SOMC : A Object-Level Data Augmentation for Sea Surface Object Detection

Jianqiu Hu1,*, Jiazhou He1, Pan Jiang1, Yuwei Yin1

1Jiangsu Automation Research Institute, Lianyungang 222061, China

*email: jianqiu@mail.ustc.edu.cn

Abstract. The deep learning model is a data-driven model and more high-quality data will bring it better results. In the task of Unmanned Surface Vessel’s object detection based on optical images or videos, the object is sparser than the target in the natural scene. The current datasets of sea scenes often have some disadvantages such as high image acquisition costs, wide range of changes in object size, imbalance in the number of different objects and so on, which limit the generalization of the model for the detection of sea surface objects. In order to solve problems of insufficient scene and poor effect in current sea surface object detection, an object-level data augmentation for sea surface objects called SOMC is proposed. According to different scenarios faced by the USV when performing autonomous obstacle avoidance, patrol and other tasks, SOMC generates suitable scenarios by mixing and copying targets conveniently, providing the possibility of unlimited expansion of the sea surface object. The experiment selected images in the video taken by the camera on top of the USV. A sufficient amount of comparative experiment prove that the SOMC integrates with existing excellent data augmentations and achieved an improvement in the detection effect, which proves the effectiveness and practicability of the SOMC in the perception task of the USV.

1. Introduction

The ocean area accounts for about 71% of the total surface area and the vast ocean needs to be developed and utilized by humans. The efficient perception of sea surface objects is the first step for us to understand the ocean. At this stage, with the widespread application of high-resolution optical sensors, such as cameras, it is necessary to automatically interpret huge quantities of sea surface images. Among them, the amount of video captured by the USV is large and the processing is difficult, however, in the harsh environment, the picture of the sea surface is small and costly. How to deal with the sea-surface image efficiently is a common concern.

Since 2014, deep learning represented by convolutional neural networks has been widely used in large-scale common life scene image detection, which brings opportunities for automated processing of large quantities of sea surface optical data. At present, in the field of sea-surface object detection, methods are mainly divided into two categories. The first category is traditional image processing technology based on artificial design features. For example, Bayes' method based on gray distribution characteristics of the sea surface[1], template matching method based on ship target[2], machine learning based on object-based combination of features[3]. The second category is deep learning method based on automatic feature extraction of convolutional neural networks, such as Yolo[4], Faster R-CNN[5], CenterNet[6], etc. These detection algorithms based on deep learning are used in pattern recognition tasks widely and the detection accuracy is higher than that of the first type of algorithm. However, the detection algorithm based on deep learning has the problem of weak scene
generalization ability and insufficient detection speed. In the actual deployment of the algorithm, the use of appropriate data augmentation methods[7][8] is an efficient and convenient way to improve model detection performance and prevent overfitting. This is especially obvious for some small datasets.  

In the real engineering environment, it is found that if image enhancement methods of ordinary scenes, such as flip and rotation, are simply applied[9] and the complex environment of the sea is ignored, the detection model training is inefficient and the accuracy improvement is limited, because objects with different scales and different perspectives do not get enough attention, then the model tend to fall into over-fitting. Besides, the previous data augmentation[10] of sea surface objects was too complicated and inconvenient in engineering applications. For example, the segmentation algorithm is used to segment the ship and then copy it to the image to improve the effect of the detection model[10][11]. The image data generated by this approach has a good visualization effect. At the same time, when the generated ship data is sufficient, the detection result is also improved. However, this augmentation required segmentation model and segmentation data which require more workload. Therefore, in order to solve these problems above, we propose a data augment called SOMC for a variety of sea scenes, which simply achieve a mix of multiple sea surface objects and backgrounds. Contributions of this paper are as follows:

1. We propose an object-level data augmentation method called SOMC for sea surface object detection to alleviate the problems of model overfitting, fewer complex scenes and insufficient labeled samples.

2. We create the Yellow Sea civil maritime dataset through the USV, including the port and open sea channel scenes under various conditions. On this dataset, SOMC improves the generalization ability of the model in different scenarios.

3. SOMC does not require additional complicated network structure for feature learning and it combines with other excellent classic data augmentation to speed up model training, make the model better converge, and achieve better results at a small cost.

2. Methodology

Aiming at the situation encountered by the USV during the autonomous driving and the defects of the existing dataset, this paper designs a simple data augmentation based on the specific scenario to achieve data expansion. The overall strategy flowchart is shown in figure 1.

![Diagram of SOMC application](image_url)

**Fig. 1** Diagram of SOMC application

2.1. Sea area selection

In a real marine environment, complex weather conditions will cause serious interference to image processing and affect the object detection performance of the USV. In this paper, the sea-sky-line detection is used to select the sea area roughly, and the strategy used for sea-sky-line detection is domain gray feature estimation[12]. The detection results of the sea area under various weather conditions are shown in figure 2.
In the subsequent data augmentation of different scenes, for some pictures in the training set, we copy and paste object frames and it restricts the midpoint of the bottom edge of the frame to the sea surface area, achieving the enhancement of the instance under any sea conditions.

2.2. Tracking and encountering scenes
In the process of the USV entering and leaving the port or performing tracking tasks, there are many encounters and tracking scenes. In these scenes, there are problems such as dramatic changes in the size of objects. When the distance between the USV and the object to be detected is far, the size of the object is small. They are easy to be mistakenly detected as other categories. When the distance is close, the object size is large and the observation angle of the USV changes drastically.

In response to these situations, SOMC uses the scale change strategy to make up for the shortcomings of few object samples of extreme sizes and uses horizontal flip or truncation strategy to make up for the shortcomings of fewer samples from any perspective. In this process, the upside-down operation is not used and the horizontal truncation is reflected in the subsequent mixing strategy in section 2.4, which could simulate the scene where the target is incomplete at the edge of the picture. At this time, the SOMC strategy is determined by three parameters: the degree of maximum scale transformation, scale enhancement probability, and horizontal flip probability. The SOMC effect is shown in figure 3.

2.3. Own ship shaking scenes
The USV will encounter different sea conditions on the sea and sometimes the hull shakes violently. Correspondingly, the observed sea horizon also changes drastically. The common method is to rotate the background by a certain angle, while SOMC uses the rotation of the object to simulate this situation, which can simulate the target's rotation caused by the USV horizontal distance or bad sea conditions to a greater extent. At this time, the SOMC strategy has newly added the parameter of the maximum rotation angle of the target. The effect of SOMC is shown in figure 4.

In this case, the rotation causes the object frame to change. The input is the coordinates of the maximum point and the minimum point of the original object frame. At the same time, the output is the minimum point coordinates of the maximum point of the rectangular object frame after the rotation
transformation. In the rotation transformation, the transformation formulas for the four vertices of the original rectangular target frame are shown in equations (1) and (2) and the calculation for selecting the coordinates of the maximum point of the newly generated target frame is shown in equation (3).

$$x_{\text{rot}} = (x_i^\text{ori} - (x_{\text{min}}^\text{ori} + x_{\text{max}}^\text{ori}) / 2) \times \cos(\theta) - (y_i^\text{ori} - (y_{\text{min}}^\text{ori} + y_{\text{max}}^\text{ori}) / 2) \times \sin(\theta) + (x_{\text{min}}^\text{ori} + x_{\text{max}}^\text{ori}) / 2$$

$$y_{\text{rot}} = (x_i^\text{ori} - (x_{\text{min}}^\text{ori} + x_{\text{max}}^\text{ori}) / 2) \times \sin(\theta) + (y_i^\text{ori} - (y_{\text{min}}^\text{ori} + y_{\text{max}}^\text{ori}) / 2) \times \cos(\theta) + (y_{\text{min}}^\text{ori} + y_{\text{max}}^\text{ori}) / 2$$

$$x^\text{rot}_{\text{min}} = \min \{x_i^\text{rot}\}; y^\text{rot}_{\text{min}} = \min \{y_i^\text{rot}\}; x^\text{rot}_{\text{max}} = \max \{x_i^\text{rot}\}; y^\text{rot}_{\text{max}} = \max \{y_i^\text{rot}\}$$

2.4. Mixed strategy
We need to mix different scenarios and the specific mixing strategy is as follows: first, the scene generated pictures are processed together with the original pictures, checking if the object and label in the image match correctly. When generating a new image, we add a variety of noise, blur and other operations to the object frame randomly. Next, we divide each image into four small slices according to the X-axis direction and select the picture slice with the largest number of objects in slices. After that, the incomplete object label in the slice is processed reasonably. Finally, the four picture slices are combined into a mixed picture as a new image to train. The advantage of mixed strategy is to increase the diversity of image data and improve the detection efficiency of sparse targets. When the BN normalization layer[13] is used for calculation, the statistical mean and variance of the features in each feature layer are closer to the original entire dataset. The result of mixed strategy data enhancement is shown in figure 5.

![Fig. 5 Diagram of SOMC under mixed strategy](image)

3. Experiments

3.1. Dataset
We created the Yellow Sea civilian maritime dataset (YSCMD). Most of the sources of the ship data set are the real sea conditions collected by the USV’s optical sensors in the Yellow Sea, and the rest are a small number of similar scene pictures open sourced on the Internet. We obtain images and videos from different sea locations and channels, at the same time we collect different scenes, including mist, strong light, low light, complex near-shore background, incomplete near-distance targets, and blurred distant targets, etc.
The dataset is divided into 5 categories of civilian sea surface objects. The first type is production and operation ships, such as fishing boats, etc; the second type is commercial transportation ships, such as cargo, etc; the third type is competitive viewing types, such as sailing ships, etc; the fourth type is ships that carry tourists, such as passenger ships, speedboats, etc; the fifth category is floating objects on the water. Each type of object is named after the object that accounts for the most. The five types of objects in the dataset as shown in figure 6. The image size in the dataset is 1920×1080. The details of the dataset are shown in table 1. The unit of average width and height in table 1 is pixel.

![Fig. 6 Diagram of different categories of ship in YSCMD](image)

### Tab. 1 Statistics table of the YSCSD

| Classes       | Trainset | Testset | Average height | Average weight |
|---------------|----------|---------|----------------|---------------|
| Fishing_ship  | 2134     | 512     | 48.3           | 90.9          |
| Cargo         | 1043     | 314     | 97.4           | 263.7         |
| Sailing Boat  | 4155     | 1217    | 87.7           | 55.2          |
| Passenger Ship| 2819     | 664     | 131.0          | 180.0         |
| Floating      | 1944     | 512     | 52.0           | 75.4          |

#### 3.2. Experimental details

The hardware information of the experimental platform is as follows: CPU is Intel Xeon Silver 4110, GPU is Nvidia GeForce RTX 2080Ti. The platform software information is as follows: the system is Ubuntu and the software framework is based on Pytorch1.7.0. The basic experimental model used is the yolov5-m model released by the ultralytics team.

The comparison experiment uses the same test conditions in the process of training and testing the data set, mainly including the same training set and test set, the same network model structure and hyperparameters, etc. By choosing different data enhancement methods or use different degrees of enhancement methods to conduct comparative experiments. The specific values of the relevant parameters in the experiment are as follows: the input size is 640×640, the learning rate is 0.01, the optimizer uses SGD, the epoch is 200, use the same warmup operation, the hyperparameters of the loss function adopt the initial value without adjustment, etc. The SOMC used in the experiment uses 1,000 mixed strategy images to be added to the training. The number of images generated by SOMC is 1000, and the training set images are equivalent to self-expanding by about 40%. According to the requirements of subsequent engineering tasks, the evaluation index we chose in the experiment is mAP@.5.

#### 3.3. Result analysis

The function of data enhancement is to expand the limited and expensive data set so that the model can achieve better detection results in different scenarios. Among them, the image augmentation method for hue and saturation belongs to pixel-level augmentation. For the above two types of methods, the parameters initially set in the model are used in the experiments in this article and no comparison test is performed.
We conduct comparative experiments on the enhancement methods of image blocks, mainly including Mixup[14] and Mosaic[9], which are currently effective methods in the basic algorithm. These methods belong to multiple image composite enhancement methods for object detection. Mixup selects 2 images at random for weighted summation and the label of the sample corresponds to the weighted summation; Mosaic combines 4 images to synthesize a new picture, which is rich in backgrounds and objects. During the experiment, SOMR explores the combination of Mixup or Mosaic to achieve better and more efficient data augmentation. The training situation of different groups of experiments is shown in figure 7, and the data of the comparative experiment table is shown in table 2 below.

![Fig. 7 Diagrams of training conditions for different parameters](image)

**Tab. 2** Experimental results of multiple augmentations combined with SOMC

| Method | 1  | 2  | 3  | 4  | 5  | 6  |
|--------|----|----|----|----|----|----|
| Mosaic | 0.5| 0.5| 0.5| 0.5| 0.5| 0.5|
| Mixup  | 0  | 0  | 0.5| 0.75| 0.75| 0.75|
| SOMC   | 0  | 1  | 0  | 1  | 0  | 1  |
| Trains | 0.60| 0.76| 0.75| 0.87| 0.87| 0.80|
| Testset| 5  | 1  | 3  | 7  | 0  | 6  |

| Method | 7  | 8  | 9  | 10 | 11 | 12 |
|--------|----|----|----|----|----|----|
| Mosaic | 0.75| 0.75| 0.75| 0.75| 1  | 1  |
| Mixup  | 0.75| 0.75| 1  | 0  | 0  | 0  |
| SOMC   | 0  | 1  | 0  | 1  | 0  | 1  |
| Trains | 0.84| 0.91| 0.86| 0.91| 0.82| 0.88|
| Testset| 2  | 0  | 3  | 5  | 6  | 1  |

In table 2, the first row of values indicates the serial number of the experiment; the second row to the fourth row indicate the probability of using this augmentation; the fifth row indicates the evaluation result on the training set when the model is trained to 200 epochs, corresponding to figure 7; the sixth row indicates the test result on the test set after the model is trained.
According to the experimental results on the fifth row of table 2 on the test set, we can see that the method combining SOMC with different degrees of mosaic and mixup methods has achieved better detection results than the previous methods, reflecting the effectiveness of SOMC in detecting sparse targets on the sea. It can be seen from figure 7 and the fourth row of table 2 that the model combined with SOMC can spend less epochs to achieve better detection results. The model converges faster, and at the same time it is easier to converge to a better result, which enhances the generalization of the model.

Related work should focus on the data quality itself. To achieve better engineering practice results, more attention should be paid to the wrongly divided samples and difficult-to-detect targets. For this, more attention should be paid to the construction of the data set. In subsequent actual projects, sometimes it is necessary to process data in a small number of extreme scenarios. The SOMC method can quickly generate a large amount of data in similar scenarios, which has certain practical value in engineering.

4. Conclusion
Aiming at different scenarios encountered by USV in actual automatic driving, this paper proposes a sea surface object enhancement method called SOMC, which is effective and does not involve complex hyperparameter adjustments. In the real sea surface target detection experiment, combined with the excellent enhancement method at this stage, better detection results can be obtained faster and the generalization ability of algorithm detection can be improved. In the future, SOMC is suitable for sea-surface target detection scenarios, combined with traditional data enhancement strategies or image generation methods using generative adversarial networks. In engineering practice, it will have a certain value in improving the detection effect of sparse targets and the construction of auxiliary datasets.

References
[1] Proia N and Pagè V 2009 Characterization of a Bayesian ship detection method in optical satellite images IEEE Geoscience and Remote Sensing Letters 7(2) pp 226-230
[2] Chao X, Dongping Z, Zhengning Z and Zhiyong F 2014 BgCut: automatic ship detection from UAV images The Scientific World Journal
[3] Selvi M and Kumar S 2011 A novel approach for ship recognition using shape and texture International Journal of Advanced Information Technology.
[4] Tao L, Bo P, Lei Z, Wei Y and Xiaoqiang S 2021 Sea Surface Object Detection Algorithm Based on YOLO v4 Fused with Reverse Depthwise Separable Convolution (RDSC) for USV Journal of Marine Science and Engineering 9(7) p 753
[5] Shaoqing R, Kaiming H, Girshick R and Jian S 2015 Faster r-cnn: Towards real-time object detection with region proposal networks Advances in neural information processing systems 28 pp 91-99
[6] Kaiwen D, Song B, Lingxi X, Honggang Q, Qingming H and Qi T 2019 Centernet: Keypoint triplets for object detection In Proceedings of the IEEE/CVF International Conference on Computer Vision pp 6569-6578
[7] Shorten C and Khoshgoftaar T 2019 A survey on image data augmentation for deep learning Journal of Big Data 6(1) pp1-48
[8] Ghiasi G, Yin C, Srinivas A, Rui Q, Tsung-Yi L, Cubuk, E, Le Q and Zoph B 2021 Simple copy-paste is a strong data augmentation method for instance segmentation In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition pp 2918-2928
[9] Bochkovskiy A, Chien-Yao W and Hong-Yuan Mark L 2020 Yolov4: Optimal speed and accuracy of object detection arXiv preprint arXiv:2004.10934
[10] Kaiming H, Gkioxari G, Dollar P and Girshick R 2017 Mask r-cnn In Proceedings of the IEEE international conference on computer vision pp 2961-2969
[11] Shin H, Lee K and Lee C 2020 Data augmentation method of object detection for deep learning
in maritime image In 2020 IEEE International Conference on Big Data and Smart Computing pp 463-466

[12] Pu H, Bo L, Tingting R and Yufei C 2014 A fast sea-level line extraction and object detection method for infrared sea image In International Symposium on Optoelectronic Technology and Application 2014: Infrared Technology and Applications p 930007

[13] Ioffe S and Szegedy C 2015 Batch normalization: Accelerating deep network training by reducing internal covariate shift In International conference on machine learning pp 448-456

[14] Hongyi Z, Cisse M, Dauphin Y, Lopez-Paz D 2017 mixup: Beyond empirical risk minimization arXiv preprint arXiv:1710.09412.