Indian Journal of Geo Marine Sciences
Vol. 50 (11), November 2021, pp. 960-968

Sea-surface object detection scheme for USV under foggy environment

T Zhang[a], X Liu*[,b], Y Li[b] & M Zhang[b]

[a]College of Information Science and Engineering, Chongqing Jiaotong University, Chongqing – 400 074, China

[b]College of Mechanical and Electrical Engineering, Harbin Engineering University, Harbin – 150 001, China

*[E-mail: liuxing20080724@gmail.com]

Received 31 August 2021; revised 30 November 2021

Sea-surface target detection is investigated for the visual image-based autonomous control of an Unmanned Surface Vessel (USV). A traditional way is to dehaze for sea-surface images in the previous target detection algorithms. However, it would cause a problem that the image dehaze performance and detection speed are difficult to be balanced. To solve the above problem, a YOLO (You Only Look Once) based target detection network with good anti-fog ability is proposed for sea-surface target detection. In this proposed method, the target detection network is trained off-line to obtain a good anti-fog ability and the target detection is performed on-line. A hazed sample generation model is built based on atmospheric single scattering inverse model to obtain sufficient samples for the off-line training in the proposed method. And then, the target detection network is trained based on the generated samples to obtain good anti-fog ability according to a new learning strategy. Finally, comparative experimental results demonstrate the effectiveness of the proposed target detection algorithm.

[**Keywords**: Anti-fog enhancement, Detection accuracy, Sea-surface target, Target detection network]

Introduction

Sea-surface target detection is one of the core technologies of autonomous control of an Unmanned Surface Vessel (USV)[1-3]. In the sea-surface environment, the images collected by the vision system of USV are usually foggy. At present, target detection of foggy images always includes two steps, *i.e.*, dehaze at first and then target detection[4]. There are mainly two methods to process hazed images at present[5,6], including dark channel prior and neural network.

An image dehaze method is proposed based on dark channel prior[7], and then some improved versions of the dark channel prior algorithms, *e.g.*, guided filtering method, were proposed successively in He *et al.*[8]. These algorithms have good performances in terms of dehaze, but it is not suitable for visual image processing for USVs due to the real-time requirement. Cai *et al.*[9] proposed the first deep learning-based dehaze neural network named Dehaze-Net. In this method, images are dehazed by integrating the propagation and mapping mode of dehaze medium into the network structure, which presented good dehaze performance, but images cannot be real-time processed. A lightweight dehaze network named AOD-Net (All-in-One Dehazing Network) is proposed[10]. In addition, Wang *et al.*[11] proposed the SC-R-CNN (Scene Classification Region Convolutional Neural Network) to detect objects in foggy weather. This algorithm was applied to satellite remote sensing images of sea surface on foggy weather to test the performance of the developed algorithm.

YOLO (You Only Look Once), as the representative of the one-stage algorithm, has good detection speed[12,13]. A target detection algorithm was provided by combining with DenseNet and YOLOv3 to improve the detection accuracy for USVs under different environment conditions[14]. An improved YOLOv3 based sea-surface target detection algorithm was proposed for USVs[15], which can provide a good balance between the detection speed and detection accuracy. In order to improve the YOLO network to obtain good anti-fog ability, but not to add extra computation burden, a common way is to find suitable and sufficient samples to train the YOLOv3 network. However, most of the samples used for sea-surface target detection are fog-free. In addition, sea-surface images are affected by many factors, including fog concentration, different target objects. As a result, it is difficult to find suitable and sufficient training samples for the YOLOv3 with good anti-fog ability.

According to the above analysis, a YOLO based target detection network with good anti-fog ability is developed for autonomous control of USVs under foggy environment so as to obtain good balance between the detection accuracy and detection speed, where the target detection network is trained to off-line obtain the anti-fog ability first and then target detection is performed on-line. In order to overcome the problem of the insufficient samples, a hazed sample generation model is built based on atmospheric single scattering inverse model, where sea-surface images under fog-free condition are hazed and transformed into the samples with different visibilities. Then, the target detection network is trained based on the generated samples to obtain good anti-fog ability, where a new learning strategy is obtained by the single target detection network YOLOv2, and then this strategy is used to train the YOLOv3 with multi-scale mapping target detection network, to overcome the difficulties in anti-fog enhancement learning in the YOLOv3. Finally, comparative experiments are performed on many hazed images to verify the advantages of the developed scheme in terms of detection accuracy in comparison with other traditional schemes.

Materials and Methods

Procedures of the proposed target detection method

At present, dehaze preprocessing is always used for the target detection in a sea-surface image obtained under fog condition. However, the balance between the dehaze performance and detection speed should be considered when adopting this strategy to detect sea-surface target. Specifically, the dehaze method with fast processing speed has poor dehaze effect, and vice versa. In order to meet the real-time requirements of autonomous control of USVs and obtain satisfactory detection accuracy, a YOLO based target detection network with good anti-fog ability is developed in this paper, shown as Figure 1.

The description about the procedures is given as follows:

(1) The samples with different visibilities are generated through images without fog under the action of a suitable model.

(2) The target detection network YOLOv3 with multi-scale mapping has good detection performance. However, since YOLOv3 has a complex structure and strict requirements on training samples, it is difficult to obtain good learning strategy if directly training YOLOv3 with the samples. In this paper, YOLOv2 with relatively simple structure is trained off-line according to the generated samples to obtain a satisfactory learning strategy, and then the strategy is also used to train YOLOv3 off-line.

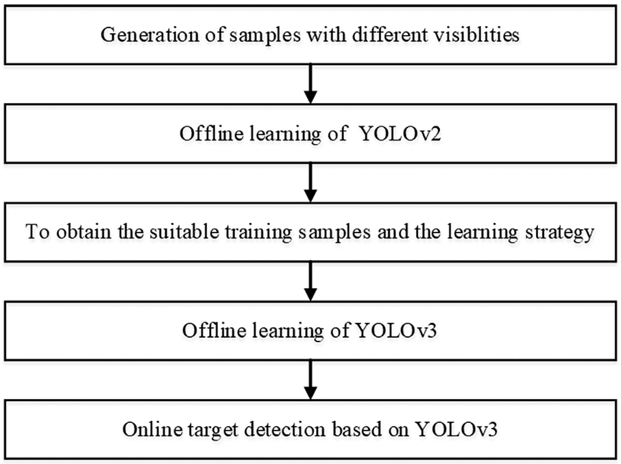
Generation of samples with different visiblities
Offline learning of YOLOv2
To obtain the suitable training samples and the learning strategy
Offline learning of YOLOv3
Online target detection based on YOLOv3

Fig. 1 — Basic procedures of the proposed method

(3) YOLOv3 after training is used to on-line detect surface target according to the new images.

Generation of training samples

The existing methods to generate fog-scene images includes confrontation neural network, virtual scene construction (or special simulator), and atmospheric single scattering model. Among these methods, the atmospheric single scattering model can establish the mapping relationship between fog-free and fog images with known 3D information by constructing the imaging relationship of light source, object, medium and camera[16].

Here, to obtain the samples to train the target detection network to have anti-fog ability, a fog-scene image generation method based on the atmospheric single scattering model combined with the structure characteristics of sea-surface images. At first, it expounds the principle of generating fog-scene images by the atmospheric single scattering model and then the approximate distance of each pixel relative to the shooting position is obtained by analyzing the specific structure characteristics. Finally, the distance information is used to generate fog-scene images in different contrast ratio, treated as samples.

(1) Atmospheric single scattering model

The atmospheric single scattering model is expressed by the following equations[9,17]

$$I(x)= J(x)t(x) + A_{\infty}\,(1- t(x)) \quad \ldots (1)$$

$$t(x) = \exp(-\beta(\lambda)\, d(x))t(x) \quad \ldots (2)$$

Where, $I(x)$ is the hazed image collected by the camera; $J(x)$ the real image to be recovered; $t(x)$ is the medium transmission; A_{∞} is the global atmospheric light; $\beta(\lambda)$ is scattering coefficient; $d(x)$ is distance from the scene point to the camera; λ is wavelength of light; x is pixels in the observed image $I(x)$.

From the Eqs. 1 – 2, it can be seen that t(x) is only related with pixels and scattering coefficient. The global atmospheric light can be estimated from adaptation through the overall brightness of the samples. And the scattering coefficient is only related with the size of the scattered particles in the atmosphere and usually is considered as a constant in the model. If the distance of each scene point in an image relative to the camera is obtained, the hazed picture of the specified visibility can be generated through the fog-free image.

(2) Estimated distance of each pixel

In this paper, we manually select 1466 images with typical features of sea-surface environment from the MSCOCO database[18], and most of the sea-surface images are taken on the ground with the average height being 3 m. Then the distance from camera to the sea-sky boundary line can be estimated by sea-sky region segmentation. The angle between the optical axis and the sea level can be calculated according to the camera shooting height h and the distance from camera to the sea-sky boundary line. Finally, the distance d(x) from the scene point to the camera is obtained by structural geometric relation, which will be introduced into the model Eq. 2 to generate the hazed images with the specified visibility.

(3) Generation of fog-scene samples with different visibilities

According to the atmospheric single scattering model and the estimated distance of each pixel to the camera, we can generate the hazed images with different visibilities. Some of them are shown in Figure 2, where the readings of the scale VR represent visibility with unit m.

(4) Determination of appropriate visibility to training detection network

To obtain the satisfactory training performance, it is necessary to set the upper and lower limits of visibility during the generation of hazed images. Now inputting the hazed images into the detection network YOLOv2, the following results are obtained. For the case with the visibility being 300 m, the detection confidence rate of YOLOv2 decreases by 12 % in average, at this time, the corresponding haze degree is light. For the case with the visibility being 200 m, the detection confidence rate of YOLOv2 decreases by 35 % in average, being the medium haze degree. For the case with the visibility being 100 m, the detection confidence rate of YOLOv2 decreases by 86 % in average, with high haze degree. In this paper, we use the atmospheric single scattering model to generate the haze images with visibilities being 300 m, 200 m

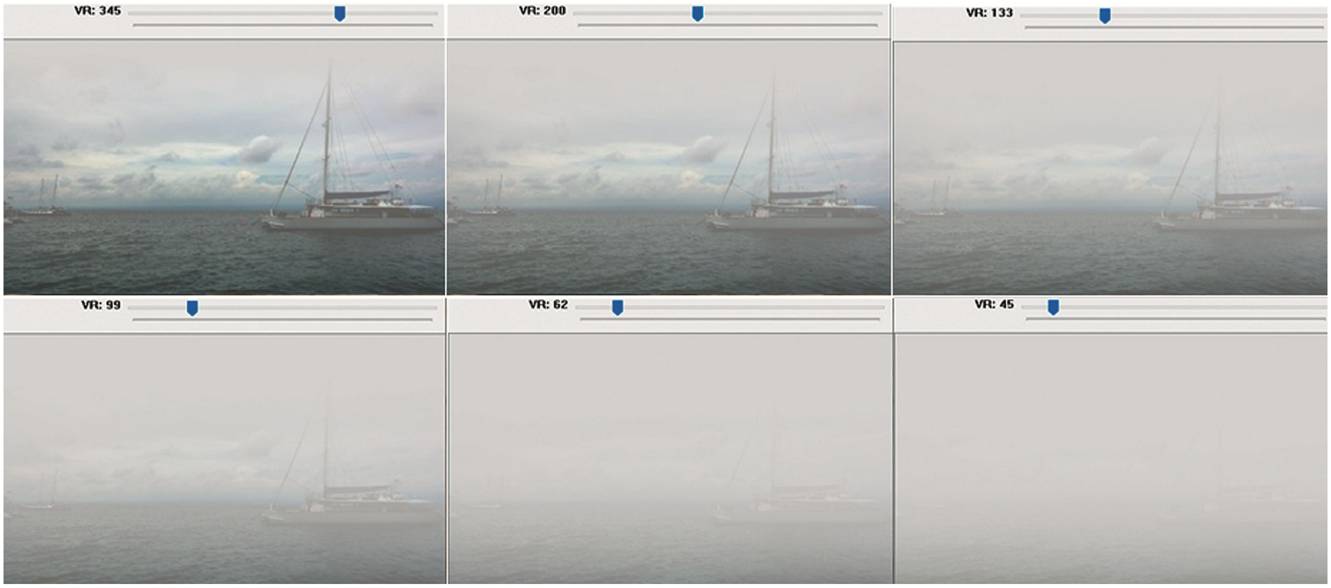

Fig. 2 — Hazed images with different visibilities

and 100 m, respectively. Part of the generated images are shown in Figure 3.

Learning strategy

This section is to present a learning strategy of target detection network. Specifically, we input the generated samples with different visibilities into the YOLOv2 network, to obtain a basic learning strategy. And then, the learning strategy is used to train the YOLOv3 network to acquire better detection performance.

Simple description of target detection network

The target detection technology based on deep learning can be divided into two categories, including signal target detection network and multiple target detection networks. The single target detection network can directly infer the target category and location without regional optional steps. For the multiple target detection networks, it has good detection performance, but requires more time, which does not adapt to the mobile platform such as unmanned ships with high real-time requirement[19]. Among the single target detection neural networks, the YOLO and its variant and SSD (Single Shot MultiBox Detector) have the best detection accuracy and real-time performance[20,21]. Considering the accuracy and complexity of the model, YOLO is selected as the detection network. YOLOv3 has more complex model structure and inference rules, but the detection accuracy is also better, compared with YOLOv2. In this paper, the basic learning strategy is obtained by using the YOLOv2 network with less parameters, and then the learning strategy is used to the YOLOv3 with more parameters to achieve the better anti-fog ability.

Learning strategy of target detection network YOLOv2

In this subsection, the target detection network YOLOv2 is trained according to the hazed images to obtain a learning strategy.

(1) Brief introduction of YOLOv2

The output of the YOLOv2 used in this paper is a C23 feature map with a shape of 19×19×425. Each point on the C23 feature map has five anchor frames[22]. The five anchor frames are obtained by

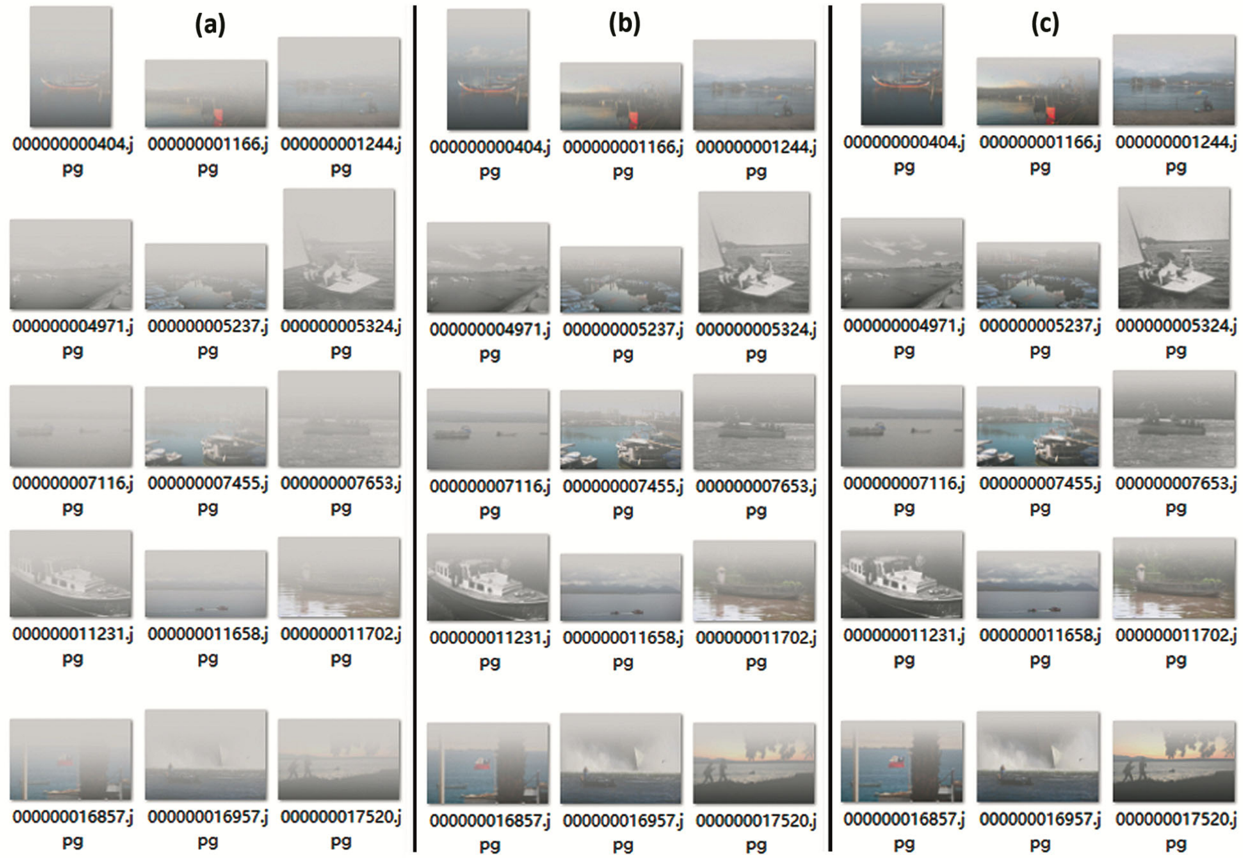

Fig. 3 — Part of hazed samples for training: a) Part of sample pictures in 100 m visibility, b) Part of sample pictures in 200 m visibility, and c) Part of sample pictures in 300 m visibility

clustering among all the sizes of the samples. When the object falls into the center of the 19×19 feature map, one of the five anchor frames that are most similar to the shape of the object is responsible for the position prediction of the object. The object position offset is the deviation corresponding to the anchor frame. According to the output probability resulted from the C23 feature map, the network filters out the overlapping target detection box by using the non-maximum suppression method, and then the detection box with confidence probability higher than the threshold is regarded as the final output result.

(2) Determination of training samples

In order to make the target detection network have the satisfactory anti-fog ability, different combinations of sea-surface images with different visibilities are used to train the target detection network, so as to obtain an optimal learning strategy.

By using the target detection network YOLOv2 which has been initially trained by the ordinary images, the sea-surface images with 100 m visibility, 200 m visibility and 300 m visibility are used to train the YOLOv2, respectively. From the experiment results, it is found that the training sample in 100 m visibility can have gradient explosion. Therefore, the sea-surface images with 200 m visibility and 300 m visibility are selected to train the YOLOv2.

(3) Learning strategy and results

To get the satisfactory training performance, different learning strategies are designed as follows. The first two strategies to train the YOLOv2 only according to the sea-surface images with 200 m visibility and 300 m visibility, respectively. The third one is to train the YOLOv2 by the samples in 300 m visibility at first and then trained by the samples in 200 m visibility, which is also called as gradient learning strategy. The last one is to train the YOLOv2 by the samples in 300 m visibility, 200 m visibility, and their corresponding haze-free images, which is also denoted as mixed learning strategy.

Here, two indexes, including average precision (mAP) and recall, are used to evaluate the performance of the detection network[16]. The training results of the four different learning strategies are shown in Table 1.

Table 1 show that the mixed learning strategy has the best detection performance in terms of “mAP” and “Recall”, in comparison with the other learning strategies. Therefore, this paper also selects the mixed learning strategy to train the detection network.

After train the YOLOv2 based on the mixed learning strategy, the next is to train the YOLOv3 with more complex network.

Training of the YOLOv3

Since the YOLOv3 has the more complex network, we use the mixed learning strategy to train the YOLOv3.

(1) Brief introduction of YOLOv3

Compared with the YOLOv2, the YOLOv3 is improved mainly by composing backbone neural network with residual structure and feature pyramid network with multi-scale mapping[23]. The gray part represents the backbone neural network with residual structure, and the three output parts of the network represent the characteristic pyramid network. Feature maps of each feature pyramid network are equivalent to one output of the YOLOv2. The difference is that the output feature maps have different scales, specifically; objects of different sizes are outputted by feature maps of different scales.

The main idea of residual structure is to use residual connection to solve the problem of gradient disappearance and gradient explosion in deep neural network training. Its structure module is given in the reference[24]. Because of this “residual” connection, when the gradient of the intermediate layer is close to 0, the backpropagation of the gradient with a partial derivative of 1 can also be guaranteed, remaining the concept of the gradient.

The feature pyramid network takes advantage of the topological characteristics of the feature map. And the feature maps of the loop from bottom to top become smaller and the encoded information is more and more advanced, while the feature maps of the loop from top to bottom become larger and the resolution is higher and higher. In this way, the feature maps in smaller size at the top of the pyramid are suitable for processing complex features, while the feature maps in higher

Table 1 — Results of different learning strategies

	Original network	Sample A[a]	Sample B[b]	Sample C[c]	Sample D[d]
mAP	28.61%	40.03%	46.36%	42.73%	76.26%
Recall	42.13%	52.91%	60.10%	59.33%	81.45%

a: only samples in 200 m visibility; b: only samples in 300 m visibility; c: Gradient learning strategy; and d: Mixed learning strategy resolution at the bottom of the pyramid are suitable for processing objects with smaller and simpler actual size. By arranging several feature maps in different sizes to form pyramids with different size gradients, the detection effect of the network on multi-scale objects can be greatly improved [25].

(2) Training effects of the YOLOv3

Compared with the YOLOv2, YOLOv3 has stronger feature expression ability. The specific training results are shown in Table 2.

Table 2 — Training effects of YOLOv2 and YOLOv3 by using the mixed learning strategy

Target detection network	Indicators	Pre-training	After training
YOLOv2	mAP	28.6%	76.2%
	Recall	42.1%	81.4%
YOLOv3	mAP	31.7%	78.8%
	Recall	52.2%	89.9%

It can be seen from Table 2 that the YOLOv3 after trained by the mixed learning strategy have strong anti-fog ability.

Experimental verification

In order to verify the effectiveness of the proposed target detection algorithm, experimental verifications are presented in this section. In this paper, two of the most representative dehaze methods are selected for comparative experiments: the dehaze method based on dark channel prior[5] and the dehaze method based on AOD convolutional neural network[8]. The test images for experiments are from the Internet.

Performance of the dark channel prior method

The experimental results of the dehaze enhancement method based on the dark channel prior are shown in Figure 4. The left, middle and right groups of Figure 4 show the detection effects of samples with different visibilities, respectively. From top to bottom in Figure 4, the pictures are,

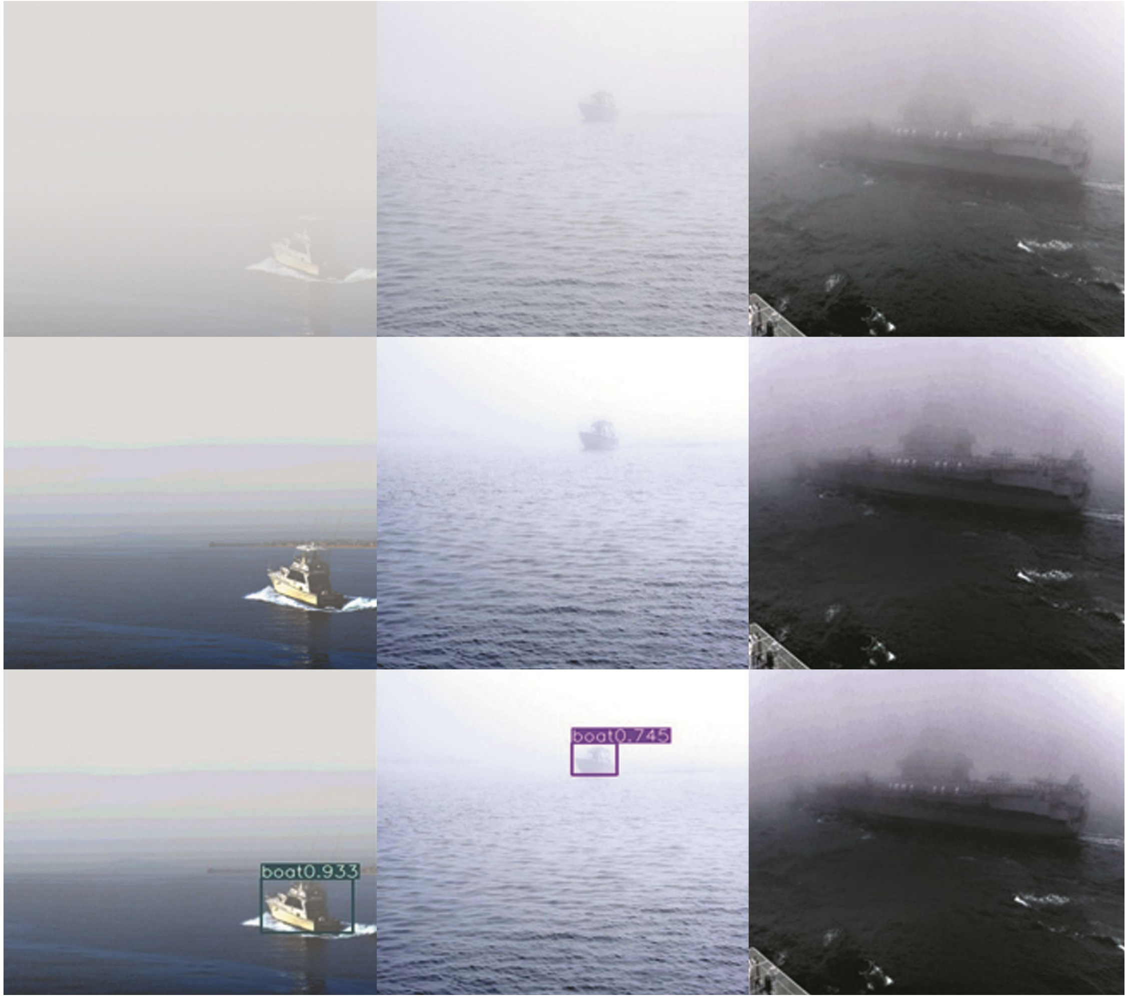

Fig. 4 — Detection results of dark channel prior based dehaze enhancement method

respectively, original images, images after dehaze, and detection results according to dehaze images.

It can be seen from Figure 4 that the accuracy rate of target detection network adopting dehaze enhancement method based on dark channel prior is improved to 29.2 % after dehaze processing. However, this method requires relatively more processing time, up to 1.524 s, *i.e.*, its real-time performance is poor.

Performance of AOD network method

The experimental results of the detection method based on AOD network are shown in Figure 5. The left, middle and right groups of Figure 5 show the processing results on samples with different visibilities, respectively. From top to bottom in Figure 5, the pictures are, respectively, original images, images after dehaze, and detection results according to dehaze images.

As can be seen from Figure 5 that the improvement in the accuracy of target detection network is poor, only 15.4 %, however, the additional time based on AOD network is less, only 0.037 s.

Detection performance of the developed method

The developed method based on the improved YOLOV3 network is tested using the same fog-scene test dataset of YOLOv2. Part of the image results are shown in Figure 6.

The top, middle and bottom groups of Figure 6 are the results for sea-surface images with different visibilities. And the left, middle and right groups of Figure 6 are, respectively, original images, images of YOLOv2 without anti-fog enhancement, and

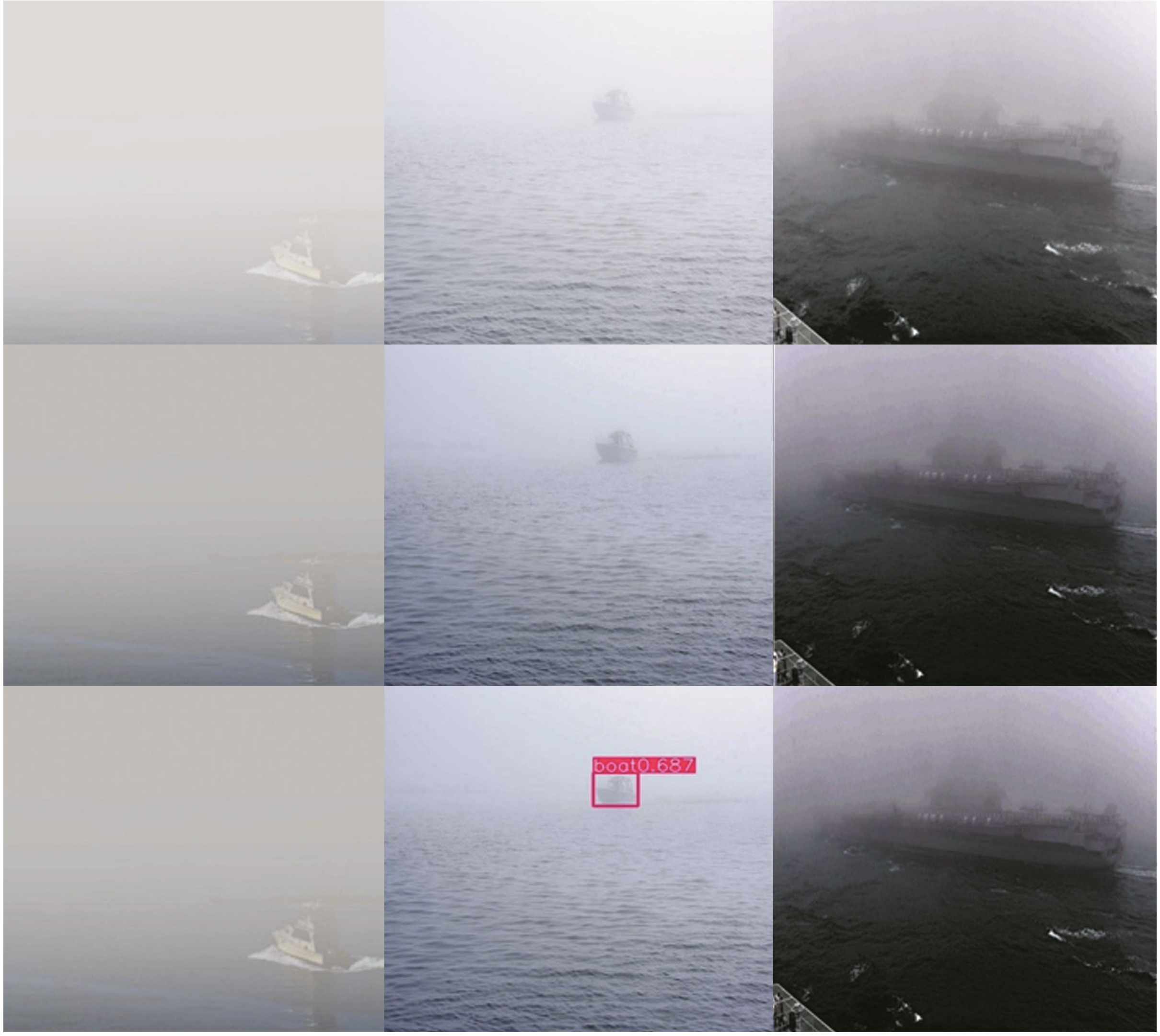

Fig. 5 — Detection effect of dehaze enhancement method based on AOD network

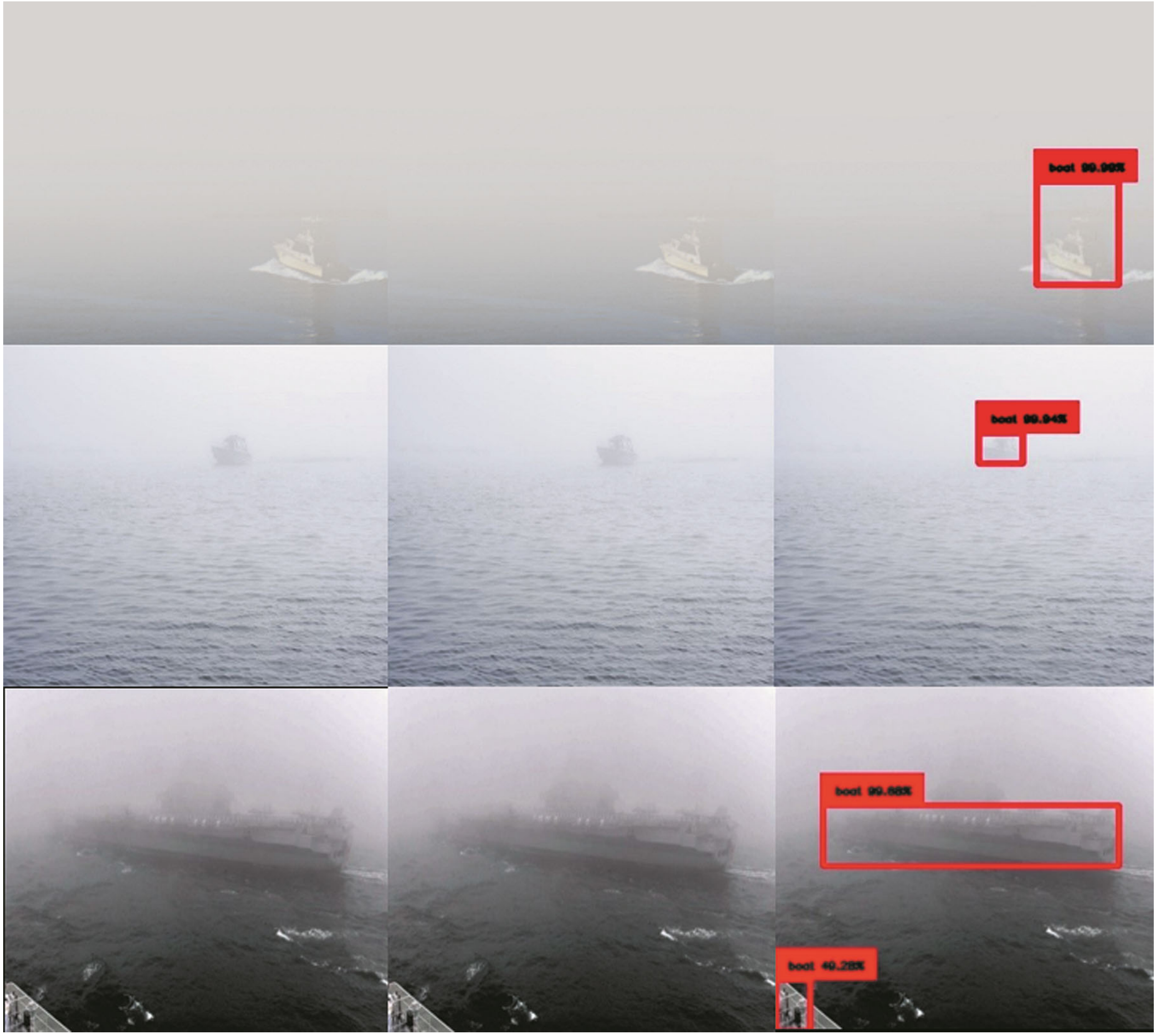

Fig. 6 — Detection effect of YOLOv3 network

YOLOv2 with anti-fog enhancement. The detection accuracy rates are 99.9, 99.4, and 99.7 %, respectively. Compared with network YOLOv2, the detection accuracy is improved. In addition, this method does not require dehazing at first, so there is no additional time.

Conclusions

Under the background of autonomous control of USVs based on visual images, this paper investigates visual detection method for sea-surface targets. Sea-surface visual images are mostly hazed. The existing methods are to using the dehaze processing at first and then detect the targets, which has a problem that the dehaze performance and processing time are difficult to be balanced. Hence, a YOLO based target detection network with good anti-fog ability is proposed for autonomous control of USVs under foggy environment. Specifically, the target detection network is offline trained to have satisfactory anti-fog ability and then is used to detect targets online. The comparative experimental results show that the proposed method has higher detection accuracy, and no additional time is required. The results satisfy the real-time requirements of autonomous control of USVs.

In this paper, the training samples are taken from MSCOCO database, and the test samples are from Internet. In future, we need to use the surface-sea images captured from USVs to perform target detection. In addition, it investigates how to differentiate between the training samples of fog-scene images with the blurry images.

Acknowledgments

This work is supported by the National Natural Science Foundation of China under Grant 52001039.

Conflict of Interest

Authors declare that there is no conflict of interest in this study.

Author Contributions

First author contributed for data collection, review of literature, data processing and drafting of the manuscript; other authors contributed for review of research methodology, data analysis, manuscript review and editing. All authors have read and approved the final version of the manuscript.

References

1 Wang Z, Yang S, Xiang X, Vasilijevic´ A, Miškovic´ N, *et al*., Cloud-based mission control of USV fleet: architecture, implementation and experiments, *Control Eng Pract*, 106 (2021) p. 104657.

2 Yu C, Liu C, Lian L, Xiang X & Zeng Z, ELOS-based path following control for underactuated surface vehicles with actuator dynamics, *Ocean Eng*, 187 (2019) p. 106139.

3 Xiang G & Xiang X, 3D trajectory optimization of the slender body freely falling through water using Cuckoo Search algorithm, *Ocean Eng*, 235 (2021) p. 109354.

4 Wang S, Zhang Y & Zhu F, Monocular visual slam algorithm for autonomous vessel sailing in harbor area, *25th Saint Petersburg International Conference on Integrated Navigation Systems (ICINS)*, (2018) pp. 1-7.

5 Han M, Lyu Z, Qiu T & Xu M, A review on intelligence dehazing and color restoration for underwater images, *IEEE Trans Syst Man Cybern: Syst*, 50 (5) (2020) 1820–1832.

6 Xuan L & Mingjun Z, Underwater color image segmentation method via rgb channel fusion, *Opt Eng*, 56 (2) (2017) p. 023101

7 Kaiming H, Jian S & Xiaoou T, Single image haze removal using dark channel prior, *IEEE Trans Pattern Anal Mach Intell*, 33 (12) (2011) 2341-2353.

8 He K, Sun J & Tang X, Guided image filtering, *IEEE Trans Pattern Anal Mach Intell*, 35 (6) (2013) 1397–1409.

9 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) (2016) 5187–5198.

10 Li B, Peng X, Wang Z, Xu J & Feng D, AOD-Net: All-in-one dehazing network, *IEEE Int Conf Comput Vis (ICCV)*, 2017, pp. 4080-4088.

11 Wang R, You Y, Zhang Y, Zhou W & Liu J, Ship detection in foggy remote sensing image via scene classification r-cnn, *IEEE Int Conf Netw Infrastruct Digit Content (IC-NIDC)*, (2018) 81-85.

12 Huang H, Sun D, Wang R, Zhu C & Liu B, Ship target detection based on improved yolo network, *Math Probl Eng*, (2020) 1–10.

13 Liu T, Pang B, Zhang L, Yang W & Sun X, Sea surface object detection algorithm based on yolov4 fused with reverse depthwise separable convolution (rdsc) for usv, *J Mar Sci Eng*, 9 (7) (2021) p. 753.

14 Li Y, Guo J, Guo X, Liu K, Zhao W, *et al.*, A novel target detection method of the unmanned surface vehicle under all-weather conditions with an improved yolov3, *Sensors*, 20 (17) (2020) p. 4885.

15 Liu T, Pang B, Ai S & Sun X, Study on visual detection algorithm of sea surface targets based on improved yolov3, *Sensors*, 20 (24) (2020) p. 7263.

16 Nayar S K & Narasimhan S G, Vision in bad weather, *7th IEEE Int Conf Comput Vis*, (1999) 820-827.

17 Chen L, Papandreou G, Kokkinos I, Murphy K & Yuille A L, DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs, *Trans Pattern Anal Mach Intell*, 40 (4) (2018) 834–848.

18 Kumar R, Visual linguistic model and its applications in image captioning, *SN Comput Sci*, 1 (3) (2020) p. 124.

19 Wang N, Wang Y & Er M J, Review on deep learning techniques for marine object recognition: Architectures and algorithms, *Control Eng Pract*, (2020) p. 104458.

20 Wang S-H, Lv Y-D, Sui Y, Liu S, Wang S-J, *et al.*, Alcoholism detection by data augmentation and convolutional neural network with stochastic pooling, *J Med Syst*, 42 (1) (2017) 1-11.

21 Wang Z & Feng Y, Fast single haze image enhancement, *Comput Electr Eng*, 40 (3) (2014) 785–795.

22 Andreae M O & Crutzen P J, Atmospheric aerosols: Biogeochemical sources and role in atmospheric chemistry, *Science*, 276 (5315) (1997) 1052–1058.

23 Suarez P L, Sappa A D & Vintimilla B X, Cross-spectral image dehaze through a dense stacked conditional gan based approach, *14th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*, (2018) 358-364.

24 He K, Zhang X, Ren S & Sun J, Deep residual learning for image recognition, *IEEE Conf Comput Vis Pattern Recognit (CVPR)*, (2016) 770-778.

25 Xiang J & Zhu G, Joint face detection and facial expression recognition with mtcnn, *4th Int Conf Inf Sci Control Eng (ICISCE)*, (2017) 424-427.