Tech Science Press

MCMOD: The Multi-Category Large-Scale Dataset for Maritime Object Detection

Zihao Sun[1,*], Xiao Hu[2], Yining Qi[2], Yongfeng Huang[2] and Songbin Li[3]

[1]Institute of Remote Sensing and Geographic Information System, Peking University, Beijing, 100871, China
[2]Department of Electronic Engineering, Tsinghua University, Beijing, 100084, China
[3]Institute of Acoustics, Chinese Academy of Sciences, Beijing, 100190, China
*Corresponding Author: Zihao Sun. Email: sunzihao211@gmail.com

Abstract: The marine environment is becoming increasingly complex due to the various marine vehicles, and the diversity of maritime objects poses a challenge to marine environmental governance. Maritime object detection technology plays an important role in this segment. In the field of computer vision, there is no sufficiently comprehensive public dataset for maritime objects in the contrast to the automotive application domain. The existing maritime datasets either have no bounding boxes (which are made for object classification) or cover limited varieties of maritime objects. To fulfil the vacancy, this paper proposed the Multi-Category Large-Scale Dataset for Maritime Object Detection (MCMOD) which is collected by 3 onshore video cameras that capture data under various environmental conditions such as fog, rain, evening, etc. The whole dataset consists of 16,166 labelled images along with 98,590 maritime objects which are classified into 10 classes. Compared with the existing maritime datasets, MCMOD contains a relatively balanced quantity of objects of different sizes (in the view). To evaluate MCMOD, this paper applied several state-of-the-art object detection approaches from computer vision research on it and compared their performances. Moreover, a comparison between MCMOD and an existing maritime dataset was conducted. Experimental results indicate that the proposed dataset classifies more types of maritime objects and covers more small-scale objects, which can facilitate the trained detectors to recognize more types of maritime objects and detect maritime objects over a relatively long distance. The obtained results also show that the adopted approaches need to be further improved to enhance their capabilities in the maritime domain.

Keywords: Object detection; marine vehicles; deep learning; performance evaluation

1 Introduction

With the increasing variety of ships, marine environmental management is now facing an urgent challenge which is to recognise different types of maritime vessels (MVs) rapidly and precisely. Maritime object detection is of great importance in numerous fields including military, smart cities, public security, and transportation. Traditionally, the detection of MVs is accomplished manually by maritime staff monitoring the marine environment through videos or monitors, and this approach is time-consuming and inefficient. Besides, manual detection is heavily affected by adverse weather conditions such as fog, rain, and insufficient light. With the explosive developments of deep learning and computer vision, object detection can now be accomplished by artificial neural networks. Nowadays, this technology has been widely used in various domains such as generic object detection, pedestrian detection, fruit ripeness prediction [1], and face detection [2]. Given the satisfactory performance of computer vision research in these applications, maritime object detection can be achieved efficiently and accurately by applying and improving state-of-the-art methods. Maritime object detection includes two tasks: localize the MVs in the view and classify the MVs.

The performances of deep learning algorithms highly rely on a large amount of training data, as sufficient training samples ensure their robustness and accuracy. The emergence of large datasets has contributed to the vigorous development of deep learning [3]. Numerous large-scale datasets have been published to support training object detection algorithms, such as VOC2007 [4], CIFAR-10 [5], and COCO [6]. However, these datasets focus on generic object detection, and they deem all kinds of vessels as 'ship' or 'boat' without breaking down the vessel types. In recent years, with the arousing attention to maritime object detection, MARVEL [7], VAIS [8], Seaships [9], and Mcships [10], which are datasets of vessels, have been published. These maritime object datasets either contain few MV types or do not provide bounding boxes or have not been publicly available.

This paper presents MCMOD, a new multi-category large-scale dataset of MVs, which is developed for training maritime object detection algorithms. Three onshore cameras were set up on fixed platforms in Hainan, China, recording high-resolution (1080 × 1920 pixels) video data. The 3 cameras operate 24 h a day and keep rotating and zooming in or out of view, thus the dataset is of high diversity of scenes. MCMOD currently consists of 16166 labelled images captured from raw video data, and the images are under different varieties of weather conditions including foggy, rainy, and evening. 98590 maritime objects are classified into 10 varieties, namely fishing vessels, speedboats, engineering ships, cargo ships, yachts, sailboats, buoys, rafts, cruise ships, and others.

To evaluate MCMOD, 4 state-of-art detectors from computer vision research get utilized on it. This paper compared the performances of these detectors and discussed their advantages as well as weakness and identified some possible directions for further studies.

The remainder of this paper is organized as follows: related works are presented in Section 2. Section 3 describes the dataset design and analyse the dataset. Section 4 shows the experiments and results along with the discussion among them. And an inductive conclusion is drawn in Section 5.

2 Related Works

Maritime Object Detection: Detection and identification of MVs is a time-consuming and tedious task, and numerous attempts have been made to simplify and free up human resources in the past few years. Traditionally, the detection is accomplished by naked-eye observation through monitors or videos until computer-aided methods came up. The idea of computer-aided approaches is to utilize computers to extract features from images and then use classifiers to accomplish the classification and

localization of MVs. Tello et al. [11] proposed a ship detection algorithm based on wavelet theory, using the difference in statistical behaviour between vessels and sea to interpret the information through the wavelet coefficients from synthetic-aperture radar (SAR) images in 2005. And 5 years later, Zhu et al. [12] extracts vessels from spaceborne optical images by using a semisupervised hierarchical classification approach based on shape and grey intensity features. Hwang et al. [13] integrated a visual camera with radar to detect MVs based on foreground and background segmentation using an average filter and Kalman filter. In 2013, Yang et al. [14] used background sea surface features and a linear classifier to combine pixel and region characteristics to extract vessels. These methods required manual feature selection and are not robust when disturbances happen.

Recently, researchers used deep convolutional neural networks (DCNNs) and deep learning techniques to detect MVs. Kim et al. [15] utilizes Faster R-CNN [16] to detect ships, and then recover the missing ships by using Intersection over Union (IOU) of the bounding boxes from a sequence of images. Marie et al. [17] uses Fast R-CNN [18] to find ships from Regions of Interest (ROIs) which are extracted earlier by statistical machine learning methods. As for the assessment criteria, Prasad et al. [19] points out that the conventional assessment metrics for generic object detection are not suitable in the maritime setting. He proposes new bottom edge proximity metrics to assess maritime object detection.

Object Detection Datasets: In the field of computer vision, many datasets for generic object detection and specific object detection already exist. Imagenet [20] offers 14 million well-labelled images covering 22 object categories. It sorts the objects in a tree structure innovatively. CIFAR-10 [5] and PASCAL [21] are also famous datasets in computer vision tasks, however, they contain a small number of categories and do not sub-classify 'ships' or 'boats'. To facilitate the research of face detection, Yang et al. [22] presents WIDER FACE which contains 32,203 images along with 393,793 labelled faces. Besides, CityPersons [23] (pedestrian detection), CUB-200-2011 [24] (bird detection), and FSNS [25] (street sign detection) are designed for their corresponding object detection domain. These large-scale datasets can support and facilitate object detection research.

In the past years, several maritime datasets have been developed, but few of them are suitable for academic and research purposes in maritime object detection scenarios. Gundogdu et al. [7] proposed MARVEL which contains 2 million images of MVs along with their attributes. And VAIS, presented by Zhang et al. [8], consists of more than 1,000 visible and infrared ship images among 6 vessel types. However, these 2 datasets do not include bounding boxes for object detection as they are created for object classification. Seaships [9] do have bounding boxes but it only covers 6 common ship categories. Zheng et al. [10] proposed another large-scale ship dataset (McShips), which contains both civilian ships and warships, but the dataset is not yet publicly available. To this end, this article creates the MCMOD dataset, which will be discussed in the following section.

3 MCMOD

In this paper, 3 cameras with changeable views are set up in the coastal area of Hainan Province, China to capture the maritime objects, and then the raw image data under different weather conditions (cloudy, rainy, foggy, evening, etc.) are intercepted to improve the robustness under all weather conditions of the subsequent model training results. Fig. 1 shows 6 examples of different weather conditions, in which (a) to (d) are annotated images of MCMOD while (e) and (f) are detection results generated by a trained detection model.
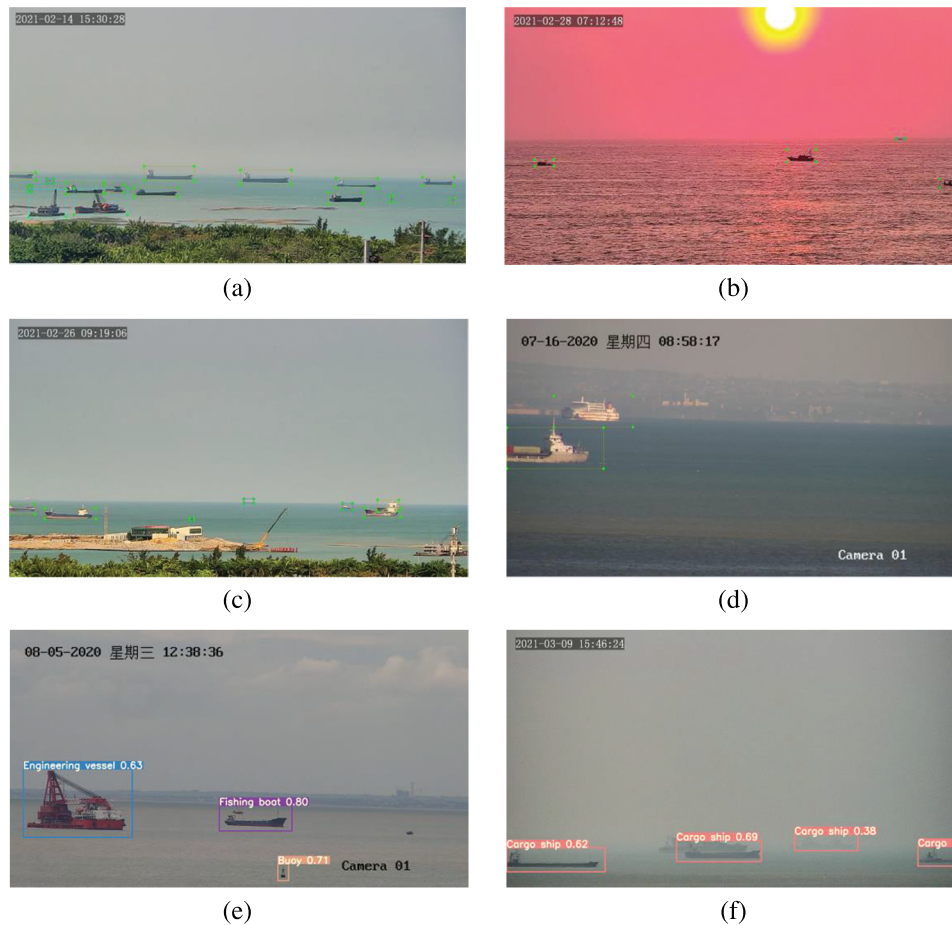
Figure 1: Example images of MCMOD

The self-built dataset is labelled with the original image data in PASCAL-VOC format using a labelling tool called LabelImg, and the 10 labelling types and the number of each type are shown in Fig. 2 below. Fig. 2 indicates that MCMOD is not balanced between the number of different kinds of vessels. This is caused by the frequency of occurrence of each type of MVs, however, it is not an adverse phenomenon as it fits the wild regularity on the contrary.

Fig. 2 also presents the number of large objects (64×64 pixels $<$ area), medium objects (32×32 pixels $\leq$ area $\leq 64 \times 64$ pixels), and small objects (area $< 32 \times 32$ pixels). MCMOD contains 54,579 large objects, 18,594 medium objects and 25,417 small objects. Compared to the SeaShips dataset (only 7,000 images are available, with only 11 small objects and 195 medium objects), MCMOD is relatively balanced in different object sizes. With a larger number of small and medium objects, the MCMOD dataset can support trained detectors to recognize MVs over longer distances.

This paper compares the MCMOD dataset against 3 other existing object detection datasets which 2 are designed for generic object detection (VOC2007 & CIFAR-10) and 1 is designed for maritime object detection (SeaShips). The differences are shown in Table 1.

According to Table 1, it can be told that MCMOD covers more MVs in one single image than the SeaShips dataset, as small and medium objects in the view are also annotated.
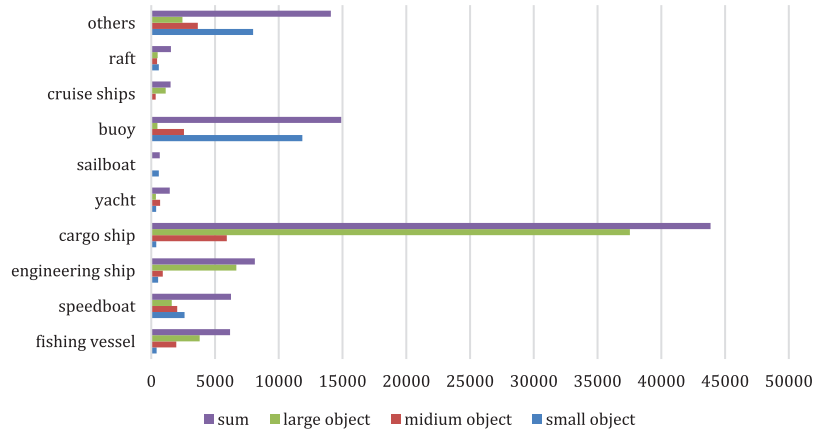
Figure 2: The number of annotated objects

Table 1: Comparison with other datasets

Dataset	Images[a]	Object[b]	Types	Image size
VOC2007	363	791	1 (boat)	Random
CIFAR-10	6,000	6,000	1 (ship)	32×32
SeaShips[c]	31,455	40,077	6[d]	1920×1080
MCMOD	16,166	98,590	10	1920×1080

Notes: [a]Images means the number of images containing ship/boat. [b]Objects means the number of total annotated objects. [c]The statistical data of seaships is from the author's paper as only 7000 images are available. [d]6 Types from seaships are ore carrier, bulk cargo carrier, general cargo ship, container ship, fishing boat and passenger ship.

4 Results and Analysis

4.1 Evaluation Metrics

There are some typical quantitative indicators for evaluating an object detection model. This paper adopted the commonly used evaluation metrics which will be briefly described below.

4.1.1 Intersection Over Union

The overlap ratio Intersection Over Union (IOU) can be computed as:

$$IOU = \frac{|b \cap b_g|}{|b \cup b_g|} \tag{1}$$

b is the predicted Bounding Box and b_g is the ground truth box. IOU is not smaller than a predefined threshold ε, or it belongs to the background. A typical value of ε is 0.5.

4.1.2 Average Precision

Average Precision (AP) is computed over different recalls achieved by varying the confidence:

$$Ap = \int_0^1 P(R)dR \tag{2}$$

AP averaged over all categories is mean Average Precision (mAP). In this paper, mAP and mean Average Recall (mAR) were averaged over multiple IOU values. Specifically, this paper used 10 IOU $\in [0.50:0.05:0.95]$, which is from COCO [6].

Moreover, AP_{50} is AP at IOU $= 0.50$.

AP_{75} is AP at IOU $= .075$.

AP_S is AP for small objects.

AP_M is AP for medium objects.

AP_L is AP for large objects.

4.1.3 Frames Per Second

Frames Per Second (FPS) is the number of frames that the detector can detect per second which is to measure speed.

4.2 Experimental Settings

This paper performed experiments based on an object detection toolbox MMDetection [26] under Ubuntu 16.04.6 LTS, CUDA V10.1, and CUDNN 7.6 systems. All the experiments run with one NVIDIA GTX 1080Ti GPU and 2.10 GHz Intel(R) Xeon(R) CPU E5-2620 v4. This paper implemented 4 methods using the PyTorch 1.5.1 library [27] with Python 3.7.11.

All models were pre-trained on ImageNet and then finetuned for maritime object detection. This paper trained a Fast-RCNN detector and a Faster-RCNN detector with a ResNet101 backbone. For SSD, the VGG16 net was used as the backbone. For YOLO v3, darknet53 was adopted. Most hyper-parameters were kept the same for a fair comparison.

4.3 Detection Methods

4.3.1 Detection Results

As MCMOD is a new dataset, this paper evaluated the state-of-the-art object detection methods on MCMOD. Four widely used detectors were carefully selected: Fast R-CNN [18], Faster R-CNN [16], SSD [28] and YOLOv3 [29].

Table 2 reports the performances of all selected models on MCMOD. As can be seen from Table 2, two-stage methods Fast R-CNN and Faster R-CNN achieve considerable effects on precisions, as all highest AP are achieved by these two methods. In contrast to 6 and 10 FPS of Fast R-CNN and Faster R-CNN, one-stage (unified) methods SSD and YOLO v3 are significantly better in detection speed as they achieve 28 and 44 FPS respectively. This difference is caused by how the approaches detect objects. Two-stage methods divide the detection procedure into two stages, in which the first step is to produce region proposals, followed by position refinement and classification of region proposals [30]. And one-stage methods generate class probabilities and position coordinates directly without constructing region proposals. Thus, one-stage methods show their advantages in detection speed while two-stage methods perform better in precision and accuracy which fits the results.

Table 2: Detection results (AP) on the MCMOD

Method	Epochs	mAP	AP50	AP75	APS	APM	APL	mAR	FPS
Fast R-CNN (ResNet-101)	24	**0.351**	0.657	**0.343**	**0.235**	**0.350**	0.430	0.434	6
Faster R-CNN (ResNet-101)	24	0.344	**0.691**	0.239	0.227	0.335	**0.495**	**0.464**	10
SSD512 (VGG16)	24	0.299	0.647	0.182	0.172	0.294	0.385	0.416	28
YOLO v3 608 (Darknet-53)	273	0.259	0.620	0.176	0.187	0.286	0.326	0.385	**44**

4.3.2 Error Analysis

For analysing errors, this paper introduces a series of precision recall (PR) curves averaging overall categories which are inspired by DEOD [31]. As the evaluation setting changes, each precision recall curve is guaranteed to be strictly higher than the previous. The curves are as follows:

C75: precision recall curve at $IOU = 0.75$.

C50: precision recall curve at $IOU = 0.50$.

Loc: precision recall curve after localization errors removed.

Sim: precision recall curve after super-category errors are removed.

Oth: precision recall curve after all class errors are removed.

BG: precision recall curve after all background errors are removed.

FN: precision recall curve after all remaining errors are removed.

Fig. 3 shows the precision recall curves of the SSD detector at 4 scales (all, small, medium, and large). For all objects, AP_{50} is 0.647 and ignoring localization errors raises AP to 0.720. After removing all class confusion, AP rises to 0.802. Removing background errors slightly increases AP to 0.848. In conclusion, imprecise localization and class errors are the main reason for confusion.
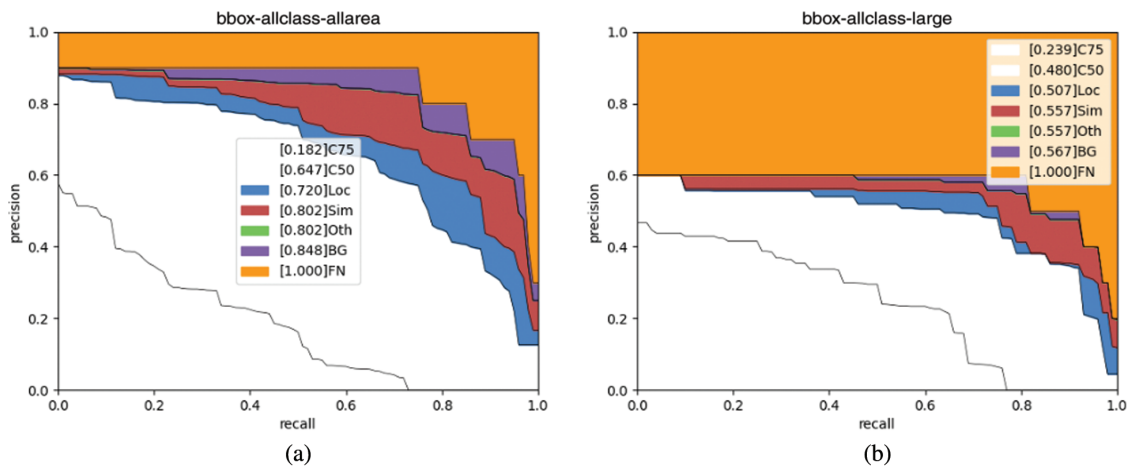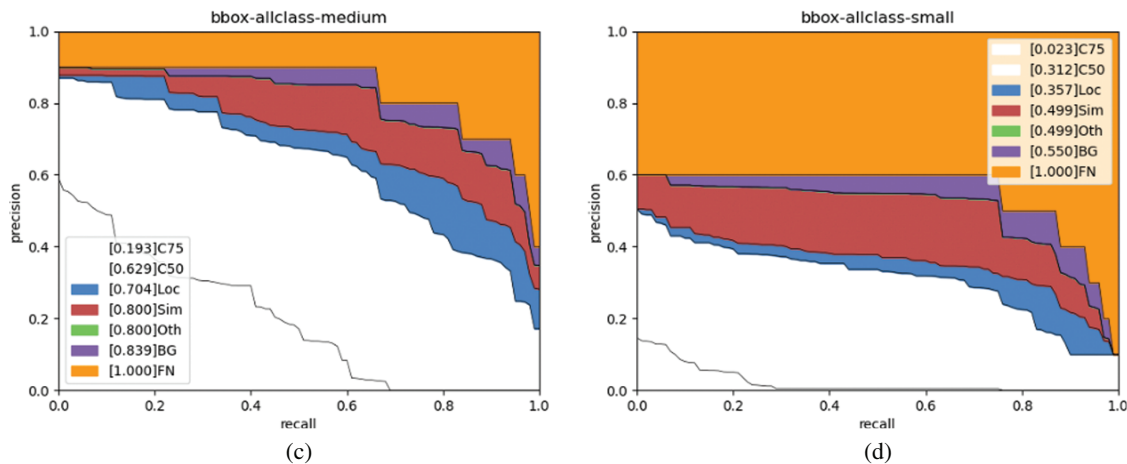

(a)

(b)

Figure 3: (Continued)

Figure 3: PR curves of MCMOD (SSD)

Comparing different scales, it is found that the performance of smaller object detection (0.023 C75 and 0.312 C50) is much worse than the medium (0.193 C75 and 0.629 C50) and large (0.239 C75 and 0.480 C50) object detection. One important factor is those small objects may become vague and have insufficient information after many forward convolutional layers. However, during the implementation of maritime object detection, it is supposed that MVs can be recognised over a relatively long distance, which means that the detector can recognise small objects. As a result of this, improving the detection performance of small objects remains a challenge.

4.3.3 Class-Wise Analysis

Table 3 shows the class-wise performance of each detector on the MCMOD. In the defined 10 maritime object categories, the performances of buoys along with yachts, cruise ships and rafts are worse than the other 6 types. The reason seems to be that the number of these 4 categories is relatively insufficient in datasets and their area are generally small. On the contrary, fishing vessels, speedboats, engineering ships and cargo ships can achieve much better results due to their relatively large number and large scale. Three 0 AP of the cruise ships can be found in Table 3, which may be caused by the scarce number of cruise ships in MCMOD and the imbalanced distribution of images containing cruise ships when randomly dividing the training and validation sets. Another reason is the cruise ships are always far from our cameras (Fig. 4 shows the examples of cruise ships which are highlighted by red circles), and the trained model is hard to distinguish the cruise ships from the background since they are small in view and the contrast between the cruise ships and the background is relatively low. Table 3 points out the existing imbalance between different types of MVs, which should be improved or settled in further development.

In addition, to further evaluate the class-wise performance, the Confusion Matrix [32] is introduced.

Fig. 5 shows the Confusion Matrix of SSD on conditions that the IOU is set to 0.5 and the confidence score threshold is set to 0.3. The confusion matrix shows most categories can be detected correctly. As mentioned above, the performances of small objects, namely yachts, cruise ships and rafts are worse than the other 6 categories. Specifically, it is found that 20% of the raft were identified as speedboats. An intuitive explanation is that these two categories are both small and look nearly the same.

Table 3: Class-wise detection results (AP) on MCMOD

Method	Fast R-CNN (ResNet-101)	Faster R-CNN (ResNet-101)	SSD512 (VGG16)	YOLO v3 608 (Darknet-53)
Fishing vessels	0.483	0.476	**0.495**	**0.469**
Speedboats	0.425	0.414	0.419	0.4
Engineering ships	0.434	0.416	0.399	0.333
Cargo ships	0.657	**0.637**	0.462	0.459
Yachts	0.329	0.349	0.19	0.267
Sailboats	**0.7**	0.5	0.5	0.22
Buoys	0.104	0.165	0.185	0.062
Cruise ships	0	0.101	0	0
Rafts	0.203	0.213	0.197	0.206
Others	0.172	0.167	0.143	0.172

(a)   (b)
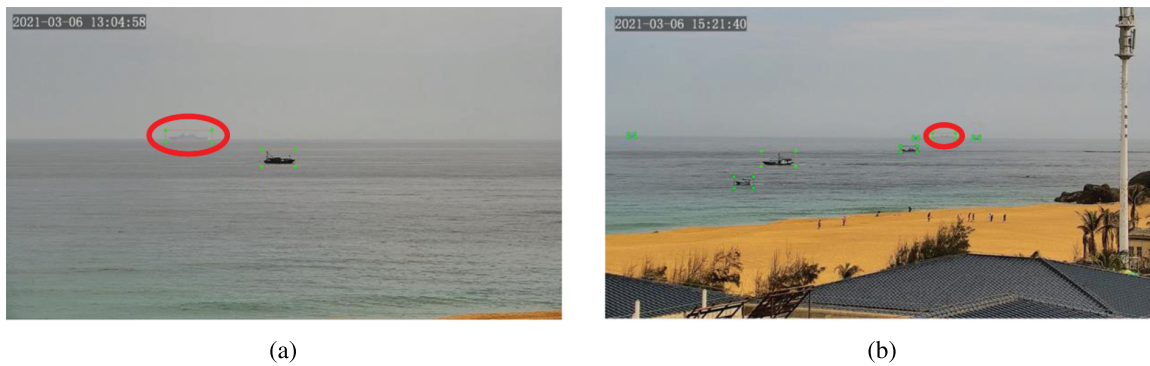
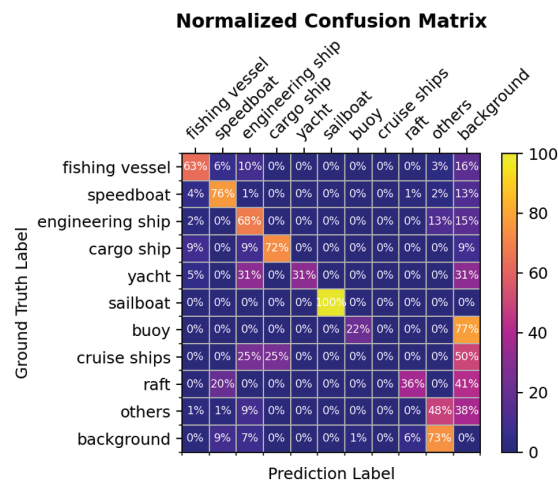**Figure 4:** Images containing cruise ships

Figure 5: Confusion matrix of SSD

4.4 Comparison with SeaShips Dataset

This paper selected SSD to evaluate the detection performance of MCMOD in comparison with SeaShips. Unfortunately, only 7,000 images of the SeaShips dataset are available. The training implementations keep the same.

Table 4 shows the maritime object detection results of experiments. The Recall, Precision and mAP are relatively low in comparison with the SeaShips dataset. The reason is presumably that MCMOD has more categories and more small objects. Although Table 4 shows the better general performance of Seaships, it also indicates the research value in small-scale object detection of MCMOD, which caters to the realistic needs of recognising MVs over a relatively long distance.

Table 4: Detection result of SSD on MCMOD and SeaShips

Method	mAP	AP50	AP75	APS	APM	APL	mAR	FPS
Seaships	**0.676**	**0.955**	**0.787**	0.077	**0.360**	**0.693**	**0.741**	28
MCMOD	0.299	0.647	0.182	**0.172**	0.294	0.385	0.416	28

To prove the MCMOD dataset's research value of small-scale objects, a comparison between SSD's performance of MCMOD and SeaShips was conducted. One important reason is that SSD adopts a structure that helps detect small objects significantly.

Fig. 6 shows the SSD detector's precision recall curves of SeaShips at 4 scales (all, small, medium, and large). Compared to MCMDO (Fig. 3), the detection performance of large and middle objects is better, but the performance of small objects is worse. This greatly proves that the MCMOD dataset shows strong advantages over detecting small objects. This article hopes this will foster further research and innovation.
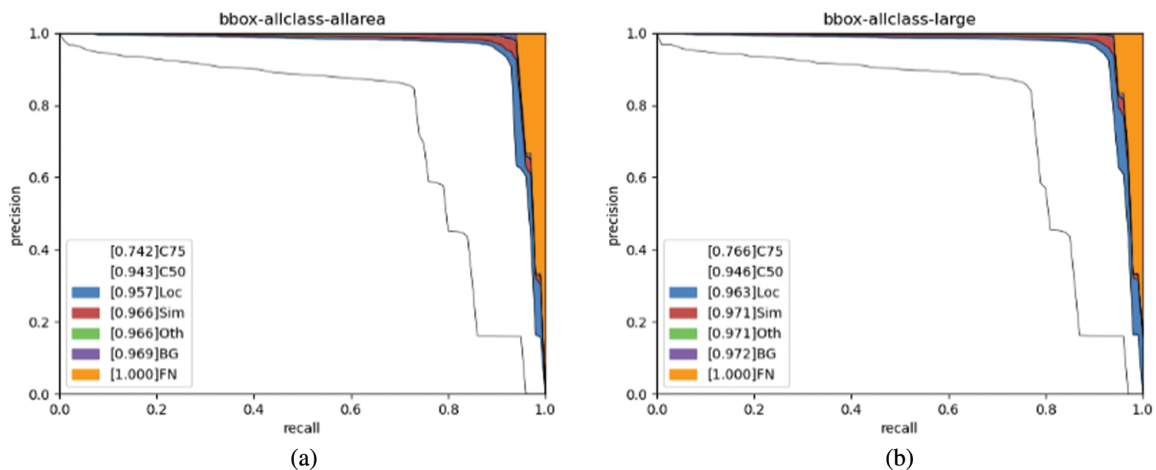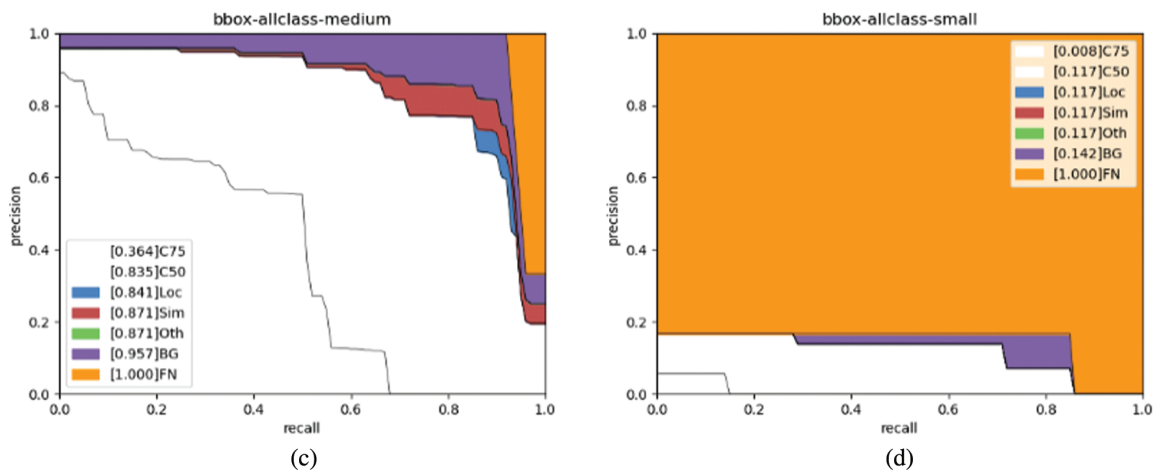

Figure 6: (Continued)

Figure 6: PR curves of SeaShips

5 Conclusion

In this paper, a multi-category large-scale dataset for maritime object detection named MCMOD is proposed. 16166 images along with 98,590 annotated maritime objects are classified into 10 classes under various environmental conditions. To evaluate MCMOD, this paper used 4 state-of-the-art object detection methods for training and testing. The obtained results show that the approaches need to be further improved to enhance their capabilities in the maritime domain and this paper points out a potential research direction of small object detection which can facilitate maritime staff to recognise MVs over a relatively long distance. In addition, the MCMOD dataset was compared with other datasets like SeaShips and the results indicate that the MCMOD dataset shows more value on small objects (the area it takes in the view, not the size in the real world). Overall, MCMOD classifies more types of MVs and covers more small-scale objects, which can facilitate the trained detectors to recognise more types of MVs and detect MVs in a relatively long distance. Further research can pay attention to the following aspects: 1) With the help of MCMOD, enhance the detection algorithms' capabilities in the maritime domain and small objects; 2) Extend the existing dataset with on-board data, in case of some types of MVs do not sail close to the seashore, which makes them remain small in the on-shore views; 3) Extend the existing dataset with some other types of data such as Near Infrared (NIR) images to improve the capabilities under sightless conditions. Further research will focus on developing a new algorithm for maritime object detection referring to existing various small object detectors such as Feature-Enhanced RefineDet [33] in the future and will try to exploit the potential of the MCMOD dataset for small object detection. Detecting small objects would be particularly meaningful and challenging. On the other hand, a large number of maritime object detection models based on deep learning have been proposed, but due to the lack of universal evaluation criteria, it is difficult to compare different improved models [34]. Thus, the development of universal evaluation criteria is also an important part of future work. As a large-scale dataset which covers most types of MVs (except military ones), the proposed MCMOD will foster further research and innovation in the field of maritime object detection.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] W. H. Cho, S. K. Kim, M. H. Na and I. S. Na, "Fruit ripeness prediction based on DNN feature induction from sparse dataset," *Computers, Materials & Continua*, vol. 69, pp. 4003–4024, 2021.

[2] Q. M. Ilyas and M. Ahmad, "An enhanced deep learning model for automatic face mask detection," *Intelligent Automation and Soft Computing*, vol. 31, pp. 241–254, 2022.

[3] C. Song, X. Cheng, Y. Gu, B. Chen and Z. Fu, "A review of object detectors in deep learning," *Journal of Artificial Intelligence*, vol. 2, pp. 59–77, 2020.

[4] M. Everingham, L. Van Gool, C. K. Williams, J. Winn and A. Zisserman, "The pascal visual object classes (voc) challenge," *International Journal of Computer Vision*, vol. 88, no. 2, pp. 303–338, 2010.

[5] A. Krizhevsky and G. Hinton, "Learning multiple layers of features from tiny images," *Handbook of Systemic Autoimmune Diseases*, vol. 1, no. 4, 2009. [Online]. Available: http://www.cs.utoronto.ca/~kriz/learning-features-2009-TR.pdf.

[6] T. Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona *et al.,* "Microsoft coco: Common objects in context," in *European Conf. on Computer Vision*, Zurich, Switzerland, pp. 740–755, 2014.

[7] E. Gundogdu, B. Solmaz, V. Yücesoy and A. Koc, "Marvel: A large-scale image dataset for maritime vessels," in *Asian Conf. on Computer Vision*, Taipei, Taiwan, pp. 165–180, 2016.

[8] M. M. Zhang, J. Choi, K. Daniilidis, M. T. Wolf and C. Kanan, "VAIS: A dataset for recognizing maritime imagery in the visible and infrared spectrums," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) Workshops*, Boston, MA, USA, pp. 10–16, 2015.

[9] Z. Shao, W. Wu, Z. Wang, W. Du and C. Li, "Seaships: A large-scale precisely annotated dataset for ship detection," *IEEE Transactions on Multimedia*, vol. 20, no. 10, pp. 2593–2604, 2018.

[10] Y. Zheng and S. Zhang, "Mcships: A large-scale ship dataset for detection and fine-grained categorization in the wild," in *2020 IEEE Int. Conf. on Multimedia and Expo (ICME)*, London, United Kingdom, pp. 1–6, 2020.

[11] M. Tello, C. López-Martínez and J. J. Mallorqui, "A novel algorithm for ship detection in SAR imagery based on the wavelet transform," *IEEE Geoscience and Remote Sensing Letters*, vol. 2, no. 2, pp. 201–205, 2005.

[12] C. Zhu, H. Zhou, R. Wang and J. Guo, "A novel hierarchical method of ship detection from spaceborne optical image based on shape and texture features," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 9, pp. 3446–3456, 2010.

[13] J. J. Hwang, S. H. Pak and G. Y. Park, "Marine object detection based on kalman filtering," *Journal of Information and Communication Convergence Engineering*, vol. 9, pp. 347–352, 2011.

[14] G. Yang, B. Li, S. Ji, F. Gao and Q. Xu, "Ship detection from optical satellite images based on sea surface analysis," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 3, pp. 641–645, 2013.

[15] K. Kim, S. Hong, B. Choi and E. Kim, "Probabilistic ship detection and classification using deep learning," *Applied Sciences*, vol. 8, no. 6, pp. 936, 2018.

[16] S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017.

[17] V. Marie, I. Bechar and F. Bouchara, "Real-time maritime situation awareness based on deep learning with dynamic anchors," in *2018 15th IEEE Int. Conf. on Advanced Video and Signal Based Surveillance (AVSS)*, Auckland, New Zealand, pp. 1–6, 2018.

[18] R. B. Girshick, "Fast R-CNN," in *2015 IEEE Int. Conf. on Computer Vision (ICCV)*, Santiago, Chile, pp. 1440–1448, 2015.

[19] D. K. Prasad, H. Dong, D. Rajan and C. Quek, "Are object detection assessment criteria ready for maritime computer vision?," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 12, pp. 5295–5304, 2019.

[20] J. Deng, W. Dong, R. Socher, L. -J. Li, K. Li *et al.,* "Imagenet: A large-scale hierarchical image database," in *2009 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Miami, FL, USA, pp. 248–255, 2009.

[21] M. Everingham, S. A. Eslami, L. Van Gool, C. K. Williams, J. Winn *et al.,* "The pascal visual object classes challenge: A retrospective," *International Journal of Computer Vision*, vol. 111, no. 1, pp. 98–136, 2015.

[22] S. Yang, P. Luo, C. -C. Loy and X. Tang, "Wider face: A face detection benchmark," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, Nevada, USA, pp. 5525–5533, 2016.

[23] S. Zhang, R. Benenson and B. Schiele, "Citypersons: A diverse dataset for pedestrian detection," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, pp. 3213–3221, 2017.

[24] C. Wah, S. Branson, P. Welinder, P. Perona and S. Belongie, "The caltech-ucsd birds-200-2011 dataset," *Computation & Neural Systems Technical Report*, 2010-001, California Institute of Technology, Pasadena, CA, 2011.

[25] R. Smith, C. Gu, D. -S. Lee, H. Hu, R. Unnikrishnan *et al.,* "End-to-end interpretation of the French street name signs dataset," in *European Conf. on Computer Vision*, Amsterdam, The Netherlands, pp. 411–426, 2016.

[26] K. Chen, J. Wang, J. Pang, Y. Cao, Y. Xiong *et al.,* "MMDetection: Open mmlab detection toolbox and benchmark," arXiv preprint arXiv:1906.07155, 2019.

[27] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury *et al.,* "Pytorch: An imperative style, high-performance deep learning library," *Advances in Neural Information Processing Systems*, vol. 32, pp. 8026–8037, 2019.

[28] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed *et al.,* "Ssd: Single shot multibox detector," in *European Conf. on Computer Vision*, Amsterdam, The Netherlands, pp. 21–37, 2016.

[29] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," arXiv preprint arXiv:1804.02767, 2018.

[30] J. Wang, T. Zhang, Y. Cheng and N. Al-Nabhan, "Deep learning for object detection: A survey," *Computer Systems Science and Engineering*, vol. 38, no. 2, pp. 165–182, 2021.

[31] D. Hoiem, Y. Chodpathumwan and Q. Dai, "Diagnosing error in object detectors," in *European Conf. on Computer Vision*, Firenze, Italy, pp. 340–353, 2012.

[32] J. T. Townsend, "Theoretical analysis of an alphabetic confusion matrix," *Perception & Psychophysics*, vol. 9, no. 1, pp. 40–50, 1971.

[33] L. Zhao and M. Zhao, "Feature-enhanced RefineDet: Fast detection of small objects," *Journal of Information Hiding and Privacy Protection*, vol. 3, no. 1, pp. 1–8, 2021.

[34] R. Zhang, S. Li, G. Ji, X. Zhao, J. Li *et al.,* "Survey on deep learning-based maine object detection," *Journal of Advanced Transportation*, vol. 2021, Article ID 5808206, 18 pages, 2021. https://doi.org/10.1155/2021/5808206.