improvement module.

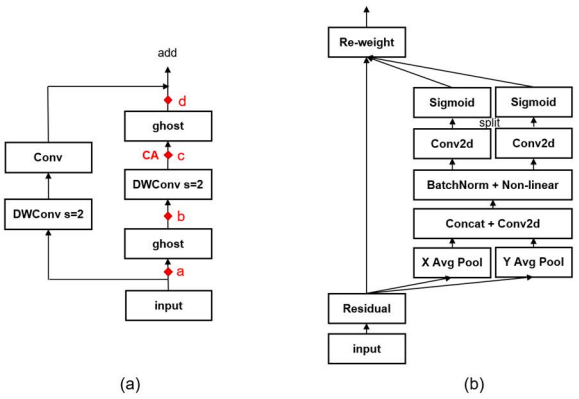

FIGURE 3. CA-Ghost module. (a) CA-based Ghost, where a, b, c, and d denoted in red represent the positions of embedded CA. The testing results indicate that Ghost-c performs best, so we select it as CA-Ghost; (b) Structure of a CA block.

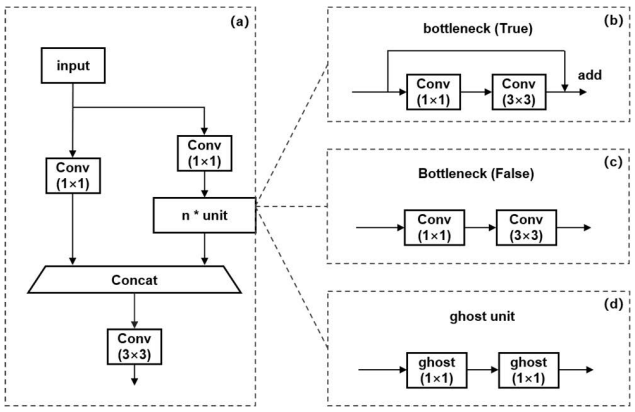

FIGURE 4. Three types of C3 modules are applied to the Backbone and the Neck. (a) Original C3 module; (b) C3 module for feature extraction in the Backbone of YOLOv5s; (c) C3 module for feature fusion in the Neck of YOLOv5s; (d) C3Ghost module in the Neck of Light-SDNet.

efficient than the sequential structure, so we adopt a parallel design in the proposed attention module.

D. HEAD

In the detection Head, three sets of output feature maps are detected to generate a final output vector with class probability scores, bounding boxes, and confidence scores. According to the Non-Maximum Suppression (NMS), the output of the three detection layers are screened to obtain the final detection results.

The loss function of the proposed method consists of three parts: classification loss (cls_loss), localization loss

TABLE 1. Comparison of original YOLOv5s and Light-SDNet.

	YOLOv5s					Light-SDNet				
	Type	Unit-Num	kernel/Stride	Output	Params	Type	Unit-Num	kernel/Stride	Output	Params
	Conv	-	6*6/2	320*320*32	3520	Conv	-	6*6/2	320*320*32	3520
B	Conv	-	3*3/2	160*160*64	18560	**CA-Ghost**	-	3*3/2	160*160*64	**6664**
A	C3	1	-	160*160*64	18816	C3	**2**	-	160*160*64	29184
C	Conv	-	3*3/2	80*80*128	73984	**CA-Ghost**	-	3*3/2	80*80*128	**20472**
K	C3	2	-	80*80*128	115712	C3	**3**	-	80*80*128	156928
B	Conv	-	3*3/2	40*40*256	295424	**CA-Ghost**	-	3*3/2	40*40*256	**69592**
O	C3	3	-	40*40*256	625152	C3	**4**	-	40*40*256	789504
N	Conv	-	3*3/2	20*20*512	1180672	**CA-Ghost**	-	3*3/2	20*20*512	**253848**
E	C3	1	-	20*20*512	1182720	C3	**2**	-	20*20*512	1839104
	SPPF				656896	SPPF				656896
	Conv	-	1*1/1	20*20*256	131584	**DWConv**	-	1*1/1	20*20*256	**1024**
	Upsample	-	Stride = 2	40*40*256	0	Upsample	-	Stride = 2	40*40*256	0
	Concat	-	-	40*40*512	0	Concat	-	-	40*40*512	0
	C3	1	-	40*40*256	361984	**C3Ghost**	**2**	-	40*40*256	**219584**
						PA				**8868**
P	Conv	-	1*1/1	40*40*128	33024	**DWConv**	-	1*1/1	40*40*128	**512**
R	Upsample	-	Stride = 2	80*80*128	0	Upsample	-	Stride = 2	80*80*128	0
E	Concat	-	-	80*80*256	0	Concat	-	-	80*80*256	0
D	C3	1	-	80*80*128	90880	**C3Ghost**	**2**	-	80*80*128	**56544**
I	Detect-1					Detect-1				
C						**PA**				**2260**
T	Conv	-	3*3/2	40*40*128	147712	**DWConv**	-	3*3/2	40*40*128	**1408**
I	Concat	-	-	40*40*256	0	Concat	-	-	40*40*256	0
O	C3	1	-	40*40*256	296448	**C3Ghost**	**2**	-	40*40*256	**186816**
N	Detect-2					Detect-2				
						PA				**8868**
	Conv	-	3*3/2	20*20*256	590336	**DWConv**	-	3*3/2	20*20*256	**2816**
	Concat	-	-	20*20*512	0	Concat	-	-	20*20*512	0
	C3	1	-	20*20*512	1182720	**C3Ghost**	**2**	-	20*20*512	**603008**
	Detect-3					Detect-3				

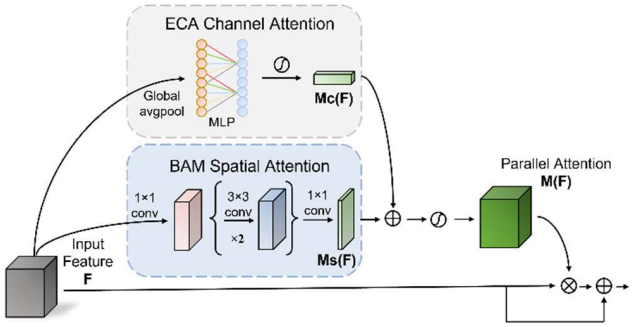

FIGURE 5. Structure of a parallel attention module. The attention module is implemented by combining the ECA and the BAM spatial attention blocks.

(loc_loss), and confidence loss (obj_loss), the formula is as follows,

$$Loss = \lambda_1 L_{cls} + \lambda_2 L_{loc} + \lambda_3 L_{obj} \quad (1)$$

where λ_1, λ_2 and λ_3 are coefficients to weight the loss contribution with values of 0.5, 0.05, and 1.0, respectively. The confidence loss and the classification loss are calculated by combining the BCE (Binary Cross Entropy) loss with the logistic loss, and the CIoU loss is used to evaluate the localization loss of the predicted box and the ground-truth box.

IV. EXPERIMENTAL RESULT AND ANALYSIS

To assess the performance of Light-SDNet, we compare it with other methods qualitatively and quantitatively. To get good detection results in both normal weather and severe weather conditions, we propose an end-to-end mixed-data training strategy. The proposed training strategy has also been implemented to demonstrate good performance under poor imaging conditions.

A. IMPLEMENTATION DETAILS

1) DATASET DESCRIPTION AND SETTINGS

We use the public ship dataset named SeaShips [51] as the original dataset. The dataset includes 7000 images that cover

6 types of ships, such as ore carriers, bulk carriers, general cargo ships, container ships, fishing boats, and passenger ships. The maritime images originate from video camera surveillance systems that track all ships near shore. Bad weather conditions such as fog, rain, and low light, tend to markedly deteriorate the quality of the images captured by a marine surveillance system, so we constructed three degraded datasets based on the classic SeaShips dataset. The extended degradation datasets are shown in Section C. For the SeaShips dataset and its degraded dataset, they are randomly divided into training, validation, and test sets with a 3:1:1 ratio for the experiments. The SeaShips_fog dataset is used to explore the model structure, while the SeaShips dataset and its degraded one are used to measure the impact of severe weather conditions on ship detection performance. The effectiveness of the proposed hybrid training strategy is verified below.

2) EXPERIMENTAL ENVIRONMENT AND PARAMETER SETTINGS

Our ship detection experiments use Pytorch (1.8.0) software library installed in Ubuntu 18.04. Specifically, all experiments are performed on a computer with an Intel(R) Xeon (R) Silver 4210R CPU @2.40 GHz and NVIDIA GeForce RTX 3090 GPU. For the optimal hyperparameters used in our network, the base learning rate, momentum and weight decay are, respectively, set to 0.01, 0.937, and 0.0005. In all experiments, the size of input images is 640 $\times$ 640 pixels, epoch is set to 300, and the batch size is set to 16. All the remaining parameters take the default values in the original YOLOv5.

B. METRICS

We follow the same criteria as PASCAL VOC [52] to evaluate the performance of Light-SDNet.

Precision is used to evaluate whether the prediction of ships is accurate, which reflects the proportion of actually positive samples over all predicted positive samples. *Recall* is used to evaluate whether all ships in the test dataset have been predicted correctly, which reflects the proportion of positive samples predicted correctly by the model over the total positive samples. *F1 score* is the harmonic average of precision and recall. *mAP@0.5* and *mAP@.5:.95* are comprehensive indicators to measure the precision and robustness of ship detection. *Correct detecting ratio (CDR)* is the proportion of correctly predicted samples over all samples. *False alarm ratio (FAR)* is the proportion of negative cases that are incorrectly classified as positive over all predicted positive samples.

Besides the above evaluation indicators, we provide the model parameters, FLOPs, and training time to verify the advanced nature of Light-SDNet. The less model parameters and FLOPs are, the lower the cost of the detection model is.

C. DATA AUGMENTATION AND THE HYBRID DATA TRAINING STRATEGY

The complex maritime environments such as fog, rain, and low light, typically enable the captured images to be blurred and blocked, bringing the huge difficulties to automatic ship surveillance. To explore the impacts of weather conditions on ship detection, we synthetically simulated the degraded images and constructed three degraded datasets that simulate fog, rain, and low light environments based on the classic SeaShips dataset. Great efforts would be devoted to the practical application of ship detection under severe weather conditions via this study. We also propose an end-to-end hybrid data training algorithm aimed at achieving ideal detection performance in normal and multiple severe weather conditions.

1) GENERATING SYNTHESIZED DEGRADED IMAGES

In the extended SeaShips_fog dataset, sea hazy images are generated based upon the atmospheric scattering model expressed as (2).

$$I(x, y) = J(x, y) \cdot t(x, y) + A \cdot t(1 - t(x, y)) \tag{2}$$

where $I(x, y)$ is a hazy image, $J(x, y)$ is a haze-free image, A is the global atmosphere light and $t(x, y)$ is the medium transmission map, decaying exponentially with the increased distance, which is formulated as (3).

$$t(x, y) = e^{-\beta d(x,y)} \tag{3}$$

where β is the medium attenuation coefficient and $d(x, y)$ is the scene depth. Several synthetically-degraded samples are shown in Figure 6. These hazy images with different concentrations are generated via adjusting the atmospheric light A and transmission map t. To simplify the process, the hazy images are synthesized by randomly taking and $t \in \hat{E}\{0.20, 0.25, 0.30, 0.35, 0.40, 0.45, 0.50\}$ and $A \in \{0.75, 0.80, 0.85, 0.90\}$.

The rainy images can be synthesized via superimposing the simulated raindrop trajectories on a clear image. Thus, a synthetically-degraded image $Z(x, y)$ with rain streaks can be formulated as follows

$$Z(x, y) = J(x, y) + B(x, y) \tag{4}$$

$J(x, y)$ is a latent sharp image and $B(x, y)$ is the raindrop noise layer. As shown in Figure 7, different rainy images can be synthesized by adjusting the lengths and angles of rain streaks. In the experiments, the number of raindrops is set to 800, the length of the rain streaks is ranged between 20 to 80 pixels, and the angle of the rain streaks is randomly chosen between -50 and 50.

A low-light maritime image is synthesized based on the Retinex theory, assuming an original image S is a product of the reflection image R and the illumination image L, i.e.,

$$S(x, y) = R(x, y) \cdot L(x, y) \tag{5}$$

where R may be seen as the latent sharp image, L represents the various intensities of light on the objects that are spatially

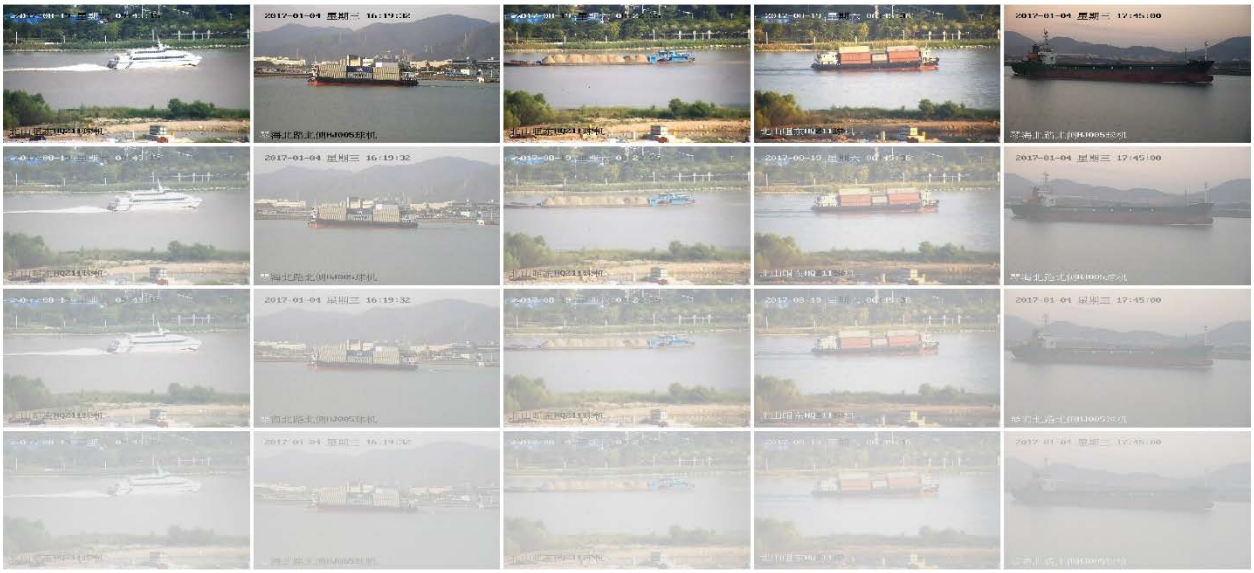

FIGURE 6. The example images of haze-degraded produced using (2) in the SeaShips dataset. From top to bottom: original sharp images, haze-degraded images with t = 0.50, t = 0.35, and t = 0.20(A is uniformly set to 0.9), respectively. Among them, t denotes the transmission map, and A is the atmospheric light.

FIGURE 7. Examples of rain-degraded images in the SeaShips dataset. From top to bottom: original images, rainy images with (RL = 20, RA = 0°), (RL = 80, RA = 50°), and (RL = 80, RA = −50°), respectively. The rain streak length (unit: pixel) and angle (unit: °) are expressed as RL and RA, respectively.

smooth. To synthesize the low-light maritime images, we first convert the original RGB images into HSV images. The ocean images are visually degraded via multiplying the V layer of the original images by different attenuation coefficients $\omega \in (0, 1)$. As shown in Figure 8, the low-light images are generated with $\omega = 0.1, 0.2, 0.3, 0.4$ and 0.5, respectively.

2) THE HYBRID DATA TRAINING STRATEGY

Table 2 shows the results of Light-SDNet trained with three degraded datasets synthesized artificially to evaluate the impact of different imaging conditions on ship detection (i.e., mAP@.5/mAP@.5:.95), including normal, hazy, low-light and rainy conditions. The results shown in Table 2 reveal that the precision of ship detection will increase markedly if the imaging conditions of model training and testing datasets maintain the same. Inspired by this [53], we propose a hybrid data training strategy. Each image has a probability of 3/4 to be randomly added with varying degrees of fog or rain or be converted to a low-light image before being input to the network for model training. To detect ships more effectively in dense fog and very low light conditions, we generate a wider range of fog concentrations and lower illumination

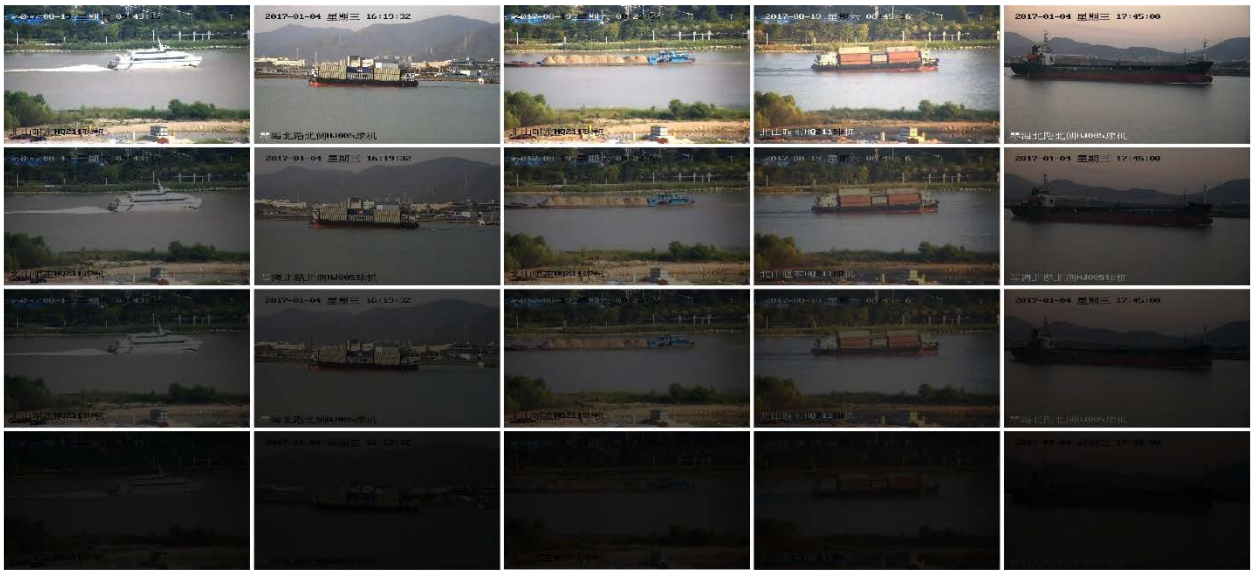

FIGURE 8. Examples of low-light maritime images in the SeaShips dataset. From top to bottom: original images, low-light maritime images with $\omega = 0.5$, $\omega = 0.3$, and $\omega = 0.1$, respectively.

TABLE 2. The effects of different adverse weather conditions on ship detection. (mAP@0.5/mAP@.5:.95).

Test dataset	Training dataset				
	Sharp images	Hazy images	Low-light images	Rainy images	Hybrid data
Sharp images	0.982/0.782	0.987/0.808	0.990/0.822	0.991/0.822	0.988/**0.826**
Hazy images	0.598/0.398	0.983/0.779	0.534/0.339	0.660/0.449	0.992/**0.825**
Low-light images	0.901/0.714	0.992/0.821	0.985/0.778	0.856/0.667	0.992/**0.824**
Rainy images	0.960/0.738	0.972/0.753	0.979/0.758	0.990/0.786	0.988/**0.810**

levels to simulate ocean scenes. A rainy scene is additionally simulated to complete the task well in the rainy days. In Algorithm 1, we describe the hybrid data training process in detail.

D. QUANTITATIVE EVALUATION

To verify the effectiveness of the proposed Light-SDNet, the quantitative evaluation was performed by comparison with state-of-the-art (SOTA) algorithms, including the YOLO series and the SOTA lightweight Backbone series. YOLO series include YOLOv3-lite, YOLOv4-lite, YOLOv5n and YOLOv5s, while popular lightweight Backbones include GhostNet [22], EiffcientNet-lite [15], MobileNetv3s [16], and ShuffleNet-v2 [21], specifically to replace the Backbone of YOLOv5s. The performance comparison was carried out on the synthetic Seaships fog dataset.

Figure 9 shows the curves of mAP, Precision, and Recall for all detectors during model trainings. As shown in Figure 9(a) and (b), Light-SDNet is better than other models due to its slightly higher mAP values, while Figure 9(c) shows that the performance of YOLOv3-tiny and YOLOv4-tiny is much lower than that of other models since YOLOv5 improves their feature extraction network and data augmentation techniques. As can be seen from Figure 9, all curves rise gently and converge rapidly, thereby indicating that the model is well trained without overfitting.

Algorithm 1 Training Procedure for Light-SDNet

Input: Original image of ship detection
Output: Calibration results of detection in adverse weather conditions
Initialize Light-SDNet $D \wedge \theta$ with random weights θ.
Set the training stage: *num_epochs* = 300, *batch_size* = 16.
Prepare the normal dataset Seaships_trainval.
for i in *num_epochs* **do**
  repeat
    Take a batch images M from Seaships_trainval.
    for j in *batch_size* **do**
      if *random.randint*(0, 3) > 0 **then**
        Generate the foggy image $M(j)$ by (2) and (3), where A = *random. choice* ([i/100 *for* i in *range* (60, 95, 5)]), t = *random.choice*([i/10 *for* i in *range*(1, 6)])
        //for foggy conditions
        Generate the rainy-image $M(j)$ by (4),
        Where L = *random.randint*(20, 80), A = *random.randint* (−50, 50) //for rain conditions
        Generate the low-light image $M(j)$ by (5),
        where ω = *random. choice* ([i/10 *for* i in *range* (1, 6)])
        //for low-light conditions
      end if
    end for
    Update Light-SDNet $D \wedge \theta$ according to detection loss.
  until all images have been fed into training models
end for

Table 3 shows that Light-SDNet achieves the best performance among the YOLO series since it improves mAP performance by 1.8% compared with the original YOLOv5s and

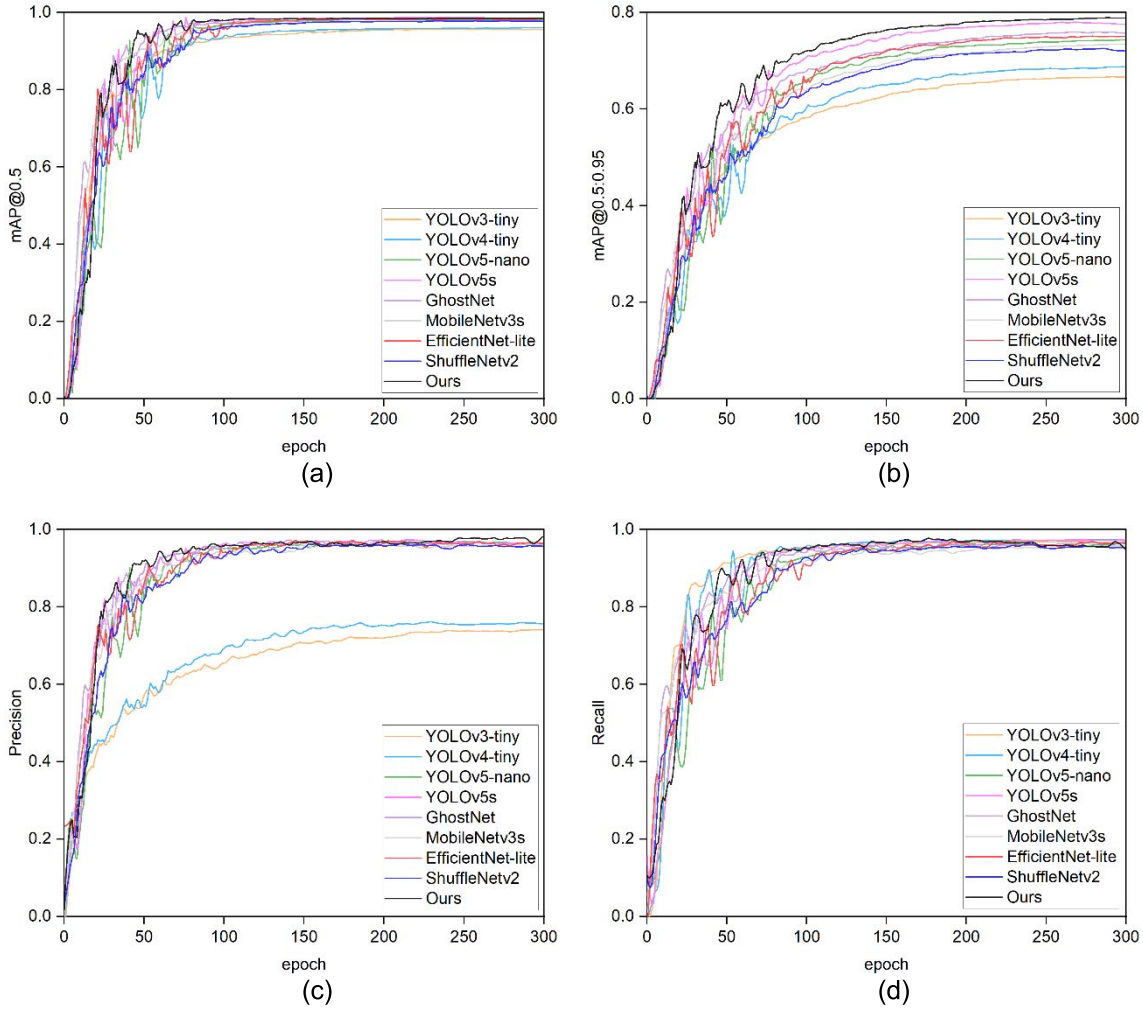

FIGURE 9. Comparison of Light-SDNet with other models on the Seaships dataset. (a)mAP@0.5; (b)mAP@0.5:0.95; (c)Precision; (d)Recall.

TABLE 3. Performance comparison on SeaShips dataset (mainly in YOLO series).

Methods	Precision	Recall	F_1 score	mAP@ .5:.95	CDR	FAR	Inference (ms)	Params (M)	FLOPs (G)	Train time(h)
YOLOv3-tiny	0.735	0.968	0.836	0.673	-	-	0.5	8.68	12.9	5.734
YOLOv4-tiny	0.763	0.969	0.854	0.696	-	-	0.7	5.89	16.16	5.732
YOLOv5-nano	0.957	0.957	0.957	0.729	0.985	0.043	1.3	1.77	4.2	5.887
YOLOv5s	0.955	**0.972**	0.963	0.761	0.987	0.045	1.6	7.24	16.5	5.868
Light-SDNet	**0.976**	0.957	**0.966**	**0.779**	**0.988**	**0.024**	2.0	4.93	12.9	6.843

10.6% compared with YOLOv3-tiny. Moreover, the size of Light-SDNet is only 4.93 MB, accounting for 68.1%, 83.7%, and 56.8% of YOLOv5s, YOLOv4-tiny and YOLOv3-tiny, respectively. As shown in Table 4, Light-SDNet achieves detection accuracy higher than other lightweight Backbone networks though its model size is not the least among the comparative models. The comparison also reveals that the detection accuracy of Light-SDNet is the highest for ship detection in adverse weather conditions. For the detection speed of the model, the inference time of Light-SDNet is 2.0 ms per image (500 fps) (fps, frames per second), indicating that Light-SDNet enables real-time ship detection. The detection precision and computational burden of comparative models are visualized in Figure 10. We can see that Light-SDNet achieves the cost-effective performance better than its competitors due to it using the multi-feature fusion and channel-spatial parallel attention mechanism for ship detection.

Table 5 shows the performance comparison between Light-SDNet with the hybrid training strategy and its

TABLE 4. Performance comparison on SeaShips dataset (mainly on the lightweight Backbone series).

Methods	Precision	Recall	F_1 score	mAP@ .5:.95	CDR	FAR	Inference (ms)	Params (M)	FLOPs (G)	Train time(h)
GhostNet	0.965	0.953	0.959	0.744	0.986	0.035	2.1	5.33	7.9	6.097
MobileNetv3s	0.955	0.955	0.955	0.718	0.984	0.045	0.8	3.54	6.1	5.877
EffiencientNet-lite	**0.977**	0.945	0.961	0.741	0.986	0.023	1.2	3.78	7.3	5.761
ShuffleNetv2	0.959	0.943	0.951	0.710	0.983	0.041	0.8	3.19	5.9	5.859
Light-SDNet	0.976	**0.957**	**0.966**	**0.779**	**0.988**	**0.024**	2.0	4.93	12.9	6.843

TABLE 5. Comparison of the different ship detection algorithms.

Ship(AP)	[41]	[51]	[47]	[49]	[54]	Ours
ore carrier	0.414	0.832	0.881	0.906	0.924	**0.989**
fishing boat	0.583	0.647	0.783	0.895	0.960	**0.986**
passenger ship	0.502	0.729	0.886	0.821	0.975	**0.982**
general cargo ship	0.387	0.932	0.917	0.880	**0.993**	0.989
bulk carrier	0.432	0.821	0.876	0.852	0.984	**0.990**
container ship	0.462	0.883	0.903	0.911	0.985	**0.995**
mAP@.5	0.487	0.791	0.874	0.878	0.970	**0.988**

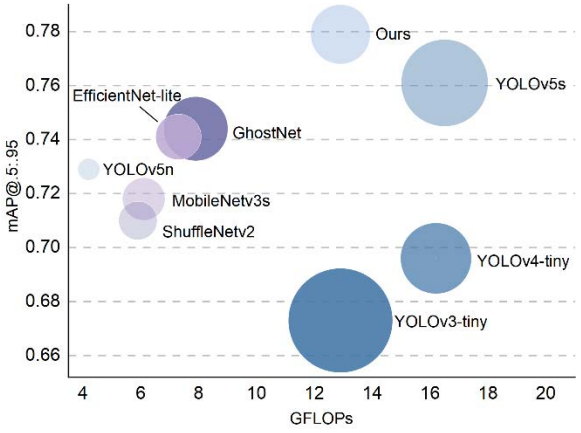

FIGURE 10. Comparison of the proposed Light-SDNet with other models. Each bubble's area represents the total number of parameters. GhostNet/EfficientNet-lite/MobileNetv3s/ShuffleNetv2 replaces the Back-bone of YOLOv5s, respectively.

competitors. We can observe that compared with its competitors, the proposed method achieves a marked improvement on the precision of ship detections in the maritime surveillance images. The reason is that Light-SDNet adopts the CA-Ghost module and the C3Ghost module guided by the attention mechanism, which achieves more fully shallow feature extraction and effective multiscale feature fusion. The hybrid training strategy is used to enhance the diversity of the training data and improve the robustness of Light-SDNet, which can further improve target detection accuracy.

E. QUALITATIVE EVALUATION

To qualitatively compare Light-SDNet with other models, we conduct experiments on synthetic sea fog dataset. Results are shown in Figures 11 and 12, where the rectangular boxes in Figures 11-12 mark the ships detected by different detectors. Special scenarios such as the simultaneous appearance of multiple ships, large overlapping areas of ships, small ships, and dense fog make it more difficult to detect ships, resulting in unreliable monitoring of maritime traffic.

Figures 11-12 indicate that most of comparative models achieves unsatisfactory results on ship detection under complex conditions due to missed detections and false detections occurring frequently. We can see that YOLOv3-tiny cannot accurately identify bulk carriers when multiple ships appear concurrently, and MobileNetv3s misidentifies bulk carriers as ore carriers. Vessel detection in severe weather conditions is a challenge. Bad weather markedly reduces the quality of the image captured by maritime surveillance systems. YOLOv3-tiny and YOLOv4-tiny cannot effectively detect ships in dense fog conditions due to it being sensitive to unstable imaging. Small ship detection is also a challenge. The size of small ships in the original image is relatively small, resulting in too few discriminative features. As a result, the detector cannot identify these small target ships accurately with blurred features after many convolutional layers. Water surface reflections and ocean waves also cause confusion and interfere with imaging due to the particular marine imaging scene, increasing the difficulty of feature extraction for small target ships. However, GhostNet and Light-SDNet show good performance for small target ship detection because the Ghost module embedded into the model compensates each other for channel information and retains more underlying information beneficial to the small target detection. As shown in Figures 11-12, other models except YOLOv4-tiny exhibit accurate detection of multiple ships with a suitable occlusion rate. In contrast, Light-SDNet can achieve robust detection of

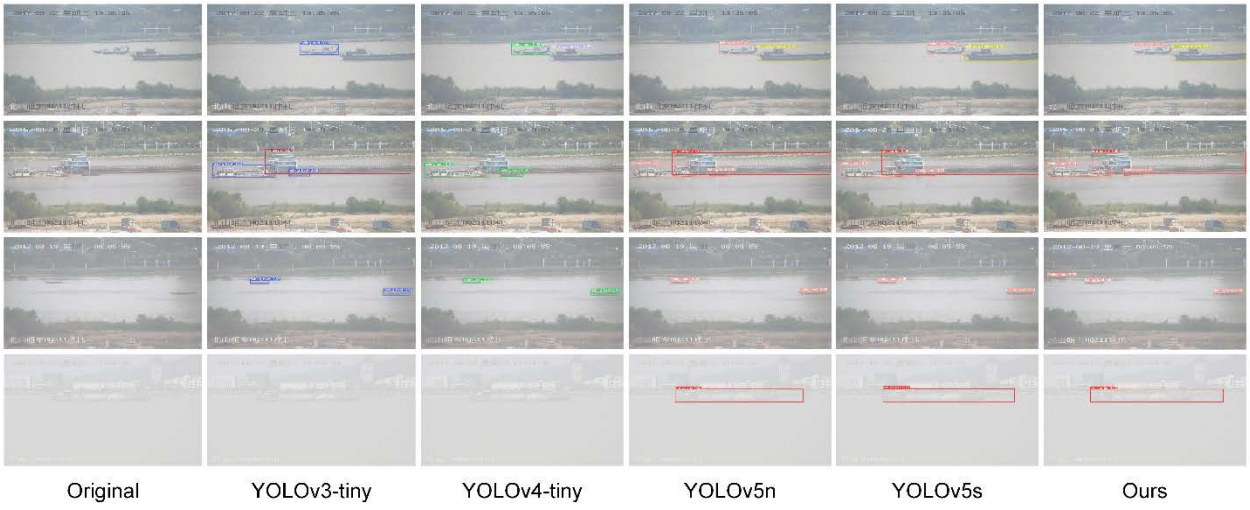

FIGURE 11. Visual comparison of the proposed Light-SDNet with the light YOLO family. From top to bottom: respectively represent detection of multiple ships, detection of ships with large overlapping areas, detection of small ships, and Ship detection in dense fog. The YOLOv3-tiny, YOLOv4-tiny, YOLOv5n, and YOLOv5s generate inaccurate detection results, while Light-SDNet can yield more satisfactory results.

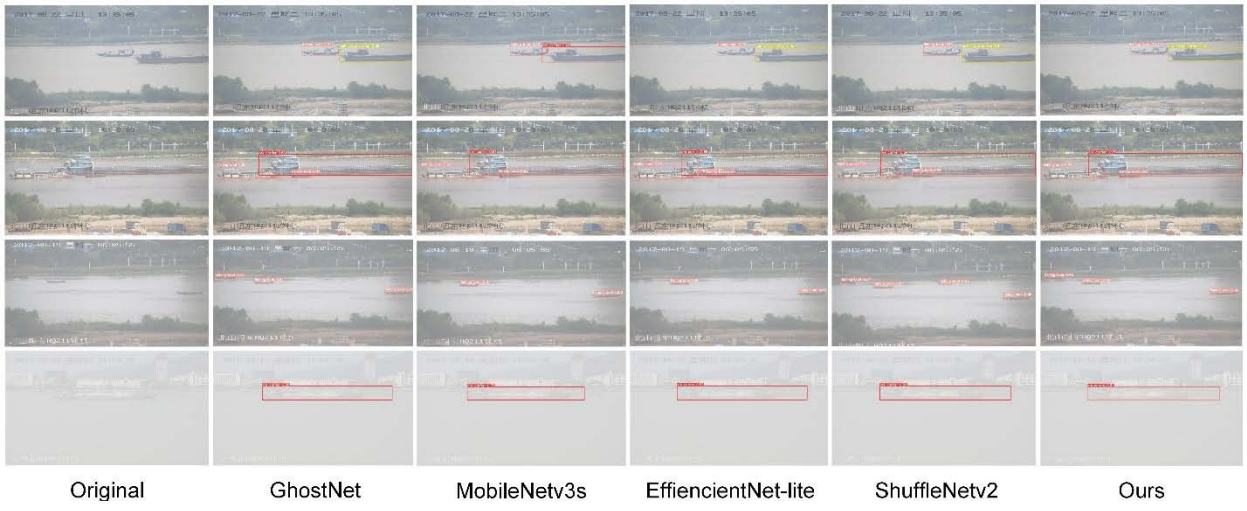

FIGURE 12. Visual comparison of the proposed Light-SDNet with four lightweight SOTA methods, in which Yolov5s's Backbone is replaced. From top to bottom: respectively represent detection of various ships, detection of ships with large overlapping areas, detection of small ships, and Ship detection in dense fog. The comparative four models can achieve accurate ship detection in dense fog conditions, while they suffer from false or missed detections in other complex environments.

moving ships under various surveillance conditions, providing strong support for maritime surveillance systems.

F. ABLATION STUDY

To find the effectiveness and efficiency of the proposed strategy, we conducted the different experiments on the components of the Light-SDNet architecture and the proposed hybrid training strategy. Details of the experimental formulations of the Light-SDNet architecture are presented in Table 6. We can observe that the original YOLOv5s yields the lowest mAP, resulting in unsatisfactory detection results. The improvement of detection accuracy brought by DWConv is not obvious, while the computational burden is reduced dramatically. As shown in Table 6, applying Ghost and C3Ghost in YOLOv5s can improve the detection accuracy markedly due to the optimization of feature maps. Moreover, the results indicate that the appropriate increase of network Depth provides 0.6% improvement in mAP performance. The comparison also shows that coordinate attention and parallel attention can improve the detection accuracy of the original YOLOv5s. Further, Light-SDNet improves the mAP performance by 1.8% compared with the original YOLOv5s. Thus, the proposed framework improves detection performance markedly by integrating multiple functional modules with YOLOv5s and appropriately increasing network depth.

To describe the impact of different types of ships on model performance, the PR curve and F1-score curve of Light-SDNet are shown in Figure 13 (a) and (b), respectively. It can

TABLE 6. Comparison of ship detection performance of different modules.

Methods	AP						mAP@ .5:.95	Precision	Recall	F_1 score	FLOPs(G)
	ore carrier	fishing boat	passenger ship	general cargo ship	bulk carrier	container ship					
YOLOv5s	0.731	0.716	0.712	0.798	0.784	0.825	0.761	0.955	0.972	0.963	16.5
+ Ghost	0.742	0.712	0.721	0.809	0.787	0.822	0.765	0.966	0.957	0.961	13.5
+ DWConv	0.736	0.704	0.719	0.810	0.777	**0.834**	0.763	0.966	0.960	0.963	12.3
+ C3Ghost	0.742	0.713	0.725	0.813	0.794	0.825	0.769	0.968	0.959	0.963	10.4
+ Depth	0.747	0.714	**0.743**	0.816	0.801	0.830	0.775	0.951	**0.971**	0.961	12.6
+ CA(c)	**0.763**	0.716	0.740	**0.820**	0.794	0.831	0.777	0.965	0.965	0.965	12.7
+ parallel attention	0.759	**0.719**	**0.743**	0.812	**0.810**	0.831	**0.779**	**0.976**	0.957	**0.966**	12.9

TABLE 7. Comparison of ship detection performance of the hybrid data training strategy.

Methods	AP						mAP@ .5:.95	Precision	Recall	F_1 score
	ore carrier	fishing boat	passenger ship	general cargo ship	bulk carrier	container ship				
Light-SDNet	0.759	0.719	0.743	0.812	0.810	0.831	0.779	0.976	0.957	0.966
+ hybrid strategy	**0.806**	**0.780**	**0.821**	**0.851**	**0.847**	**0.845**	**0.825**	**0.98**	**0.982**	**0.981**

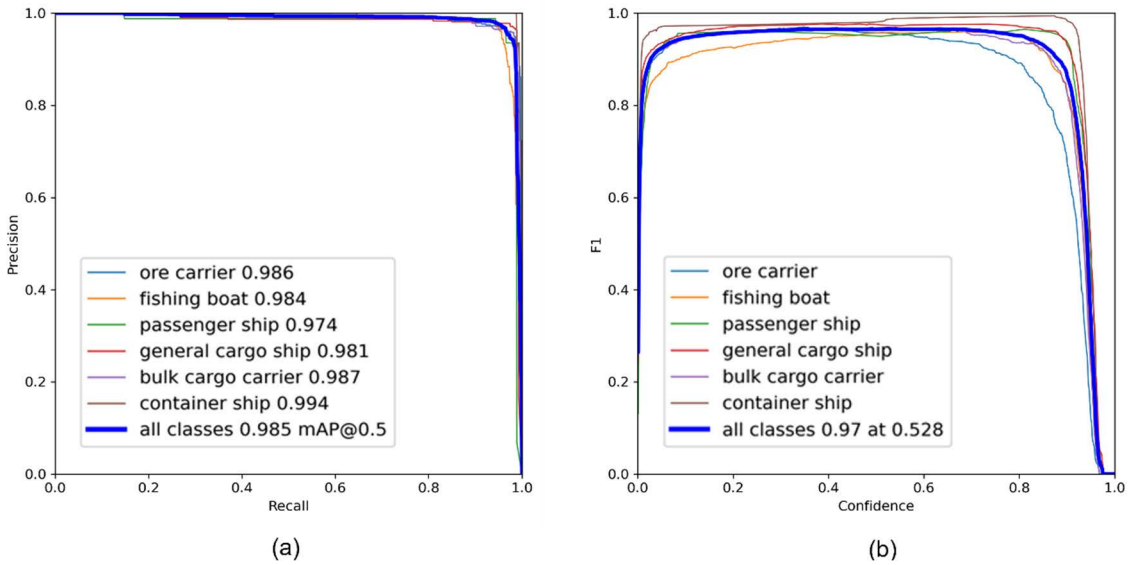

FIGURE 13. Effects of different types of ships on the performance of Light-SDNet: (a) The PR curve; (b) F1 score curve.

be derived from Figure 13 (a) that Light-SDNet is an optimal detector since it reserves high precision values along with increased Recall rate. Figure 13 (b) shows that Light-SDNet acquire the highest F1 score on the container ship images due to the salient characteristics of container ships, meanwhile, it keeps reasonable F1 score on all classes of ship images as confidence increases, especially on small fishing ships. Thus, the proposed detector possesses good generalization performance.

To validate the effectiveness of the proposed hybrid training strategy, the performance comparison of Light-SDNet with and without the proposed training method were carried out and the results are shown in Table 7. We can observe that the AP index of each type of ships is significantly improved, the recall rate is increased by 2.5%, and the mAP is increased by 4.6%. Thus, the comparison indicates that the hybrid training data strategy can improve the detection performance of the proposed detector markedly.

To examine the robustness of the proposed method, we conducted the experiments in different imaging environments, and results are shown Figure 14. It can be derived that the proposed detector can detect the moving ships accurately even if the visual quality is degraded markedly by bad weather conditions, due to the fact the proposed hybrid training

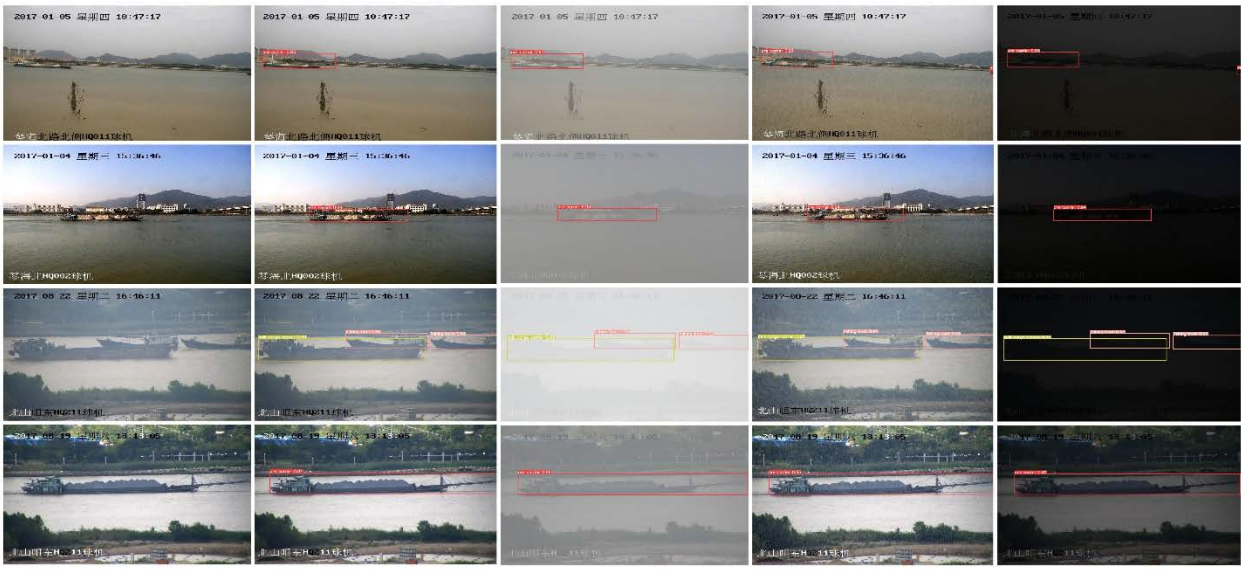

FIGURE 14. Detection results of the proposed method (i.e., Light-SDNet and the hybrid data training strategy) under different imaging conditions.

strategies with synthetically-degraded images could improve the diversity of training datasets markedly. Thus, the learning and generalization abilities of Light-SDNet are improved in practice. It can be derived from Figure 14 that the proposed method can achieve reliable, efficient, and accurate ship detection under poor imaging conditions. The reliable detection of Light-SDNet contributes to tracking maritime objects, and detecting abnormal behavior, leading to enhanced management in the intelligent maritime surveillance systems.

V. CONCLUSION

In this study, we proposed a lightweight CNN framework and a hybrid training strategy for ship detection. The proposed network makes full use of the shallow location features via introducing CA-Ghost module to improve the feature extraction capability of the Backbone. And the C3Ghost module guided by the attention mechanism has been introduced in the Neck network to achieve more effective feature fusion. In addition, we presented the hybrid training strategy to enhance the diversity of the training data and improve the robustness of Light-SDNet in the adverse weather conditions. Compared with the recently proposed SOTA models, our method achieves a balance between model complexity and detection accuracy and can detect different types of moving ships in real time with high detection accuracy. Extensive experimental results have demonstrated good detection performance of Light-SDNet under adverse weather conditions, such as hazy, rainy, and low-light conditions. This study can be extended in the following directions to make ship detection more reliable and robust.

(1) The proposed hybrid data training strategy directly synthesizes degraded images by using a simplified image generation model. However, the synthetic ocean images differ from real ones in terms of the color and structure. The next step will focus on the generation of more realistic degraded images.

(2) Accurate detection of small moving ships in a maritime surveillance system is still challenge for Light-SDNet. It is hard for monitoring camera situated at a distance from the ships to capture high-resolution maritime images, thereby leading to unreliable detection in terms of robustness and accuracy. We will promote small-scale ship detection via increasing detection Heads for small target objects [55].

(3) Bad weather typically affects the quality of the images captured by the marine surveillance systems, which causes the difficulty for accurate multi-ship detection. Thus, there is potential for imagery data combined with oceanographic radar technology to detect and classify multiple targets [56].

Although the proposed method has huge space to further improve its performance, it is still worth exploring as it can realize the accurate detection of moving ships rapidly under severe weather conditions while remaining its lightweight nature, thereby achieving a better balance between the detection accuracy and model size. Light-SDNet has the potential to be putted in practical applications to enhance maritime safety and management.

REFERENCES

[1] *Technical Characteristics for An Automatic Identification System Using Time-Division Multiple Access in the VHF Maritime Mobile Band*, Standard ITU-R M.1371, Feb. 2014. [Online]. Available: http://www.itu.int/rec/R-REC-M.1371/en

[2] L. Jiao, F. Zhang, F. Liu, S. Yang, L. Li, Z. Feng, and R. Qu, "A survey of deep learning-based object detection," *IEEE Access*, vol. 7, pp. 128837–128868, 2019.

[3] Z.-Q. Zhao, P. Zheng, S.-T. Xu, and X. Wu, "Object detection with deep learning: A review," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 11, pp. 3212–3232, Nov. 2019.
[4] W. Liu, D. Anguelov, and D. Erhan, "SSD: Single shot MultiBox detector," in *Proc. Eur. Conf. Comput. Vis.*, Oct. 2016, pp. 21–37.
[5] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788.
[6] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 7263–7271.
[7] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*.
[8] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.
[9] J. Glenn, S. Alex, and B. Jirka. (2021). *Ultralytics/YOLOv5: V6.0 (Version v6.0)*. [Online]. Available: http://doi.org/10.5281/zenodo.63715
[10] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 580–587.
[11] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.
[12] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards realtime object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.
[13] Y. Wang, X. Ning, B. Leng, and H. Fu, "Ship detection based on deep learning," in *Proc. IEEE Int. Conf. Mechatronics Autom. (ICMA)*, Aug. 2019, pp. 275–279.
[14] Y. You, J. Cao, Y. Zhang, F. Liu, and W. Zhou, "Nearshore ship detection on high-resolution remote sensing image via scene-mask R-CNN," *IEEE Access*, vol. 7, pp. 128431–128444, 2019.
[15] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. 36th Int. Conf. Mach. Learn. (ICML)*, May 2019, pp. 6105–6114.
[16] A. Howard, M. Sandler, B. Chen, W. Wang, L.-C. Chen, M. Tan, G. Chu, V. Vasudevan, Y. Zhu, R. Pang, H. Adam, and Q. Le, "Searching for MobileNetV3," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 1314–1324.
[17] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*.
[18] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4510–4520.
[19] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.
[20] X. Zhang, X. Zhou, M. Lin, and J. Sun, "ShuffleNet: An extremely efficient convolutional neural network for mobile devices," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6848–6856.
[21] N. Ma and X. Zhang, "ShuffleNet V2: Practical guidelines for efficient CNN architecture design," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 122–138.
[22] K. Han, Y. Wang, Q. Tian, J. Guo, C. Xu, and C. Xu, "GhostNet: More features from cheap operations," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 1577–1586.
[23] X. Nie, M. Yang, and R. W. Liu, "Deep neural network-based robust ship detection under different weather conditions," in *Proc. IEEE Intell. Transp. Syst. Conf. (ITSC)*, Auckland, NZ, USA, Oct. 2019, pp. 47–52.
[24] G. Graffieti and D. Maltoni, "Artifact-free single image defogging," *Atmosphere*, vol. 12, no. 5, p. 577, Apr. 2021.
[25] H. Zhang, V. Sindagi, and V. M. Patel, "Image de-raining using a conditional generative adversarial network," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 11, pp. 3943–3956, Nov. 2020.
[26] L. Chen, L. Guo, D. Cheng, and Q. Kou, "Structure-preserving and color-restoring up-sampling for single low-light image," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 4, pp. 1889–1902, Apr. 2022.
[27] M. Lin, Q. Chen, and S. Yan. *Network in Network*, 2013, *arXiv:1312.4400*.
[28] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Apr. 2016, pp. 770–778.
[29] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 936–944.
[30] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "mixup: Beyond empirical risk minimization," 2017, *arXiv:1710.09412*.
[31] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, Jan. 2014.
[32] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path aggregation network for instance segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8759–8768.
[33] C.-Y. Wang, H.-Y. M. Liao, Y.-H. Wu, P.-Y. Chen, J.-W. Hsieh, and I.-H. Yeh, "CSPNet: A new backbone that can enhance learning capability of CNN," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 390–391.
[34] D. Misra, "Mish: A self regularized non-monotonic activation function," 2019, *arXiv:1908.08681*.
[35] P. Ramachandran, B. Zoph, and Q. V. Le, "Searching for activation functions," 2017, *arXiv:1710.05941*.
[36] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, and D. Ren, "Distance-IoU loss: Faster and better learning for bounding box regression," in *Proc. AAAI Conf. Artif. Intell.*, Feb. 2020, vol. 34, no. 7, pp. 12993–13000. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view 999
[37] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-Net: Efficient channel attention for deep convolutional neural networks," in *Proc. IEEE CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2020, pp. 11531–11539.
[38] J. Park, S. Woo, J.-Y. Lee, and I.-S. Kweon, "BAM: Bottleneck attention module," in *Proc. British Mach. Vis. Conf. (BMVC)*, 2018, p. 147.
[39] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 3–19.
[40] Q. Hou, D. Zhou, and J. Feng, "Coordinate attention for efficient mobile network design," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 13713–13722.
[41] Y. Zhang, "Ship detection for visual maritime surveillance from non-stationary platforms," *Ocean Eng.*, vol. 141, pp. 53–63, Sep. 2017.
[42] G. Shi and J. Suo, "Ship targets detection based on visual attention," in *Proc. IEEE Int. Conf. Signal Process., Commun. Comput. (ICSPCC)*, Sep. 2018, pp. 1–4.
[43] Z. Chen, B. Li, L. F. Tian, and D. Chao, "Automatic detection and tracking of ship based on mean shift in corrected video sequences," in *Proc. 2nd Int. Conf. Image, Vis. Comput. (ICIVC)*, Jun. 2017, pp. 449–453.
[44] Z. Cui, X. Wang, N. Liu, Z. Cao, and J. Yang, "Ship detection in large-scale SAR images via spatial shuffle-group enhance attention," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 1, pp. 379–391, Jan. 2021.
[45] X. Ma, K. Ji, B. Xiong, L. Zhang, S. Feng, and G. Kuang, "Light-YOLOv4: An edge-device oriented target detection method for remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 10808–10820, 2021.
[46] Q. Li, L. Mou, Q. Liu, Y. Wang, and X. X. Zhu, "HSF-Net: Multiscale deep feature embedding for ship detection in optical remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 12, pp. 7147–7161, Dec. 2018.
[47] Z. Shao, L. Wang, Z. Wang, W. Du, and W. Wu, "Saliency-aware convolution neural network for ship detection in surveillance video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 3, pp. 781–794, Mar. 2020.
[48] X. Chen, Y. Yang, S. Wang, H. Wu, J. Tang, J. Zhao, and Z. Wang, "Ship type recognition via a coarse-to-fine cascaded convolution neural network," *J. Navigat.*, vol. 73, no. 4, pp. 813–832, 2020.
[49] R. W. Liu, W. Yuan, X. Chen, and Y. Lu, "An enhanced CNN-enabled learning method for promoting ship detection in maritime surveillance system," *Ocean Eng.*, vol. 235, Sep. 2021, Art. no. 109435.
[50] B. Wang and F. Huang, "A lightweight deep network for defect detection of insert molding based on X-ray imaging," *Sensors*, vol. 21, no. 16, p. 5612, Aug. 2021.
[51] Z. Shao, W. Wu, Z. Wang, W. Du, and C. Li, "SeaShips: A large-scale precisely annotated dataset for ship detection," *IEEE Trans. Multimedia*, vol. 20, no. 10, pp. 2593–2604, Oct. 2018.
[52] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and W. Zisserman, "The PASCAL visual object classes (VOC) challenge," *Int. J. Comput. Vis.*, vol. 88, no. 2, pp. 303–338, Sep. 2010.
[53] W. Liu, G. Ren, R. Yu, S. Guo, J. Zhu, and L. Zhang, "Image-adaptive YOLO for object detection in adverse weather conditions," 2021, *arXiv:2112.08088*.

[54] H. Li, L. Deng, C. Yang, J. Liu, and Z. Gu, "Enhanced Yolo v3 tiny network for real-time ship detection from visual image," *IEEE Access*, vol. 9, pp. 16692–16706, 2021.
[55] L. Zhu, X. Geng, Z. Li, and C. Liu, "Improving YOLOv5 with attention mechanism for detecting boulders from planetary images," *Remote Sens.*, vol. 13, no. 18, p. 3776, Sep. 2021.
[56] K. Kim, J. Kim, and J. Kim, "Robust data association for multi-object detection in maritime environments using camera and radar measurements," *IEEE Robot. Autom. Lett.*, vol. 6, no. 3, pp. 5865–5872, Jul. 2021.

MENGYAO ZHANG received the B.S. degree from the Department of Physics, Harbin Normal University, Harbin, Heilongjiang, China, in 2019, where she is currently pursuing the M.S. degree with the School of Physics and Electronic Engineering. Her current research interests include computer vision, image processing, and object detection.

XIANWEI RONG received the B.S. degree from the Department of Physics, Harbin Normal University, Harbin, Heilongjiang, China, in July 1996, and the M.E. degree from the School of Information and Communication, Harbin Engineering University, in 2010. He is currently a Professor with Harbin Normal University. His current research interests include image processing and machine learning.

XIAOYAN YU (Member, IEEE) received the B.S. and M.Ed. degrees from Harbin Normal University, Harbin, Heilongjiang, China, in 1998 and 2001, respectively, and the Ph.D. degree from the Department of Information Systems Engineering, Kochi University of Technology, Kochi, Japan, in 2006. She is currently a Professor with Harbin Normal University. Her current research interests include machine learning and computer vision.