Marine Vessel Detection in Sea Fog Environment Based on SSD

Yuanyuan Wang[1], Ning Wang[2(✉)], Luyuan Tang[1], and Wei Wu[1]

[1] School of Marine Electrical Engineering, Dalian Maritime University, Dalian 116026, China
[2] School of Marine Engineering, Dalian Maritime University, Dalian 116026, China
n.wang@ieee.org

Abstract. Aiming at solving the problem of low marine vessel detection accuracy in the sea fog environment, a deep learning-based anti-fog marine vessel detection method is proposed in this paper by combining defogging preprocessing with marine vessel detection model. Firstly, gated context aggregation network (GCANet) network is used to process the marine vessel image. Then, the processed image is sent to a modified SSD network, wherein anchors are tuned by statistical characteristics of the shape of marine vessel to detect the position of the marine vessel. Furthermore, to alleviate the loss of feature information due to defogging processing, channel attention mechanism based on the squeeze and excitation module (SE) is added to base convolutional layer of SSD. The comprehensive experiments and comparison results show that the proposed G-SEMSSD network is more suitable for marine vessel detection under sea fog environment.

Keywords: Deep learning · Marine vessel detection · Fog image object detection

1 Introduction

Marine safety including shipwreck, naval battle and marine accidents, *etc.*, is still a serious matter [1]. As the autonomous marine platform, underwater robotics [2,3] and unmanned surface vehicles (USVs) that can conduct long-term and large-scale marine operations have gain more attention in the fields of intelligent control [4], collision avoidance [5] and trajectory tracking [6,7]. With increasingly rapid development of computer vision and deep learning, more intelligent USVs [8] are desired to achieve high-effect marine object detection, recognition and tracking.

By far, marine vessel detection methods are roughly summarized into two kinds, i.e., traditional methods based on prior information and modern methods

This work is supported by the National Natural Science Foundation of China (Grant 52271306) and Innovative Research Foundation of Ship General Performance (Grant 31422120).

H. R. Karimi and N. Wang (Eds.): S-Cube 2022, LNICST 487, pp. 49–62, 2023.
https://doi.org/10.1007/978-3-031-34899-0_4

based on deep learning. The former strongly requires manual features such as contour, radiation, shape and moment invariants, *etc.*, thereby leading to time-consuming and empirical interventions. State-of-the-art (SOTA) deep learning-based object detectors mainly consist of two-stage and one-stage approaches. The two-stage methods generate a series of candidate boxes as samples which are classified by virtue of the convolution neural networks (CNN) [9–11]. For instance, R-CNN [12,13], SPP-net [14], Fast R-CNN [15] and Faster R-CNN [16], *etc.* The latter techniques, also named end-to-end algorithms, which mean that input an image to the network and then the probabilities and locations information of targets in the image are directly obtained. Therefore, compared to two-stage approaches, one-stage approaches are faster. The typical methods include YOLO [17] and its variants (YOLOs), i.e., YOLOv2 [18], YOLOv3 [19], YOLOv4 [20], YOLOv5, RetinaNet [21], SSD [22] and EfficientDet D0-D6 [23]. Particularly, SSD scheme features the advantages of anchors and regression from both Fast R-CNN and YOLOs, and thereby it not only reaches real-time detection but does not sacrifice accuracy.

USVs navigate in complex and unknown marine environments [24–26], such as wind turbine, ocean wave and currents, sea fog and rain, which inevitably leads to degradation of acquired image. However, deep learning methods largely depends on high quality data, for which deep learning-based marine vessel detection methods as described above are facing great challenges. At present, there is no systematic and efficient methods of marine vessel recognition to defend against the extreme weathers. Naturally, increasing related works based on CNN for dehazing algorithms have been produced, since great progresses have been made in some tasks owing to implementations of CNN. By far, the mainstreams of image defog algorithms can be roughly summarized into three categories, i.e., image enhancement, image restoration, and deep learning-based defog methods. Image enhancement algorithms aim to restore a clear image without fog by removing image noise and improving image contrast, such as histogram equalization (HE), adaptive histgram equalization (AHE), contrast limited adaptive histgram equalization (CLAHE) [27], Retinex [28] and filtering algorithms, *etc.* The second usually depends on atmospheric scattering model. For example, dark channel prior (DCP) [29], color attenuation prior (CAP) [30]. It should be noted that the second tends to outperform the former. Deep learning-based imaging defogging is to input foggy image to CNN and directly output the defoggy image. For example, DehazeNet [31], all-in-one network (AOD-Net) [32], densely connected pyramid dehazing network (DCPDN) [33], gated fusion network (GFN) [34] and gated context aggregation network (GCANet) [35]. The end-to-end networks for image dehazing are easily to be embedded into other CNN models mentioned above. Sequentially, some comprehensive results and comparisons among various defogging algorithms mentioned above are conducted in paper [36], which concluded that Faster R-CNN trained on the image processed by defogging algorithms increases detection accuracy by a range of 0.2% to 2.0%, except for DehazeNet. In conclusion, not all of defoggy algorithms are absolutely benefit to object detection, but most.

In this paper, SOTA defogging algorithm GCANet is firstly utilized to process images, and then the clean images are used to fine-tune SSD model to satisfy the specific detection task. GCANet is a defogging network based on CNN and it not only can remove the fog in a single image but has little effect on clear images. Furthermore, we have modified and optimized the SSD network: 1) designing more reasonable aspect ratio (AR) by the statistical characteristics of marine vessel; 2) inserting channel attention mechanism based on SE block to the base convolutional layer of SSD, to alleviate the loss of key features due to defogging processing. In this context, fusing GCANet, AR and SE techniques, the G-SEMSSD detection method is eventually established to defend against sea fog environment. Main contributions of this paper are summarized as follows:

* Adding an image preprocessing module named GCANet to the SSD scheme, both detection precision and recall have been significantly enhanced.
* Formulating new default boxes of marine vessel with statistical characteristics of MVDD13 dataset, the detection accuracy is further improved.
* Inserting the channel attention mechanism SE block into the base convolutional layer, the loss of key features caused by defogging processing is dramatically decreased.

The remainder of this paper is organized as follows. Section 2 briefly introduces related works on marine vessel model, including image processing module, SSD model and optimization. In Sect. 3, the comprehensive experimental results and comparisons are given and analyzed. Conclusions are laid in Sect. 4.

2 Marine Vessel Detection Model

There are many abbreviations are utilized in the paper, for convenience, the commonly used words are summarized in Table 1.

Table 1. The full names and responding abberviations included in paper.

Abbreviation	Full Name
AR	aspect ratio
SE	squeeze and excitation module
GCANet	gated context aggregation network
SSD	single shot multibox detector
MVDD13	marine vessel detection dataset including 13 categories
AP	IoU = 0.5 and for each category
mAP@.5	IoU = 0.5 and averaged across 13 categories
mAP@.75	IoU = 0.75 and averaged across 13 categories
mAP@[.5 : .95]	averaged across 10 IoU thresholds and 13 categories

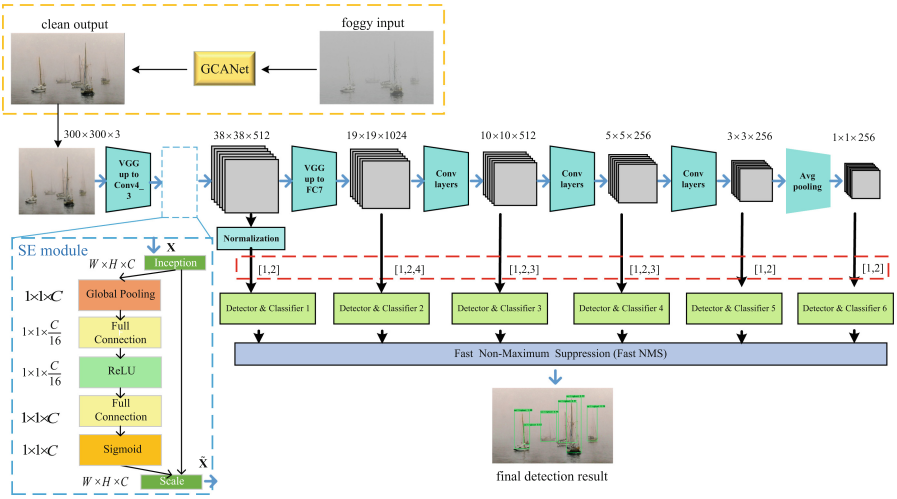


Fig. 1. The flow chart of defogging preprocessing and detection.

As shown in Fig. 1, the flow chart of marine vessel detection in foggy environment is divided into two steps: GCANet is firstly used to defog the marine vessel image. Then, the processed image is input to the modified SSD network to detect the position of marine vessel.

2.1 Image Processing Module

GCANet was proposed by Chen et al. [35]. It applies the smooth dilated convolution to extract image features, and the residual between clean and foggy image features are calculated to remove fog, instead of reling on prior information.

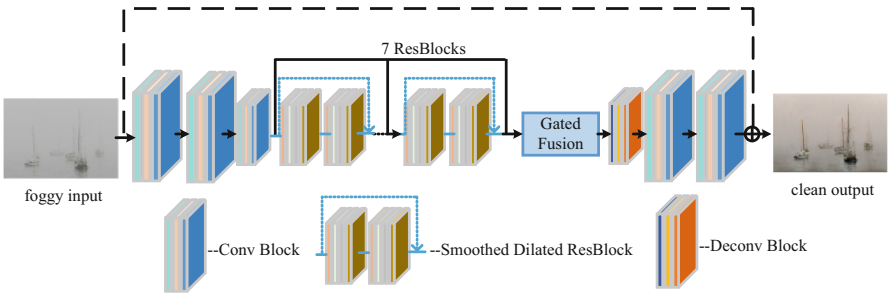


Fig. 2. Structure of GCANet.

The structure of GCANet and the defogging processing are shown in Fig. 2. Firstly, three convolutional layers are used to encode the foggy image, and only

the last one downsamples the feature maps by 1/2 once. Then seven smoothed dilated resblocks are used to aggregate image features. These features from different level are fused into a gate fusion subnetwork, which is used to calculate the residual between the foggy and the fog-free image. During the defogging processing, the residuals are adaptively adjusted by different concentrations of fog in the input image. In decoder part, one deconvolutional layer with stride 1/2 is used to upsample the feature maps to the original resolution, then the following two convolutional layers convert the feature maps back to the image space to get the final target fog residue.

Fig. 3. Comparison of defogging effect.

During the runtime, the GCANet will predict the residue between the target clean image and the hazy input image in an end-to-end way. Specifically, Fig. 3 shows the effective of GCANet, whereby the first raw is original foggy images and the below is defoggy images.

2.2 SSD Model and Optimization

By far, Faster R-CNN, YOLOs and SSD are SOTA deep learning models, which have been widely applied in object detection field. According to our previous work, SSD with a 300×300 input size significantly outperforms SOTA object detector counterparts as stated in [22], especially in terms of speed.

Default Box. The SSD framework is vividly shown in Fig. 1. VGG16 is selected as its base network, of which fc6 and fc7 are converted to two convolutional layers and the following three scale-different feature layers predict the offsets to default boxes of different scales, aspect ratios and their associated confidences.

For each bounding box, the relationships between its width (w), height (h), scale (s) and aspect ratio (α_r) are expressed as following:

$$w \cdot h = s, \qquad \frac{w}{h} = \alpha_r \tag{1}$$

According to bounding box of each target annotated in dataset, α_r can be computed. In SSD scheme [22], aspect ratios are designed as $\alpha_r \in \left\{1, 2, 3, \frac{1}{2}, \frac{1}{3}\right\}$, and then the width and height for each default box can be computed as following:

$$w = s\sqrt{\alpha_r}, \qquad h = \frac{s}{\sqrt{\alpha_r}} \tag{2}$$

Suppose that m feature maps are utilized for prediction. For each feature map $k \in [1, m]$, the scale of default boxes is computed as:

$$\begin{cases} s_k = s_{min} + \dfrac{s_{max} - s_{min}}{m-1}(k-1), \alpha_r \in \left\{2, 3, \dfrac{1}{2}, \dfrac{1}{3}\right\}, \\ s_k, and \sqrt{s_k s_{k+1}}, \qquad\qquad \alpha_r = 1 \end{cases} \tag{3}$$

where $s_{min} = 0.2$ and $s_{max} = 0.9$, meaning the lowest layer has a scale of 0.2 and the highest layer has a scale of 0.9, and all layers in between are regularly spaced.

The center of each default boxes is $\left(\dfrac{i+0.5}{|f_k|}, \dfrac{j+0.5}{|f_k|}\right)$, where $|f_k|$ is the size of the k-th square feature map, $i, j \in [0, |f_k|)$.

Channel Attention Module-SE Block. Due to low contrast and brightness of the foggy image captured in complex and changeable marine environment, the edges of targets in image are often blurred and the texture features are not obvious. As shown in Fig. 1, SE module is inserted to the base convolutional layer (conv4-3) and learns the importance of each feature channel through the attention mechanism. Previous works have shown that using feature maps from the lower layers can strengthen feature extraction, since more details of objects are contained in lower layers. The computation processing is summaried in **Algorithm 1**. In this context, the ability of the network to aggregate effective features or feature extraction will be greatly improved.

3 Experiential Results and Analysis

In order to demonstrate the effectiveness and superiority of the proposed G-SEMSSD model, we set groups of comparisonal experiments. More details of experimental procedure are described in the following subsections.

3.1 Experimental Setups

All experiments are conducted on a Linux PC platform with an Intel Core i7-8700K CPU, and a single NVIDIA TITAN Xp GPU (12GB). In addition, both training and testing environments are built in Ubuntu 18.04 operation system running CUDA 10.0 and Python 3.7. The stochastic gradient descent with the initial learning rate $2 \cdot 10^{-3}$ and ending learning rate 10^{-4} is employed. In addition, 50 and 150 epochs are set in freezing and full training stages.

Algorithm 1: Channel attention mechanism SE

Input: feature map $\mathbf{X} \in \mathbb{R}^{\acute{W} \times \acute{H} \times \acute{C}}$, where width, height and channel: $\acute{W}$, $\acute{H}$ and $\acute{C}$

1. /***Inception****/ convolutional transformation:$f(\mathbf{X}) = \mathbf{U} \in \mathbb{R}^{W \times H \times C}$

2. /***Global pooling****/ calcute mean of feature map: $z = \frac{1}{H \times W} \sum_{i=1}^{W} \sum_{j=1}^{H} u(i,j)$

3. /***Full connection****/ $fcl1 = \mathbf{W}_1\mathbf{z}$, $\mathbf{W}_1 \in \mathbb{R}^{\frac{C}{r} \times C}$, where $r = 16$

4. /***ReLU****/ $R = max(0, fcl)$

5. /***Full connection****/ $fcl2 = \mathbf{W}_2\mathbf{R}$, $\mathbf{W}_2 \in \mathbb{R}^{C \times \frac{C}{r}}$, where $r = 16$

6. /***sigmoid****/ $s = \frac{1}{\mathbf{e}^{fcl2}}$

7. /***Scale****/

Output: $\tilde{\mathbf{X}} = s \cdot u$

3.2 Experimental Dataset

In our previous work, we have established a marine vessel dataset and named MVDD13. By virtue of LabelImg tool, objects are annotated by 13 categories, i.e., *cargo*, *passenger*, *cruise*, *bulker*, *tanker*, *sailingboat*, *tug*, *fishing*, *drill*, *fire-fighting*, *containership*, *warship* and *submarine*, and their number distribution in MVDD13 and corresponding pinyin abbreviation are shown in Table 2.

In all experiments, 25,541, 2,838 and 7,095 images are used for training, validation and testing sets. The techniques including transfer learning and fine-tuning are implemented on pretrained SSD on PASCAL VOC0712.

3.3 Evaluation Metrics

In the following experiments, the evaluation metrics are used to measure the performance of marine vessel detection models.

Intersection over Union (IoU). The IoU is defined by

$$\text{IoU} = \frac{area(B_p \bigcap B_{gt})}{area(B_p \bigcup B_{gt})} \tag{4}$$

where B_p and B_{gt} are predicted and ground-truth bounding boxes, respectively.

Average Precision (AP). Given an IoU threshold, the Recall and Precision can be determined accordingly. For each ship type, a Precision-Recall (P-R) curve can be governed under an IoU threshold. In this context, the AP defines the area surrounded by the P-R curve $P(R)$ as follows:

$$\text{AP} = \int_0^1 P(R)dR \tag{5}$$

where P and R are Precision and Recall, respectively.

Table 2. Number of objects of each marine vessel category.

Category	Short of Name	Objects	Percentage
cargo	zhc	7,640	0.1871
passenger	kc	5,685	0.1392
cruise	lyl	5,307	0.1299
bulker	shc	5,270	0.1290
tanker	yl	4,589	0.1124
sailingboat	fc	3,830	0.0938
tug	tc	2,733	0.0669
fishing	yc	1,427	0.0349
drill	qzc	1,336	0.0327
firefighting	sjc	1,278	0.0313
containership	jzxc	1,074	0.0263
warship	zc	420	0.0103
submarine	qt	250	0.0061

Mean Average Precision (mAP). Intuitively, the mAP measures the average value of AP over all C categories to be detected, and evaluates the overall performance of a detector by the following equation:

$$\text{mAP} = \sum_{i=1}^{C} \text{AP}_i \Big/ C \tag{6}$$

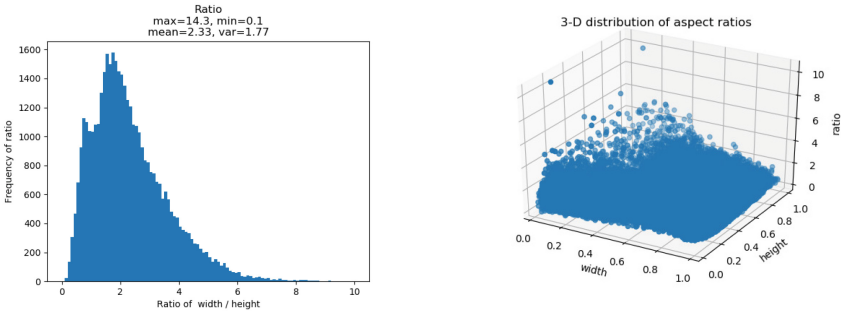
where AP_i is the AP of i-th category and C is the number of categories.

3.4 Statistical Characteristics-Based Anchors

In SSD scheme [22], aspect ratios are designed as $\alpha_r \in \left\{1, 2, 3, \frac{1}{2}, \frac{1}{3}\right\}$. It should be noted that the size and shape of prior boxes are closely related with the data, i.e., the ratios of width and height of targets need to be accurately calculated.

We can see from Fig. 4(a) and Fig. 4(b) that the distribution characteristics of aspect ratios is high in the middle and low on both sides. Similar to the normal distribution that the mean $\mu = 2.33$ and variance $\sigma^2 = 1.77$, about 68.2% of targets are in the ratio interval $[\mu - \sigma, \mu + \sigma]$, i.e., interval $[1, 4]$. In other words, most of marine vessel in MVDD13 dataset features long and narrow side. In different feature map layers of SSD scheme, the aspect ratios can be set as different values. By conducting experiments on MVDD13 dataset, the aspect ratios of each feature map layer has been reasonably set, and the specific details in the revised MSSD scheme are clearly shown in Fig. 1 and Table 3.

To thoroughly illustrate that the superiority and effectiveness of the proposed MSSD scheme, the typical SSD framework is utilized to make fair comparison,



(a) Distribution of aspect ratios (b) 3D distribution of aspect ratios

Fig. 4. Distributions of aspect ratio in MVDD13 dataset.

Table 3. Comparison of detection results.

Backbone	Model	mAP@.5 (%)	Feature-map layer					
			1	2	3	4	5	6
VGG16	SSD	88.47	[1, 2]	[1, 2, 3]	[1, 2, 3]	[1, 2, 3]	[1, 2]	[1, 2]
	MSSD	**92.17**	[1, 2]	[1, 2, **4**]	[1, 2, 3]	[1, 2, 3]	[1, 2]	[1, 2]

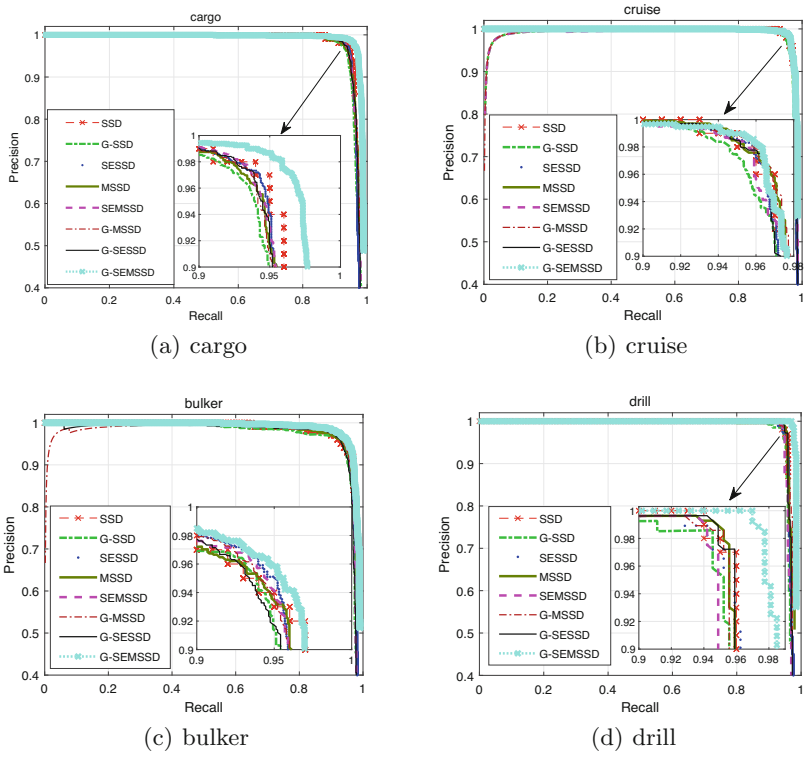
and experimental result is summarized in Table 3, whereby the VGG16 is used as backbone, from which we can clearly see that the developed MSSD scheme is 3.7% higher than SSD framework in mAP@.5. It is reasonable to conclude that the use of proper default boxes from the statistical characteristics of data can enhance detection performance of SSD to some extent. This sufficiently illustrated that fine-tunning aspect ratios is truely good to detection for SSD model.

3.5 Ablation Studies and Analysis

To thoroughly illustrate that the superiority and effectiveness of the proposed model in more details, we carried out seven controlled experiments to examine how each component affects performance. The developed models are as follows:

(1) The original SSD module trained on defoggy image by GCANet (i.e., G-SSD),
(2) The default boxes of SSD are designed according to MVDD13 dataset(i.e., MSSD),
(3) The SE module is inserted into SSD convolutional layer (i.e., SESSD),
(4) The SE module is inserted into M-SSD convolutional layer (i.e., SE-MSSD),
(5) The MSSD module trained on defoggy image by GCANet (i.e., G-MSSD),
(6) The SESSD module trained on defoggy image by GCANet (i.e., G-SESSD),
(7) The SE-MSSD module trained on defoggy image by GCANet (i.e., G-SEMSSD).

Different models generated by various design choices on SSD together with corresponding comparisons of performance in terms of mAP@.5, are shown in Table 4.



(a) cargo (b) cruise

(c) bulker (d) drill

Fig. 5. Detection results of (a) cargo (b) cruise (c) bulker and (d) drill.

In order to sufficiently demonstrate the superiority of GCANet, the training set processed by GCANet is leveraged to train models including SSD, MSSD, SESSD and SE-MSSD, i.e., G-SSD, G-MSSD, G-SESSD and G-SEMSSD. We can see that models trained on training set with GCANet are mostly better than those without GCANet, apart from G-SESSD. The reason might be that there is a conflict in the preservation of image features between GCANet and SE. In fact, as depicted in previous work [37], there is no exact theory to prove that the image after dehazing preprocessing must benefit the detection task.

Compared to original SSD, G-SSD, MSSD and SESSD models surprisingly promote by 1.00%, 3.70% and 3.96% mAP, respectively. It is reasonable to conclude that the three techniques can individually enhance detection performance of SSD to some extent. Specifically, MSSD is significantly more accurate due to the use of proper default boxes from the statistical characteristics of data. Furthermore, G-MSSD improves 3.1% mAP over G-SSD. This sufficiently

Table 4. Detection results of different models for each category.

Model	Technique			mAP @.5 (%)	Marine vessel category											
	G-CANet	AR	SE		zhc	kc	lyl	shc	yl	fc	tc	yc	qzc	sjc	jzxc	zc	qt
SSD				88.47	96.22	90.18	97.09	96.56	94.54	87.85	83.71	75.55	96.39	89.35	83.46	69.45	89.82
G-SSD	✓			89.47	96.83	90.87	97.58	96.33	96.58	89.96	84.23	76.25	96.19	90.14	82.85	78.52	86.75
MSSD		✓		92.17	97.06	92.92	98.20	96.91	**97.40**	90.22	86.67	80.73	97.24	94.04	87.90	85.22	93.73
SESSD			✓	92.43	96.92	92.74	98.16	97.23	97.30	89.71	87.48	81.16	97.14	93.71	88.56	87.17	94.36
SE-MSSD		✓	✓	92.14	97.22	93.00	98.06	96.92	97.26	89.22	86.06	80.36	96.45	95.02	88.29	86.38	93.58
G-SESSD	✓		✓	91.79	97.14	93.18	98.17	96.43	97.16	89.84	86.54	78.47	96.99	94.42	87.97	83.73	93.20
G-MSSD	✓	✓		92.57	97.00	92.97	98.20	96.93	97.33	90.35	87.07	80.49	96.90	95.44	88.37	87.49	94.82
G-SEMSSD	✓	✓	✓	**94.32**	**98.46**	**94.40**	**98.22**	**98.19**	97.19	**94.27**	**89.07**	**86.17**	**98.43**	**97.38**	**90.10**	**88.81**	**95.51**

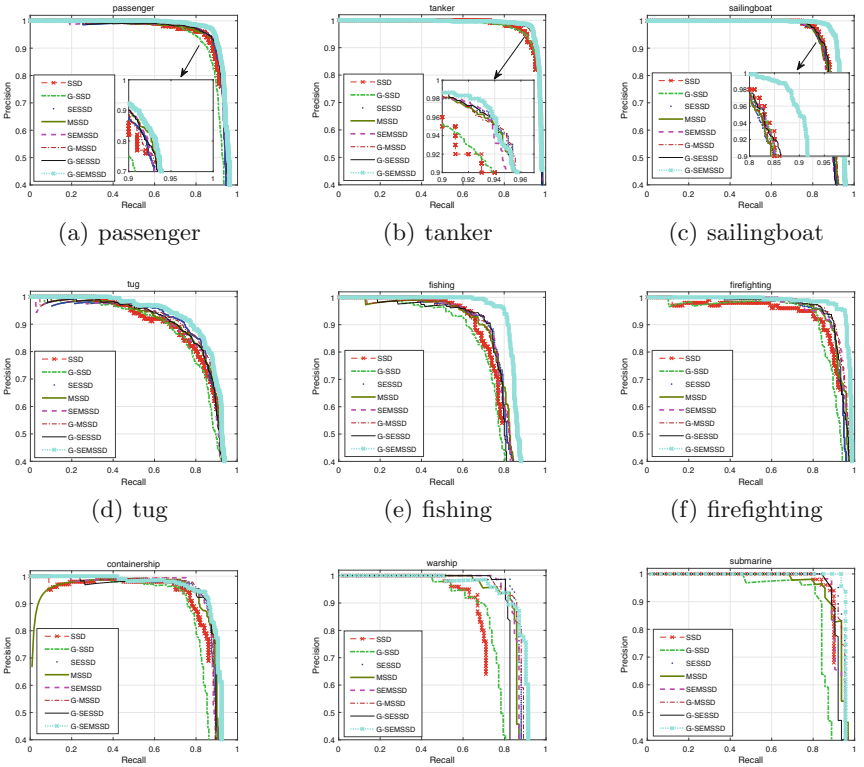
* Numbers in bold indicate the best results of each category.

illustrated that tunning aspect ratios is truely good to detection, except for similar effects between SESSD and SE-MSSD.

As shown in Table 4, G-SESSD has a significant superiority over G-SSD. Promising results are shown in G-SEMSSD model which gains a dramatic performance 94.32% mAP. To further validate the superiority of SE bolock, the comparisons between SESSD (92.43) and SSD (88.47), SE-MSSD (92.14) and MSSD (92.17), and G-SEMSSD (94.32) and G-MSSD (92.57) are designed. Ignoring the basically same results between SE-MSSD and MSSD, we can clearly see that SSD models with SE block perform really well. In this context, it implies that inserting SE block can effectively enhance the detection performance.

For each category, the effect of different models can be vividly shown in Fig. 5 and 6, wherein the P-R curves with comparisons are based on IOU with threshold 0.5. It should be emphasized that, in some range of precision (here, precision is greater than 0.4), the higher recall the better performance. In this context, for the curves distributed densely in the upper right corners, we enlarge them. We can clearly observe that by using GCANet, AR and SE strategies, the G-SEMSSD model achieves remarkably superior trade-off between precision and recall for all categories. Especially in *caro, drill, sailingboat* and *fishing*, the G-SEMSSD is far superior compared to other models. It should be noticed that these ship types own salient shape feature, i.e., width (height) is greatly more than height (width), fitting better to the new designed default boxes. In addition, we find that *cruise* performs best and robust for each model, since it not only features better aspect ratio, but is less affected by fog.

In summary, the proposed G-SEMSSD model performs remarkable superiority in detection accuracy, and it can be applied in sea fog environment.



(a) passenger (b) tanker (c) sailingboat

(d) tug (e) fishing (f) firefighting

(g) containership (h) warship (i) submarine

Fig. 6. Detection results of (a) passenger (b) tanker (c) sailingboat (d) tug (e) fishing (f) firefighting and (g) containership (h) warship (i) submarine.

4 Conclusion

In this paper, we propose a marine vessel detection method G-SEMSSD to defend against sea fog environment. By using GCANet defogging preprocessing method, a high accuracy of recognition can be achieved not only in sunny day but in foggy environment. Moreover, the new default boxes elaborately designed by statistical characteristics of data are benefit for detection task. Furthermore, inserting SE module has been devised to significantly strengthen extraction of key features in complex marine environments. Eventually, the G-SEMSSD detection method is established. Comprehensive experiments and comparisons have demonstrated that the model can achieve more accurate detection performance for marine vessel detection in sea fog environment.

References

1. Wang, N., Wang, Y., Er, M.J.: Review on deep learning techniques for marine object recognition: architectures and algorithms. Control. Eng. Pract. **118**(3), 104458 (2022)
2. Huang, H., Zhou, H., Yang, X., Zhang, L., Qi, L., Zang, A.: Faster R-CNN for marine organisms detection and recognition using data augmentation. Neurocomputing **337**, 372–384 (2019)
3. Chen, T., Wang, N., Wang, R., Zhao, H., Zhang, G.: One-stage CNN detector-based benthonic organisms detection with limited training dataset. Neural Netw. **144**, 247–259 (2021)
4. Wang, N., Gao, Y., Yang, C., Zhang, X.: Reinforcement learning-based finite-time tracking control of an unknown unmanned surface vehicle with input constraints. Neurocomputing **484**, 26–37 (2022)
5. Huang, Y., Chen, L., Chen, P., Negenborn, R.R., Van Gelder, P.H.A.J.M.: Ship collision avoidance methods: state-of-the-art. Saf. Sci. **121**, 451–473 (2020)
6. Wang, N., Qian, C., Sun, J., Liu, Y.: Adaptive robust finite-time trajectory tracking control of fully actuated marine surface vehicles. IEEE Trans. Control Syst. Technol. **24**(4), 1454–1462 (2016)
7. Wang, N., Er, M.J.: Direct adaptive fuzzy tracking control of marine vehicles with fully unknown parametric dynamics and uncertainties. IEEE Trans. Control Syst. Technol. **24**(5), 1845–1852 (2016)
8. Wang, N., Karimi, H.R., Li, H., Su, S.-F.: Accurate trajectory tracking of disturbed surface vehicles: a finite-time control approach. IEEE/ASME Trans. Mechatron. **24**(3), 1064–1074 (2019)
9. Lecun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proc. IEEE **86**(11), 2278–2324 (1998)
10. Krizhevsky, A., Sutskever, I., Hinton, G. E.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, vol. 25 (2012)
11. Russakovsky, O., et al.: Imagenet large scale visual recognition challenge. Int. J. Comput. Vision **115**(3), 211–252 (2015)
12. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Region-based convolutional networks for accurate object detection and segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **38**(1), 142–158 (2016)
13. van de Sande, K.E.A., Uijlings, J.R.R., Gevers, T., Smeulders, A.W.M.: Segmentation as selective search for object recognition. In: IEEE International Conference on Computer Vision, pp. 1879–1886 (2011)
14. He, K., Zhang, X., Ren, S., Sun, J.: Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Trans. Pattern Anal. Mach. Intell. **37**(9), 1904–1916 (2015)
15. Girshick, R.: Fast R-CNN. In: IEEE International Conference on Computer Vision, pp. 1440–1448 (2015)
16. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Trans. Pattern Anal. Mach. Intell. **39**(6), 1137–1149 (2017)
17. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: unified, real-time object detection. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 779–788 (2016)

18. Redmon, J., Farhadi, A.: Yolo9000: better, faster, stronger. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 6517–6525 (2017)
19. Redmon, J., Farhadi, A.: Yolov3: an incremental improvement, arXiv preprint arXiv:1804.02767 (2018)
20. Bochkovskiy, A., Wang, C, Liao, H.: Yolov4: optimal speed and accuracy of object detection. arXiv preprint, arXiv:2004.10934 (2020)
21. Lin, T., Goyal, P., Girshick, R., He, K., Dollar, P.: Focal loss for dense object detection. IEEE Trans. Pattern Anal. Mach. Intell. **99**, 2999–3007 (2017)
22. Liu, W., et al.: SSD: single shot MultiBox detector. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9905, pp. 21–37. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46448-0_2
23. Tan, M., Pang, R., Le, Q.V.: EfficientDet: scalable and efficient object detection. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10778–10787 (2020)
24. Wang, N., Er, M.J., Sun, J., Liu, Y.: Adaptive robust online constructive fuzzy control of a complex surface vehicle system. IEEE Trans. Cybern. **46**(7), 1511–1523 (2016)
25. Wang, X., Zhang, L., Heath, W.P.: Wind turbine blades fault detection using system identification-based transmissibility analysis. Insight-Non-Destructive Test. Condition Monit. **64**(3), 164–169 (2022)
26. Wang, N., Er, M.J.: Self-constructing adaptive robust fuzzy neural tracking control of surface vehicles with uncertainties and unknown disturbances. IEEE Trans. Control Syst. Technol. **23**(3), 991–1002 (2015)
27. Zuiderveld, K.: Contrast limited adaptive histogram equalization. In: Graphics Gems, pp. 474–485 (1994)
28. Rahman, Z., Jobson, D., Woodell, G.: Retinex processing for automatic image enhancement. J. Electron. Imaging **13**(1), 100–110 (2004)
29. He, K., Sun, J., Tang, X.: Single image haze removal using dark channel prior. IEEE Trans. Pattern Anal. Mach. Intell. **33**(12), 2341–2353 (2010)
30. Zhu, Q., Mai, J., Shao, L.: A fast single image haze removal algorithm using color attenuation prior. IEEE Trans. Image Process. **24**(11), 3522–3533 (2015)
31. Cai, B., Xu, X., Jia, K., Qing, C., Tao, D.: Dehazenet: an end-to-end system for single image haze removal. IEEE Trans. Image Process. **25**(11), 5187–5198 (2016)
32. Li, B., Peng, X., Wang, Z., Xu, J., Feng, D.: An all-in-one network for dehazing and beyond. arXiv preprint arXiv:1707.06543 (2017)
33. Zhang, H., Patel, V.M.: Densely connected pyramid dehazing network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3194–3203 (2018)
34. Ren, W., et al.: Gated fusion network for single image dehazing. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3253–3261 (2018)
35. Chen, D., et al.: Gated context aggregation network for image dehazing and deraining. In: 2019 IEEE Winter Conference on Applications of Computer Vision, pp. 1375–1383 (2019)
36. Li, C., Guo, C., Guo, J., Han, P., Fu, H., Cong, R.: PDR-Net: perception-inspired single image dehazing network with refinement. IEEE Trans. Multimedia **22**(3), 704–716 (2020)
37. Chen, X., Lu, Y., Wu, Z., Yu, J., Wen, L.: Reveal of domain effect: How visual restoration contributes to object detection in aquatic scenes. arXiv preprint arXiv:2003.01913 (2020)