



# Open-set marine object instance segmentation with prototype learning

Xing Hu<sup>1</sup> · Panlong Li<sup>1</sup> · Hamid Reza Karimi<sup>2</sup> · Linhua Jiang<sup>3</sup> · Dawei Zhang<sup>1</sup>

Received: 14 September 2023 / Revised: 26 October 2023 / Accepted: 13 May 2024 / Published online: 28 May 2024  
© The Author(s) 2024

## Abstract

The ocean world is full of Unknown Marine Objects (UMOs), making it difficult to deal with unknown ocean targets using the traditional instance segmentation model. This is because the traditional instance segmentation networks are trained on a closed dataset, assuming that all detected objects are Known Marine Objects (KMOs). Consequently, traditional closed-set networks often misclassify UMOs as KMOs. To address this problem, this paper proposes a new open-set instance segmentation model for object instance segmentation in marine environments with UMOs. Specifically, we integrate two learning modules in the model, namely a prototype module and an unknown learning module. Through the learnable prototype, the prototype module improves the class's compactness and boundary detection capabilities while also increasing the classification accuracy. Through the uncertainty of low probability samples, the unknown learning module forecasts the unknown probability. Experimental results illustrate that the proposed method has competitive known class recognition accuracy compared to existing instance segmentation models, and can accurately distinguish unknown targets.

**Keywords** Instance segmentation · Prototype learning · Marine objects · Open-set conditions

## 1 Introduction

The Earth's surface is covered by more than 70% ocean, which includes numerous biological species and natural resources. Human exploration of the ocean has been ongoing, and with the development of modern exploration technology,

such as underwater robots, many previously unknown areas of the ocean have been uncovered. However, only 5% of the ocean floor has been explored so far, with 95% remaining unknown due to the vastness of the ocean. Currently, artificial intelligence has empowered deep learning-based marine image analysis, which has become a popular research topic. Various aspects of marine analysis, such as marine object detection [1, 2], and marine animal segmentation [3], have made significant progress. However, when a well-trained model encounters a "new class" or "different knowledge", it tends to misclassify the objects. In other words, the model assigns them to pre-defined categories [4]. Unknown Marine Objects (UMOs) with unknown categories frequently appear in real ocean scenes, making it challenging not only to label many known categories [5] but also to identify the locations of UMOs. Traditional detection or segmentation models, such as Mask RCNN [6], SOLO [7], and YOLOX [8], are unable to handle these "unknown classes". In practical systems, for the sake of performance and safety, it is crucial to make every effort to reject unknown objects to prevent irreparable losses caused by classification errors, such as misidentifying a peculiar-looking branch as an underwater robot. To address the problem of object instance segmentation of marine environment objects with UMO, an open

---

✉ Hamid Reza Karimi  
hamidreza.karimi@polimi.it

✉ Linhua Jiang  
11594@zjhu.edu.cn

✉ Dawei Zhang  
dwzhang@usst.edu.cn

Xing Hu  
huxing@usst.edu.cn

Panlong Li  
lipanlong25@126.com

<sup>1</sup> School of Optical Electrical and Computer Engineering, University of Shanghai for Science and Technology, No.516 Jungong Road, 200093 Shanghai, China

<sup>2</sup> Department of Mechanical Engineering, Politecnico di Milano, 20156 Milan, Italy

<sup>3</sup> ISEP-Sorbonne Joint Research Lab, 10 Rue de Vanves, 92130 Paris, France

instance segmentation model with prototype learning is proposed. This model aims to improve the misclassification issue encountered by traditional closed-set models when they encounter unknown objects. In the proposed model, we first separate the known objects of different categories in the feature space as much as possible, while minimizing the feature differences between individual classes to obtain a robust closed set classifier. Then, on this basis, the unknown probability is predicted by low score samples. Specifically, we integrate a prototype module and an unknown learning module into the Mask-RCNN model, which imparts the model with the capability for open-set detection. The advantage of utilizing a prototype is that it can enhance the classification accuracy of closed-sets and identify the open world [9]. By adding a prototype module, known classes become more compact in the feature space, while the unknown learning module optimizes the uncertainty of low-probability samples within the classifier. During the actual testing stage, the unknown probability of an instance determines whether it is detected as an unknown object. To validate the effectiveness of our method and assess the actual impact of each module, we use the Trashcan dataset and the CH-DUTUSEG dataset to detect closed sets and open sets. The results of the model under different datasets demonstrate significant progress in the open-set index while ensuring closed-set accuracy. Our model reduces the error of taking an unknown class as a known class. The main contributions of this paper are highlighted as follows:

- This study introduces prototype learning into the open set instance segmentation model, thereby enhancing the accuracy of the model.
- An unknown learning module is incorporated, and the optimization of the unknown boundary is achieved by training low-scoring samples, thereby enhancing the models capacity for identifying unknown objects.
- Given the limited availability of marine life datasets, we extract samples from the existing marine dataset DUT-USEG and curate a novel dataset called CH-DUTUSEG for the purpose of model validation.

This article is structured as follows: In Sect. 2, we review related work. In Sect. 3, we provide a detailed introduction to the prototype module and unknown learning module. Section 4 discusses the experimental details and main results. Finally, in Sect. 5, we summarize our work.

## 2 Related work

**Application of Deep Learning in Ocean Target Analysis.** With the development of deep learning technology, numerous scholars have investigated the application of deep learning in

the marine field. In 2020, Tseng et al. [10] realized the automatic measurement network of fish body length using a CNN network. Siddiqui et al. [11] proposed a visual method based on deep learning to classify fine-grained fish. Reus et al. [12] presented a machine learning approach that uses CNN to estimate the coverage rate of seagrass by describing seagrass patches and superpixels. Ma et al. [13] utilized a fusion algorithm to collect and integrate face image resources from videos, trained a face recognition model using R-CNN, and developed an application platform for crew face recognition and positioning analysis on ships. Wang et al. [14] provided an overview of recent developments in marine biological identification and a detailed analysis of the benefits and drawbacks of deep learning in this area. In order to address the issue of underwater degradation, Chen et al. [15] proposed an underwater scene semantic segmentation network (USSSN), which may minimize artifacts and preserve the integrity of foreground objects while enhancing photos. These examples indicate the growing maturity of deep learning models in the marine field.

**Instance Segmentation Model.** The fundamental idea of instance segmentation involves first detecting instances in an image and then generating a segmentation mask for each detected instance. Among these methods, Mask RCNN [6] evolved from Fast RCNN [16] by incorporating a mask branch into the target detection network to predict instance segmentation results. PointRend [17] treats instance segmentation as a rendering problem in image processing, producing superior masking results compared to Mask RCNN. Another idea of two-stage instance segmentation is to perform pixel-level semantic segmentation first, followed by classification through clustering and other post-processing techniques [18]. Influenced by the research on single-stage target detection, single-stage instance segmentation model has also been explored. YOLACT [19] uses different layers to generate mask coefficients and prototype masks, maintaining spatial consistency and near real-time speed. Using the concept of a class activation diagram to build the case activation layer and sparse the corresponding connection, Cheng et al. [20] introduced an unique instance segmentation technique in 2022. However, these closed-set instance segmentation algorithms typically require strong supervision and struggle to reject unknown objects.

**Open Set Recognition.** The Open Set Recognition (OSR) proposal aims to overcome the limitations of models in real-world situations. OSR models are classified as either generative or discriminative depending on the modeling form [21]. SVM [22, 23] was first used in the discriminant model to minimize the risk of open sets and optimize the space occupied by unknown classes. With the development of deep learning, Zheng et al. [24] takes advantage of RPN's insensitivity to categories, and takes some candidate frames with high confidence and no labeled information as open

set position objects, and then discriminates open set objects by clustering. Through comparative learning and incremental learning, Joseph et al. [25] introduced a new field of open-world object detection and achieved open-set object detection. In the generation model, Neal et al. [26] used GANs to expand the training set samples, and generated the synthetic open set samples for model training. However, there are still some practical differences between the generated samples and the open set samples. Prototype learning is also widely used in open set identification, among which Yang et al. [9] first applied prototype learning to convolutional networks, which proved that the integration of prototypes improved the robustness of closed set classification and made it possible to identify unknown samples. Lu et al. [27] proposed a new framework for prototype mining and learning, and made open set identification after considering the multi-attributes of prototype sets.

### 3 Methodology

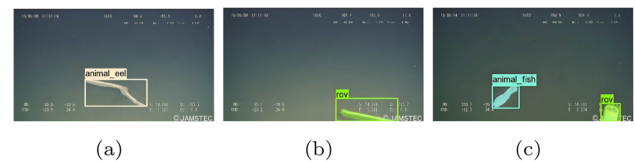
#### 3.1 Preparations

Our model design is based on two basic premises: 1) the real ocean scene is full of "unknown" possibilities, and open-set recognition is suitable for ocean scenes. 2) when faced with "unknown" objects, the traditional model will misclassify the objects, as shown in Fig. 1. Therefore, we use  $D = (x, y)$ ,  $x \in X$ ,  $y \in Y$  to represent the scene dataset, where  $x$  represents a sample instance and  $y = (c, b, m)$  represents the label of this sample instance, including category  $c$ , detection box  $b$ , and segmentation mask  $m$ . To reflect the complexity of the ocean scene and the open-set encountered in the test as much as possible, we use the  $D_{train}$  data set to train our model, in which  $D_{train}$  contains the known class  $K$ , which is expressed as  $C_K = 1, \dots, K$ . We use  $D_{test}$  data set to test our model, in which  $D_{test}$  contains known class  $k$  and other classes  $C_U$  that did not appear in training, which can be expressed as  $C_U = K + 1$ . Our goal is to make the model detect not only the known classes in  $D_{test}$ , but also the location of unknown classes, thus reducing the probability of the wrong classification.

At the same time, we consider that a picture may contain samples of both known class  $C_K$  and unknown class  $C_U$  [28]. So, we have made the following preparations: Try to avoid unknown objects in training, and then better distinguish background class  $C_B$  from unknown class  $C_U$ .

#### 3.2 Model architecture

Considering the high accuracy and robustness of Mask RCNN model, we use Mask RCNN as our baseline architecture. During baseline learning, we found that the baseline



**Fig. 1** Incorrect classification. **a** A branch is classified as an eel; **b** a branch is classified as a remotely operated vehicle (ROV); and **c** a plastic bag is classified as a fish

tends to classify unknown objects as background or known classes with low scores. This demonstrates that while the classic model has some rejection potential, it will result in incorrect classification since it lacks significant separation potential for unidentified class traits. To enhance feature separation and unknown identification. According to Fig. 3, we add a prototype module and an unknown learning module to the foundation network.

#### 3.3 Prototype module

In this section, we introduce the feature learning module, which makes different categories more separated and the same category more compact through the learnable prototype, according to Fig. 2. We classify features according to their distance scores from different prototypes. We use  $m_i$  to represent the prototype, where  $i \in 1, 2, \dots, K$  represents the known class index corresponding to the prototype. Quantitatively, we use the Euclidean distance between features and different prototypes to measure the probability score. Where the Euclidean distance is:

$$d(f(x), m_i) = \|f(x) - m_i\|_2 \quad (1)$$

Among them,  $f(x)$  represents the features extracted in the early stage, and  $d(f(x), m_i)$  represents the Euclidean distance from the sample features to the corresponding prototype. As shown in Fig. 2, during the training process, the features should be as close as possible to the corresponding prototype, hence we define  $loss_d$  as:

$$loss_d = -\frac{1}{2N} \sum_{j=1}^N d(f_j, m_i)^2 \quad (2)$$

where  $N$  is the total number of features. At the same time, we introduced classification loss to strengthen the model's robustness and improve the separation of the prototype. The stability of the feature during training is aided by classification loss, which bases its label judgment on the distance between each feature and the prototype, as shown in Fig. 2. The Euclidean distance of each feature and each category prototype is calculated to get a distance distribution matrix  $D$ :

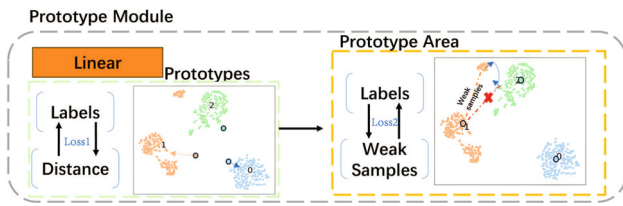


Fig. 2 Prototype module

$$D_{ij} = -d(f_j, m_i)^2 \tag{3}$$

where  $i \in 1, 2, \dots, K$  and  $j \in 1, 2, \dots, N$ . In addition, a background class prototype is kept around to filter out negative samples. Then the cross-entropy loss is applied on  $D$ , the  $loss_1$ :

$$loss_1 = -\frac{1}{N} \sum_{j=1}^N Y_j * \log \left( \frac{\exp(D_i)}{\sum_{i=0}^K \exp(D_i)} \right) \tag{4}$$

We also consider the impact of some atypical points. Figure 2 illustrates how some feature points could be rather distant from the associated prototype, which could result in incorrect classification. Therefore, in order to penalize the incorrect classification of boundary samples, we include a prototype region module. We select some low-scoring foreground and background samples (the number of Weak Samples is  $M$ ) to make cross entropy loss and loss function:

$$loss_2 = -\frac{1}{M} \sum_{j=1}^M Y_j * \log \left( \frac{\exp(D_i)}{\sum_{i=0}^K \exp(D_i)} \right) \tag{5}$$

Finally, we define the prototype loss function as:

$$loss_p = \sigma_1 * loss_d + loss_1 + \sigma_2 * loss_2 \tag{6}$$

It is worth noting that in a closed-set detection environment, we need only determine the category score based on the corresponding distance, as follows:

$$s_i(x) \propto -\|f(x) - m_i\|^2$$

$$Y_x = \arg \max_{0 \leq i \leq k} s_i(x) \tag{7}$$

The  $s_i$  above represents the score of the feature, and  $y$  represents the label measured by the sample  $x$ . In the open-set environment, if we only rely on the threshold of measurement to debug the model, it is bound to make the model, like the traditional model, unable to consider the accuracy of the closed set and the ability of open set recognition. Therefore, we consider adding an unknown learning module to ensure both as far as possible.

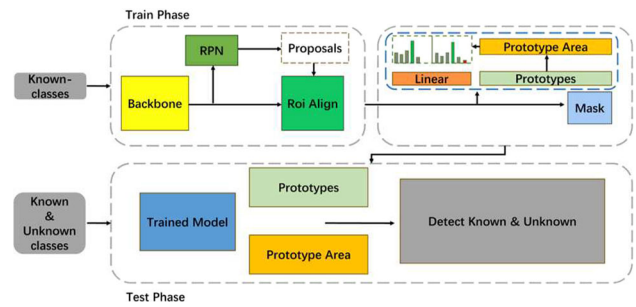


Fig. 3 The process of our method

### 3.4 Unknown learning module

This section focuses on how to make the model learn the ability to judge the unknown without sample data. Based on the premise mentioned above, we find that traditional networks tend to classify unknown objects into low-scoring backgrounds or low-scoring known categories. Based on the analysis of the above phenomena, we believe that the traditional network has a certain ability to identifying unknown objects. It's just that the traditional model regards space as a global and closed-set, and can't reject unknown objects. Therefore, we extend the  $K + 1$  classifier to the  $K + 2$  (including unknown probability) classifier in training. Our focus is on the uncertainty of low-scoring samples in the learning and training process.

During the training process, we select an equal number of low-scoring known samples and background samples as the training data. We define the probability score of each class of  $K + 2$ (including background and unknown probability) as follows:

$$s_u = \exp(soft_u) / \left( \sum_{j=0}^{K+1} \exp(soft_j) - \exp(soft_c) \right)$$

$$s_i = \exp(soft_i) / \sum_{j=0}^{K+1} \exp(soft_j), \quad i \in [0, K] \tag{8}$$

In the above formula, "soft" represents the score assigned by the extended classifier, while "c" denotes the actual label of the training sample. We utilize features from low probability samples that resemble those from unknown samples to establish the boundary for unknown samples. Given this information, the loss function is defined as follows:

$$loss_k = -\log(s_i)$$

$$loss_u = -\gamma * \log(s_u)$$

$$loss_{ul} = loss_k + \tau * loss_u \tag{9}$$

where  $\gamma$  is the degree factor of unknown probability to the probability of unreal tag, indicating the possibility of

unknown in tag, and  $\tau$  is the coefficient. We set  $\gamma = (1-s)^T * s$  to optimize the unknown probability score, where  $s$  is the true label probability.

### 3.5 Detection optimization and prediction

According to Sects. 3.3 and 3.4, in this section, we will outline our open model, as shown in Fig. 3. Firstly, to address the issue of feature separation and compactness, we have added a prototype module. Secondly, to enable the model to learn the ability to distinguish the unknown, we added the unknown learning module. We define the loss function of the whole detection part as:

$$loss_{dect} = loss_{ul} + loss_p \quad (10)$$

Firstly, the object category is assigned based on the prediction score:

$$Y_x = \begin{cases} i, & \arg \max_{1 \leq i \leq k} (soft_i) > soft_u \\ K + 1, & \arg \max_{1 \leq i \leq k} (soft_i) < soft_u \end{cases} \quad (11)$$

If it is known, we determine the category label of the known class based on the distance score of the prototype to ensure the highest possible detection accuracy for the known class. The relationship between probability and distance is as follows:

$$s_i = \rho * \exp(D_i) / \sum_{i=0}^K \exp(D_i) \quad (12)$$

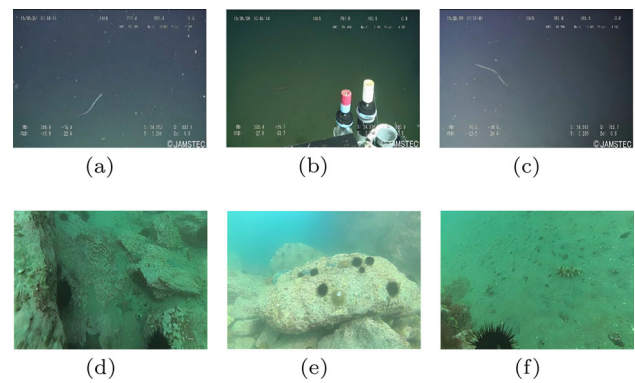
Here,  $\rho$  represents the degree coefficient of the closed set, which describes the influence of the open set on the score of the closed set. We set  $\rho = 1$ .

## 4 Experiment

### 4.1 Experimental setup

**Baseline Method.** We use the two-stage Mask RCNN network as the baseline for comparison. Simultaneously, we integrate ablation experiments to select specific experimental parameters in order to explore the influence of different modules on the result.

**Validation Metrics.** Our goal is to maximize the model's accuracy in detecting known categories, approaching the performance in a closed-set scene, while ensuring stable identification of unknown categories. Given the above objectives and specific scenarios, we use the Mean Average Precision (mAP) to assess the test accuracy of known classes. At the same time, based on our research on open-sets and the



**Fig. 4** Among them, Figures a, b, and c are samples from the Trashcan dataset, and Figures d, e, and f are samples from the CH-DUTUSEG dataset

concept of open-set classification, we use Absolute Open-Set Error (AOSE) to measure the number of unknown errors of model classification. Furthermore, we consider the relationship between open and closed sets and use Wilderness Impact (WI) to measure the degree to which unknown objects are misclassified as known categories, where  $WI = (p_k / p_{k \cup u} - 1)$ .

**Datasets.** To evaluate the model's effectiveness in real-world marine applications, we conduct experiments using the Trashcan dataset and the CH-DUTUSEG dataset. Figure 4 provides an overview of the datasets. The Trashcan dataset has 6065 training and 1147 testing images with 8 non-garbage and 14 garbage classes. Analyzing this dataset, we used 8, 14, and 22 categories to verify the prototype module functionality and Trashcan8-14 and Trashcan14-8 as two benchmarks to verify open-set performance. We select 400 images containing 1191 instances from the DUT-USEG dataset as the CH-DUTUSEG dataset.

**Setup Details.** We use ResNet-50 and Feature Pyramid Network (FPN) as the backbone of the improved model and baseline. Regarding hyperparameter settings, we set  $\sigma_1$  to 0.001,  $\sigma_2$  to 0.0001,  $\tau$  to 0.1, and T to 1.15. For optimizer and learning rate settings, we use an SGD optimizer with an initial learning rate of 0.08, a momentum of 0.9, and a weight decay of 0.0001.

### 4.2 Main results

Firstly, we verify the positive effect of adding a Prototype module on Trashcan and CH-DUTUSEG, in which the test set only contains known classes. The results are shown in Table 1.

The comparison findings in the above table indicate that adding a prototype module will increase the model closed set accuracy. We then compared the improved model to the baseline on the Trashcan8-14 and Trashcan14-8 datasets. Table 2 shows the comparison results on the Trashcan dataset. Table 3

**Table 1** Test accuracy of different methods

Datasets	Methods	$AP^b$	$AP^m$
Trashcan-8	Baseline	58.68	56.57
	Baseline+PM	<b>60.78</b>	<b>58.75</b>
Trashcan-14	Baseline	63.00	57.38
	Baseline+PM	<b>65.37</b>	<b>58.96</b>
Trashcan	MaskRCNN [29]	55.40	55.30
	SparseInst [20]	–	56.05
	Baseline+PM	<b>60.55</b>	<b>56.21</b>
CH-DUTUSEG	Baseline	59.05	48.93
	SparseInst [20]	–	50.99
	Baseline+PM	<b>60.36</b>	<b>51.16</b>

**Table 2** Comparison results of different models on Trashcan subset

Datasets	Methods	$AP^b$	$AP^m$	$WI$	$AOSE$
Trashcan 8-14	Baseline	56.42	54.73	0.429	2624
	OWOD[25]	56.13	–	<b>0.259</b>	<b>1589</b>
	Baseline+PM	56.92	54.97	0.348	2130
	Baseline+UL	56.70	54.88	0.266	1630
	Ours	<b>57.08</b>	<b>55.39</b>	0.26	1591
Trashcan 14-8	Baseline	61.69	56.17	0.279	1675
	OWOD[25]	61.05	–	<b>0.147</b>	<b>881</b>
	Baseline+PM	<b>64.31</b>	<b>58.02</b>	0.227	1361
	Baseline+UL	59.94	55.22	0.148	888
	Ours	61.91	56.26	0.171	1030

**Table 3** Comparison results of different models on CH-DUTUSEG

Datasets	Methods	$AP^b$	$AP^m$	$WI$
CH-DUTUSEG	Baseline	57.75	47.68	0.112
	OWOD[25]	57.86	–	0.092
	Baseline+PM	58.98	49.95	0.105
	Baseline+UL	57.59	46.94	<b>0.087</b>
	Ours	<b>61.09</b>	<b>52.72</b>	0.097

shows the comparison results on the CH-DUTUSEG dataset. As the CH-DUTUSEG dataset has fewer samples, we only used  $WI$  as the measurement metric.

Through the above comparison, we find that the accuracy of known classes is greatly improved by adding a prototype module, and the open set ability of models is improved by adding an unknown learning module. Finally, our model demonstrates an overall improvement compared to the baseline. Figure 5 below shows our prediction results on Trashcan.

**Fig. 5** The comparison of results between baseline (top) and improved models (bottom)**Table 4** The influence of different feature dimensions

$dim$	$AP^b$	$AP^m$
2	37.81	33.53
14	63.28	58.09
100	<b>65.37</b>	<b>58.96</b>
512	63.09	57.69

**Table 5** The impact of  $T$  setting on the results

$T$	$AP^b$	$AP^m$	$WI$
1	61.76	55.79	0.171
1.15	<b>61.91</b>	56.26	<b>0.171</b>
1.5	61.64	<b>56.41</b>	0.19

**Table 6** The effect of closed-set degree coefficient  $\rho$  on results

$\rho$	$AP^b$	$AP^m$	$WI$
1	<b>61.91</b>	<b>56.26</b>	0.171
$1 - soft_u$	60.5	55.47	<b>0.162</b>
$(1 - soft_u) * (1 + soft_u)$	61.29	55.94	0.172

### 4.3 Ablation experiments

In this section, we explore the optimal performance of the relevant modules and investigate the optimal hyper-parameters in the experiments and their possible effects. Regarding the prototype module, we discuss the potential layer's dimensions (i.e., constructed feature dimensions) impact on known class accuracy. We trained and tested on Trashcan8-14. The results are shown in the table below:

According to the results in the above table, we find that the dimension depth of the prototype features has a very important influence on the accuracy of the model. Considering the accuracy, we make it deeper than the feature dimension of the classification layer. Regarding the unknown learning module, we investigate the impact of the  $T$  setting on the results. We conducted training and testing on the Trashcan14-8 dataset. The results are presented in the table below:

Finally, we analyze the effect of  $\rho$  settings on the model by conducting training and testing on the Trashcan14-8 dataset. The results are displayed in the table below:

## 5 Conclusions

This paper proposes a novel method for open-set instance segmentation in ocean scenes. Building upon the baseline model, we introduce two learning modules, the prototype module, and the unknown learning module. These modules are designed to enhance the accuracy of closed-set classification, allowing the model to maintain stable accuracy in identifying known classes while effectively recognizing unknown classes in open-set scenes. The performance of the model is evaluated on Trashcan and CH-DUTUSEG datasets, demonstrating improved classification accuracy for closed sets and enhanced recognition capability for open sets. Finally, misclassification still exists in the improved model. This is because there is a good chance that unknown samples will be classified incorrectly as background samples. The focus of the following study will be on feature separation between background samples and unknown samples.

**Funding** Open access funding provided by Politecnico di Milano within the CRUI-CARE Agreement.

**Data availability** The datasets analyzed in this study are available upon request.

## Declarations

**Conflict of interest** The authors have no relevant financial or nonfinancial interests to disclose.

**Ethics approval and consent for participate** Not applicable.

**Consent for publication** All authors agreed on the final approval of the version to be published.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Wang, N., Wang, Y., Er, M.J.J.C.E.P.: Review on deep learning techniques for marine object recognition: Architectures and algorithms. *Control Eng Pract.* 118: 104458 (2022)
- Moniruzzaman, M., Islam, S.M.S., Bennamoun, M., Lavery, P.: Deep learning on underwater marine object detection: A survey. In: *Advanced Concepts for Intelligent Vision Systems: 18th International Conference, ACIVS 2017, Antwerp, Belgium, September 18–21, 2017, Proceedings 18*, Springer, pp. 150–160 (2017)
- Li, L., Rigall, E., Dong, J., Chen, G.: MAS3K: An open dataset for marine animal segmentation. In: *International Symposium on Benchmarking, Measuring and Optimization*, Springer, pp. 194–212 (2020)
- Zhou, D.-W., Ye, H.-J., Zhan, D.-C.: Learning placeholders for open-set recognition. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 4401–4410 (2021)
- Salehi, M., Mirzaei, H., Hendrycks, D., Li, Y., Rohban, M.H., Sabokrou, M.J.a.e.-p.: A Unified Survey on Anomaly, Novelty, Open-Set, and Out-of-Distribution Detection: Solutions and Future Challenges. *arXiv preprint arXiv:2110.14051* (2021)
- He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask r-cnn. In: *Proceedings of the IEEE international conference on computer vision*, pp. 2961–2969 (2017)
- Wang, X., Kong, T., Shen, C., Jiang, Y., Li, L.: Solo: Segmenting objects by locations. In: *Computer Vision/CECCV 2020: 16th European Conference, Glasgow, UK, August. 23–28, 2020, Proceedings, Part XVIII 16*, Springer, pp. 649–665 (2020)
- Ge, Z., Liu, S., Wang, F., Li, Z., Sun, J.J.a.e.-p.: YOLOX: Exceeding YOLO Series in 2021. *arXiv preprint arXiv:2107.08430* (2021)
- Yang, H.-M., Zhang, X.-Y., Yin, F., Liu, C.-L.: Robust classification with convolutional prototype learning. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3474–3482 (2018)
- Tseng, C.-H., Hsieh, C.-L., Kuo, Y.-F.J.B.E.: Automatic measurement of the body length of harvested fish using convolutional neural networks. *Science* **189**, 36–47 (2020)
- Siddiqui, S.A., Salman, A., Malik, M.I., Shafait, F., Mian, A., Shortis, M.R.: Arvey ESJJOMS: automatic fish species classification in underwater videos: exploiting pre-trained deep neural network models to compensate for limited labelled data. *Science* **75**(1), 374–89 (2018)
- Reus, G., Möller, T., Jäger, J., Schultz, S.T., Kruschel, C., Hase-nauer, J., Wolff, V., Fricke-Neuderth, K.: Looking for seagrass: Deep learning for visual coverage estimation. In: *2018 OCEANS-MTS/IEEE Kobe Techno-Oceans (OTO)*, IEEE, pp. 1–6 (2018)
- Ma, C., Chen, L., Yang, C., Zhang, W., Li, H.: A Deep Learning Based Personnel Positioning System for Key Cabin of Ship. In: *2019 International Conference on Intelligent Computing, Automation and Systems (ICICAS)*, IEEE, pp. 492–496 (2019)
- Wang, N., Chen, T., Liu, S., Wang, R., Karimi, H.R., Lin, Y.J.N.: Deep learning-based visual detection of marine organisms: a survey. *Neurocomputing* **532**, 1–32 (2023)
- Chen, T., Wang, N., Chen, Y., Kong, X., Lin, Y., Zhao, H.: Karimi, HRJEOAI: Semantic attention and relative scene depth-guided network for underwater image enhancement. *Eng Appl Artif Intell* **123**, 106532 (2023)
- Girshick, R.: Fast r-cnn. In: *Proceedings of the IEEE international conference on computer vision*, pp. 1440–1448 (2015)
- Kirillov, A., Wu, Y., He, K., Girshick, R.: Pointrend: Image segmentation as rendering. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 9799–9808 (2020)
- De Brabandere, B., Neven, D., Van Gool, L.J.a.e.-p.: Semantic Instance Segmentation with a Discriminative Loss Function. *arXiv preprint arXiv:1708.02551* (2017)
- Bolya, D., Zhou, C., Xiao, F., Lee, Y.J.: Yolact: Real-time instance segmentation. In: *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 9157–9166 (2019)
- Cheng, T., Wang, X., Chen, S., Zhang, W., Zhang, Q., Huang, C., Zhang, Z., Liu, W.J.a.e.-p.: Sparse Instance Activation for Real-Time Instance Segmentation. *arXiv preprint arXiv:2203.12827* (2022)

21. Geng, C., Huang, S.-J.: Recent advances in open set recognition: a survey. *IEEE Trans Pattern Anal Mach Intell* **43**(10), 3614–31 (2020)
22. Scheirer, W.J., de Rezende Rocha, A., Sapkota, A., Boulton, T.E.J.I.T.O.P.A.: Toward open set recognition. *IEEE Trans. Softw. Eng.* **35**(7): 1757–72 (2012)
23. Scherrek, M.D., Rigling, B.D.J.I.T.o.A., Systems, E.: Open set recognition for automatic target classification with rejection. *52*(2): 632–42 (2016)
24. Zheng, J., Li, W., Hong, J., Petersson, L., Barnes, N.: Towards open-set object detection and discovery. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3961–70 (2022)
25. Joseph, K., Khan, S., Khan, F.S., Balasubramanian, V.N.: Towards open world object detection. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 5830–40 (2021)
26. Neal, L., Olson, M., Fern, X., Wong, W.-K., Li, F.: Open set learning with counterfactual images. In: *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 613–628 (2018)
27. Lu, J., Xu, Y., Li, H., Cheng, Z., Niu, Y.: Pmal: Open set recognition via robust prototype mining. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 1872–1880 (2022)
28. Han, J., Ren, Y., Ding, J., Pan, X., Yan, K., Xia, G.-S.: Expanding low-density latent regions for open-set object detection. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9591–9600 (2022)
29. Hong, J., Fulton, M., Sattar, J.J.a.e.-p.: TrashCan: A Semantically-Segmented Dataset towards Visual Detection of Marine Debris. *arXiv preprint arXiv:2007.08097* (2020)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.