Computing and Informatics, Vol. 41, 2022, 739–756, doi: 10.31577/cai_2022_3_739

INTELLIGENT ANNOTATION ALGORITHM BASED ON DEEP-SEA MACROBENTHIC IMAGES

Qihang Wu, Yong Liu*, Jianyi Zhang, Yongpan Wang

College of Information Science and Technology
Qingdao University of Science and Technology
Qingdao, 266061, China
e-mail: liuyong@qust.edu.cn

Abstract. In the field of image processing, due to the need of expertise and skills in deep-sea biology and the disadvantages of high labor cost and long time consuming, it has always been a difficult task to mark the images of deep-sea benthic organisms. To solve this problem, this paper proposes a new image intelligent labeling algorithm LACP AL (Localization-Aware-Choice and Pseudo Label Active Learning) which is based on Localization-Aware Active Learning. LACP AL is an active learning framework based on Faster R-CNN, it finds the "valuable" samples from unlabeled samples by clustering algorithm for every training; it selects hard-to-identify samples for manual annotation and further optimizes the model; and it proposes an improved pseudo-labeling mechanism to expand the training set and improve the model accuracy. According to the publicly available dataset provided by 2020 China Underwater Robot Professional Contest, a series of experiments has been done to verify that our algorithm can achieve higher recognition accuracy with fewer training samples compared with the existing algorithms for Marine benthic image recognition.

Keywords: Intelligent labeling, active learning, pseudo-labeling, object detection

Mathematics Subject Classification 2010: 68U10

* Corresponding author

1 INTRODUCTION

Macrobenthic organisms are an important ecological group in the marine environment and are closely related to human life. Some of them are objects of fisheries or aquaculture, with food value, medicinal value and economic value, such as shrimps, crabs and shellfish; some of them are harmful to humans, such as fouling and drilling organisms; elucidating the patterns of changes in the abundance of benthic organisms and their relationship with the productivity and resources of marine organisms is of great significance to the development of aquatic production and the study of the ecosystem of water bodies.

With the rapid development of computer information technology, deep learning has made important breakthroughs in the field of deep-sea macrobenthos research, and traditional machine learning methods are gradually replaced by deep learning-based methods. The success of deep learning in machine vision relies on large-scale labeled data, and the labeling of image data of deep-sea macrobenthic organisms has the following problems:

1. The annotation of deep-sea macrobenthos pictures requires high expertise of the annotators, who need to have marine related knowledge, which increases the annotation cost.
2. Manual annotation can be affected by physical condition, psycho-emotional and other factors, leading to a decrease in annotation quality.

In addition, many benthic images can cause physical fatigue, leading to a decrease in annotation speed.

To cope with the huge dataset annotation effort, Su et al. [1] propose a novel incremental learning framework that combines incremental learning methods with deep convolutional neural networks to achieve fast and efficient learning. This paper investigates the existing annotation algorithms and proposes a deep active learning algorithm LACP AL (Localization-Aware-Choice and Pseudo Label Active Learning) to achieve intelligent annotation of deep-sea macrobenthic images. Classical detection methods are very time-consuming, such as OpenCV Adaboost [2] using sliding window and image pyramid to generate detection frames, or R-CNN [3] using SS (Selective Search) method to generate detection frames. Faster RCNN, on the other hand, discards the traditional sliding window and SS methods and directly uses RPN to generate detection frames, which greatly improves the speed of detection frame generation. The pseudo-labeling [4] mechanism allows us to use pseudo-labeled data while training the model, which can increase the model performance without increasing the model complexity. Our algorithm fully combines the advantages of Faster R-CNN and pseudo-labeling mechanism, and automatically selects images that are difficult to be discriminated by the model for manual labeling to ensure the labeling accuracy, which can effectively handle data annotation of deep-sea benthic images.

Our contributions are as follows:

1. For the selection of the initial training set of the model, the clustering method is used to obtain representative samples from the initial sample set to improve the robustness of the model.
2. In the sample labeling stage, a novel selection algorithm is introduced, which combines the boundary stability selection criteria and classification uncertainty selection criteria, making the model automatically select pseudo-labeled samples with stable boundaries and manually labeled samples that are difficult to discriminate, improving the training speed of the model.
3. A pseudo-labeling mechanism is introduced, and the threshold of pseudo-label selection is made to decrease with model iteration, which mean that the number of pseudo-labeled samples is gradually increased with the robustness of the model to enhance the fitting ability of the ascending model.

2 RELATED WORK

Active learning [5] is based on the principle of manually labeling a small number of samples to quickly obtain a high-quality model, and eventually using the high-quality model to predict the entire sample, with human-assisted modifications to obtain the final labeling; using this feature of active learning, it is possible to quickly obtain a large amount of high-quality labeled data. Therefore, active learning has gradually received due attention, it also have been applied to a variety of computer vision problems, including target classification [6, 7], image segmentation [8, 9], intelligent robot [10, 11], and activity recognition [12, 13] and has been applied to various data labeling tasks, such as the study of deep-sea macrobenthos, satellite images [14] and vehicle images [15], and the active learning process is shown in Figure 1.

After the great progress of deep learning in the field of target detection, researchers started to apply active learning in the field of target detection to cope with the problem of high-cost image annotation due to the demand of big data. The key of active learning lies in how to select the appropriate set of annotation candidates for manual annotation, and the method of selection is called query strategy. The most common query strategy currently used is Uncertainty Sampling, which usually has three ideas in describing the uncertainty of samples or data: Entropy [5], Least Confidence [16], and Margin [17]. Brust [18] proposed three active learning target detection methods based on uncertainty sampling: an incremental learning scheme for deep object detectors without catastrophic forgetting, active learning metrics for detection derived from uncertainty estimates, leverage selection imbalances for active learning, active learning metrics for detection derived from uncertainty estimates, leverage selection imbalances for active learning, where the underlying uncertainty measure module has low dependency and can be replaced using other active learning metrics, and the subsequent aggregation methods can be used not only in the field of target detection but also applied to image segmentation.

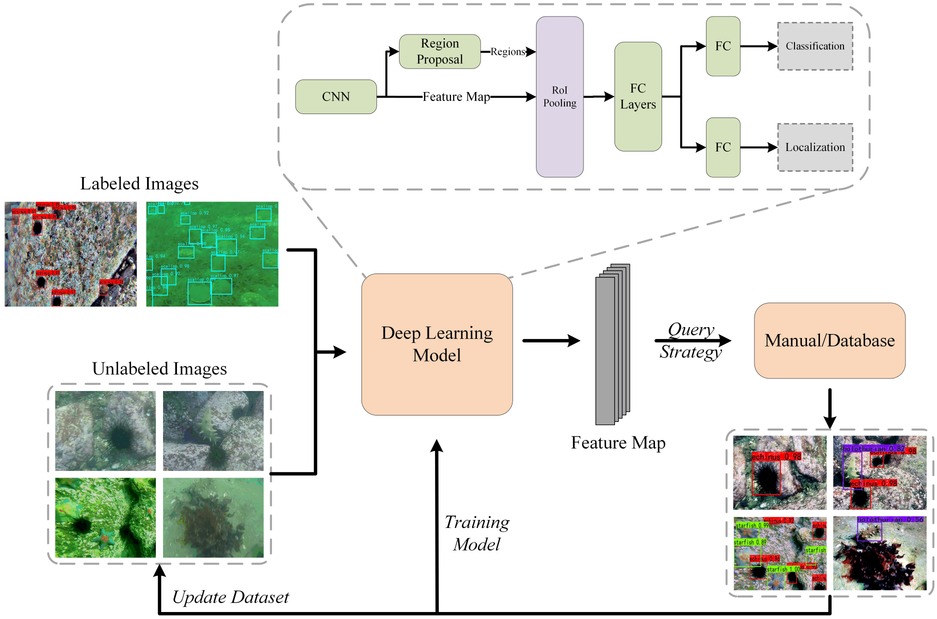


Figure 1. Active learning process

In addition, Roy et al. [19] proposed the 'query-by-committee' model, which selects the most uncertain samples for target detection, dividing active learning into two methods: black-box methods, and white-box methods. Black-box methods do not focus on the underlying network architecture and use confidence levels to select samples; White-box methods combine the method of selecting samples with the network architecture, which is quite different among different networks. The active learning strategy used exploits the detection differences between convolutional layers and requires only a small fraction of the training images to produce better target detection results, but it supports a single batch processing method. Haussmann et al. [20] proposed several active learning scoring functions, including entropy-based functions, core-based functions, etc., using function evaluation to the amount of information in each unlabeled image, and then the selected images are labeled and added to the training set. And experiments show that this sample selection method is more beneficial for model improvement enhancement compared to expert hand-picking. Kao et al. [21] added localization of the detections to active learning to consider both classification and localization results of unlabeled images when selecting unlabeled images for labeling, with better results than the traditional classification output algorithm alone. Desai et al. [22] proposed to combine active learning with multilayer supervised combined with multilayer supervision what will substantially save the annotation effort required to train the model. Iyer et al. [23] used information metrics for active learning sample selection and investigated some optimization problems related to information measures in data aggregation. Kaushal et al. [24] proposed a set of parameterized information metrics for active learning sample selection based on Iyer et al. [23] and experimentally showed that the method can improve the overall accuracy by 2 % to 10 %.

All the above methods of sample selection methods, compared to random selection, are effective in reducing the cost of labeling with improved accuracy. However, these sampling methods do not further filter the selected samples, resulting in the possible existence of similar samples among the selected samples, which can lead to underfitting at the beginning of the iteration when the samples are small, and their wasted manual labeling time far exceeds their effect on model enhancement. To solve the above problems and filter the duplicate samples to ensure that the selected samples have high uncertainty and diversity, this paper proposes a deep active learning algorithm, which introduces clustering algorithm and pseudo-label strategy in active learning and adds a noisy sample selection strategy to the localization stability method to more effectively measure the amount of information contained in the samples.

3 METHODS

In this section, we present our proposed algorithm LACP AL (Localization-Aware-Choice and Pseudo Label Active Learning) for the task of intelligent annotation of deep-sea microbenthic images. The method proposed in this paper consists of an improved active learning framework and an object detection model Faster R-CNN. After inputting the images, our model outputs the corresponding detection boxes, each with its position and size, as well as the confidence level of the category. Algorithm 1 shows the pseudo-code of this algorithm.

Algorithm 1 Localization-Aware-Choice and Pseudo Label Active Learning

```
Data: unlabeled sample set $U$, labeled sample set $L$, recognition model $M$,
  iteration count threshold $N$
Begin For: $i = 1, 2, \ldots, N$
  Train$(M, L)$; // Training recognition model $M$ using labeled sample set $L$
  $T = \mathrm{Test}(M, U)$;
  $S = \mathrm{Select}(M, U)$; // Sample selection using the improved selection algorithm
  Label$(S)$; // Labeling of samples in the set $S$
  $L = L + \mathrm{Label}(S)$
  $U = U - S$
Until the number of iterations reaches the threshold
```

3.1 Uncertainty Criteria

The algorithm is based on two uncertainty measures as criteria, Localization Tightness and Localization Stability.

1. Localization Tightness (LT): LT is used to measure the closeness between the detected bounding box and the target, and the smaller the distance between them, the more accurate the positioning. Since the true position of the target object is unknown, the localization tightness of the unlabeled image cannot be calculated by detecting the box and the true position, so we use the change from the Region Proposal to the finalized bounding box to evaluate the compactness of the bounding box. The proximity T of the j^{th} prediction frame is defined as follows:

$$\mathrm{T}\left(B_0^j\right) = \mathrm{IoU}\left(B_0^j, R_0^j\right) \tag{1}$$

where R_0^j is the bounding box of the output of the region suggestion network, which is later adjusted to obtain B_0^j.

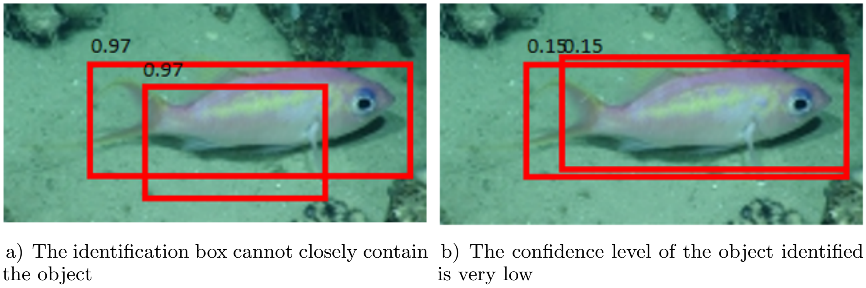
$$\mathrm{J}\left(B_0^j\right) = \left|\mathrm{T}\left(B_0^j\right) + \mathrm{P}_{\max}\left(B_0^j\right) - 1\right|. \tag{2}$$

After calculating the proximity T of all predicted frames, the confidence of the optimal class of predicted frames is introduced, and the calculation of proximity and confidence is done to describe LT and defined as J. The formula is shown below:

There are two extreme cases of LT:

(a) Given a prediction frame, it is if the optimal class confidence is 1 ($P_{max} = 1$), but it cannot wrap tightly a real object ($T = 0$).

(b) On the contrary, the prediction frame can closely wrap a real object ($T = 1$), the confidence of the classification result is very low (uncertain).



a) The identification box cannot closely contain the object b) The confidence level of the object identified is very low

Figure 2.

2. Localization Stability (LS): LS is used to describe the stability of the detection frame localization, which is essentially an expression of whether the detected bounding box is sensitive to the noise of the input image. To evaluate the stability of the localization, we added different levels of Gaussian noise to the image pixel values and calculated how the detected bounding boxes varied with the noise.

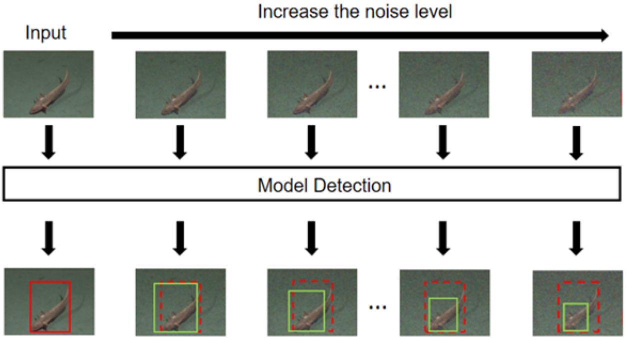


Figure 3. The process of calculating the stability of the prediction frame, where green is the prediction frame and red is the reference frame

First, the current model is used to detect the bounding box in the original image and uses it as a reference frame, the j^{th} reference frame is noted as B_0^j. The corresponding noise samples are generated using Gaussian noise of N energy levels for each sample separately, and then the model is used to obtain prediction frames on the noisy images. Observing the bounding boxes obtained from different noise levels, if the bounding box does not change noticeably throughout the noise level, it can be inferred that the model is stable against noise of this reference frame. Thus, the localization stability of each reference frame can be defined as the average of the IoU of the reference frame and the corresponding frame at all noise levels. Given N levels of noise, the stability of this sample reference frame is expressed by the formula shown in Equation (3):

$$S_B\left(B_0^j\right) = \frac{\sum_{n=1}^{N} \mathrm{IoU}\left(B_0^j, C_n\left(B_0^j\right)\right)}{N} \tag{3}$$

where, B_0^j is the j^{th} reference box, $C_n()$ indicates the highest IoU value among all the detected boxes with overlap.

Based on the former, this localization stability of unlabeled images is defined based on the stability and confidence weighting of each reference frame. The weight of each reference frame is the probability of its highest confidence class to select the frame with high probability but high uncertainty about its position as the foreground object, defined to measure the stability of a single sample, as shown in Equation (4):

$$S_I\left(I_i\right) = \frac{\sum_{j=1}^{M} P_{\max}\left(B_0^j\right) S_B\left(B_o^j\right)}{\sum_{j=1}^{M} P_{\max}\left(B_o^j\right)} \tag{4}$$

where, B_0^j is the j^{th} reference box, M is the number of reference frames.

3.2 Active Learning Initial Sample Selection Method Based on Clustering

In this section, clustering methods will be introduced to active learning for reducing the manual annotation cost. When the sample selection strategy selects a batch of samples from the unlabeled sample set, which may contain a class of similar images, submitting this batch of similar images to manual annotation will increase the manual annotation cost and have limited improvement on the target detection model. Therefore, after sample selection, a clustering algorithm is added to filter the samples with high similarity and provides more representative samples to manual annotation to effectively improve the active learning efficiency.

For a set of unlabeled images, the problem of image clustering lies in dividing the images into different subsets according to their content: clustering two images representing similar objects together and dividing images representing objects with different properties into different subsets. Solutions for image clustering focus on feature selection and processing of complex features. For example, in [25], images are represented using a Gaussian mixture model with fitted pixels and clustering is performed using the information bottleneck (IB) method [26].

In the LACP AL algorithm proposed in this paper, Faster R-CNN [27] is selected as the target detection model. Faster R-CNN is a typical algorithm for two-stage detection, which is an improvement of its predecessor Fast R-CNN and uses Region Proposal Network (RPN) to generate Region of Interest (ROI) [28] based on Fast R-CNN, which replaces the sliding window and selective search strategy with slower computation speed. Before active learning starts, we use a pre-trained feature extraction network of Faster R-CNN to extract features from the data, and the network structure is shown in Figure 4. The CR module is composed of Conv connected ReLU. CRMX module is composed of several CR connection MaxPooling.

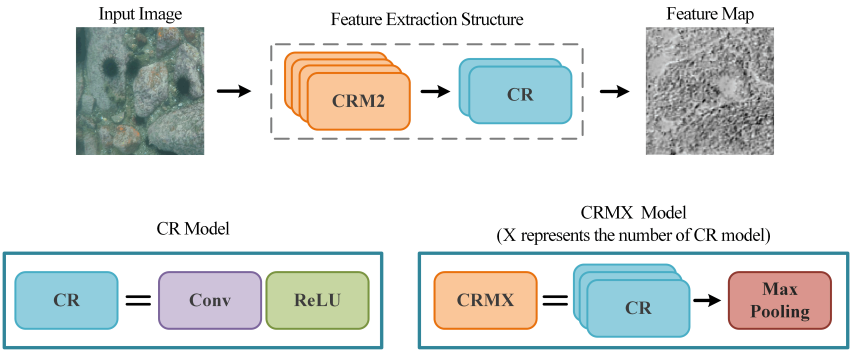


Figure 4. Feature extraction network structure

After feature extraction, the feature map is clustered using the k-means algorithm. K-means algorithm divides the samples into K clusters according to the distance between samples for a given set of samples. The data within the clusters are connected as closely as possible, while making the distance between clusters as large as possible, in the following steps:

1. Randomly initialize k cluster center coordinates;
2. Calculate the distances of all objects in the data set to the k cluster centers, and divide the data points into the nearest clusters;
3. For each cluster, recalculate the center of mass of the cluster as the average of the coordinates of the nodes in the current cluster;
4. Repeat steps 2 and 3 until convergence.

Set the hyperparameters n, and then randomly select n samples from each class for manual labeling. The training set constructed using this method has balanced samples and contains rich information, which helps to improve the recognition accuracy in the initial training of the model.

3.3 Improving Active Learning Noise Sample Label Selection Mechanism

In the metric uncertainty method localization stability, we add different levels of noise to the sample, and subsequently detect the noisy samples, and measure the uncertainty of the sample according to the stability of the noisy sample detection results. In the case of low stability, it contains a situation where the accuracy of some of the noise samples of the sample is high and the accuracy of only a small number of noise samples is very poor, which eventually leads to a low stability of the sample. Such samples with low stability can in fact already be identified by the model and do not need to be selected for manual labeling. To address this situation, we propose a new noise sample selection mechanism.

For each sample, first calculate its average optimal class label confidence generated by adding noise for m samples, as shown in Equation (5):

$$a_i = \frac{1}{m}\sum_{j=1}^{m} p_i^j \tag{5}$$

where m is the number of noise samples corresponding to each sample, p_i^j is the confidence level of the optimal class label in the j^{th} noise sample corresponding to sample i. If $a_i > 0.5$, set parameter b to select top b % of the noisy samples; otherwise, select the latter b % of the noisy samples. For the case where most noisy samples have high confidence, a few extreme noisy samples are removed to improve the stability of the samples; for the case where most noisy samples have low confidence, a few better-performing samples are removed to reduce the stability of the samples, making the sample selection method more inclined to select that type of samples. The time complexity $O(m^2)$ for calculating the stability of the samples of the whole dataset is high, and the time complexity $O(b^2m^2)$ when calculating the selected samples, which greatly saves the computation time.

3.4 New Pseudo-Label Selection Mechanism

The active learning incremental training method has limited effect on model enhancement due to the small number of available samples in the early iterative training phase. To solve this problem, this paper introduces the pseudo-label mechanism into active learning to improve the model performance during supervised learning by means of label-free samples.

We use the model to recognize the unlabeled data and get the recognition results and use the recognition results as pseudo-label for training the model in the next training round. However, in the first iteration of active learning, the recognition accuracy of the model is poor, and most of the pseudo-labels have errors with the real labels, and the use of the pseudo-label mechanism has no positive effect on the training of the model. To solve this problem, this paper proposes a novel pseudo-label selection mechanism: Let the product of the optimal class confidence and stability of the samples be K, set a threshold δ for the pseudo-label mechanism, and set a high threshold δ in the early stage of active learning iteration to select only samples with K greater than δ for assigning pseudo-labels; as the number of model training increases, gradually reduce δ to ensure the pseudo-label accuracy while selecting more samples for training the model.

The specific method is to first select high confidence samples and high stability samples from the sample set, if the product of their optimal class confidence and stability is greater than the threshold δ. Then these samples are assigned pseudo-labels with explicit prediction. Then set the hyperparameter z as the lower limit of the threshold δ, the threshold δ decreases continuously with the number of iterations, and keep the threshold value not changing when the threshold δ decreases to z to ensure that the selected pseudo-labels have high accuracy.

Set the update threshold δ at the end of the t^{th} iteration:

$$\delta = \begin{cases} \delta_0, & t = 0, \\ \delta - d_r * t, & t > 0 \end{cases} \tag{6}$$

where δ_0 is the initial threshold value and d_r is the threshold decay rate. This method enables the pseudolabel selection threshold to decrease with the number of iterations to avoid using too many wrong pseudolabels to affect the model training process.

4 EXPERIMENTS AND RESULTS

The parameters of Faster R-CNN model are shown in Table 1, and the following parameters are set: learning rate is 0.001, momentum is 0.9, learning rate decay coefficient gamma is 0.1, and the number of learning rate decay steps is 100. The dataset uses the public dataset provided by the 2020 China Underwater Robot Professional Contest (CURPC), and the CURPC target categories include "holothurian", "echinus", and "echinoid". The CURPC target categories include "Sea cucumber", "Sea urchin", "Sea star" and "Scallop".

Layer Type	Kernel Size/Stride	Output Size
convolution	$7 \times 7/2$	$112 \times 112 \times 96$
relu and norm		
max pool	$3 \times 3/2$	$56 \times 56 \times 96$
convolution	$5 \times 5/2$	$28 \times 28 \times 256$
relu and norm		
max pool	$3 \times 3/2$	$14 \times 14 \times 256$
$14 \times 14 \times 384$		
relu		
convolution	$3 \times 3/1$	$14 \times 14 \times 384$
relu		
convolution	$3 \times 3/1$	$14 \times 14 \times 256$
relu		

Table 1. Faster R-CNN network structure

4.1 Experimental Environment and Data Set

In this paper, mean Average Precision (mAP), Precision (PR) and Omission Ratio (OR) are used to evaluate the model performance. In addition, to complete the sample density and quantitative counting statistics, both Mean Average Error (MAE) and Root Mean Squared Error (RMSE), which are commonly used in density counting work, are used as evaluation indexes. The formulas of PR, OR, MAE and RMSE are shown in Equation (7).

$$\begin{aligned} PR &= \frac{1}{n}\left(\sum_{i}^{n}\frac{t_i}{t_i+f_i}\right), \\ OR &= \frac{1}{n}\left(\sum_{i}^{n}\frac{m_i}{t_i+f_i+m_i}\right), \\ MAE &= \frac{1}{n}\sum_{i}^{n}|k_i'-t_i|, \\ RMSE &= \sqrt{\frac{1}{n}\sum_{i}^{n}(k_i'-t_i)^2} \end{aligned} \tag{7}$$

where t_i denotes the number of correctly detected targets, f_i denotes the number of incorrectly detected target classes, m_i denotes the number of undetected targets, $\{k_i'\}_{i=1}^{n}$ denotes the number of model-predicted targets in each test image, and i$\in$[1,n],$\{\mathrm{t}_i\}_{i=1}^{n}$ denote the number of target truth values in each image.

4.2 Comparison Experiments

In this paper, experiments were conducted on the CURPC dataset using four methods using LAAL, Discriminative Active Learning, Learning Loss for Active Learning, and LACP AL, respectively, and the experimental termination condition was chosen to be 70 % of the overall data labeled. Table 1 measures the final effect of the experiments by the four metrics.

Method	PR	OR	MAE	RMSE
LAAL	0.799	0.142	13.13	19.67
Discriminative Active Learning	0.7637	0.156	15.57	23.95
Learning Loss for Active Learning	0.786	0.142	13.60	21.83
Ours	0.810	0.126	11.21	14.17

Table 2. Comparison of experimental results of different models

As shown in Table 2, the LACP AL model achieved 81.07 % accuracy, which is 4.76 % better than the Discriminative Active Learning method, 3.05 % lower leakage rate, 1.92 % lower mean absolute error, and 9.78 % lower root mean square error, which means that the LACP AL model can achieve more accurate detection in the more complex. This means that the LACP AL model can achieve more accurate detection of benthic organisms in the more complex seafloor environment. Figure 5 shows the comparison of LAAL and LACP AL identification results.

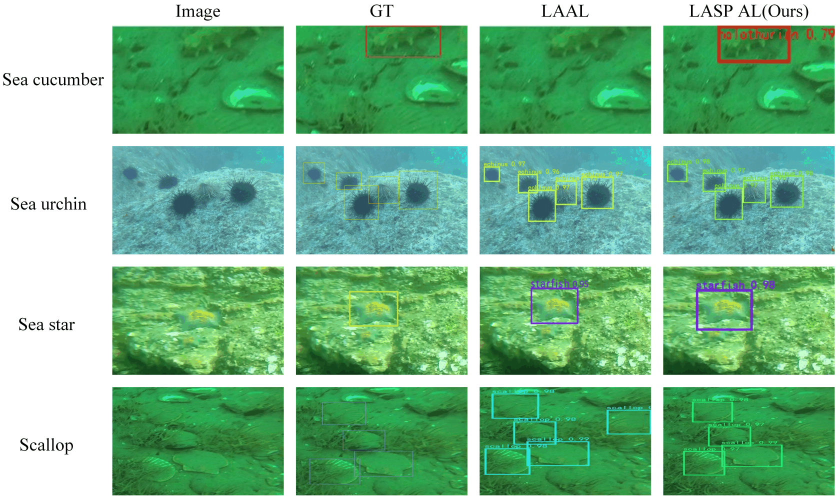


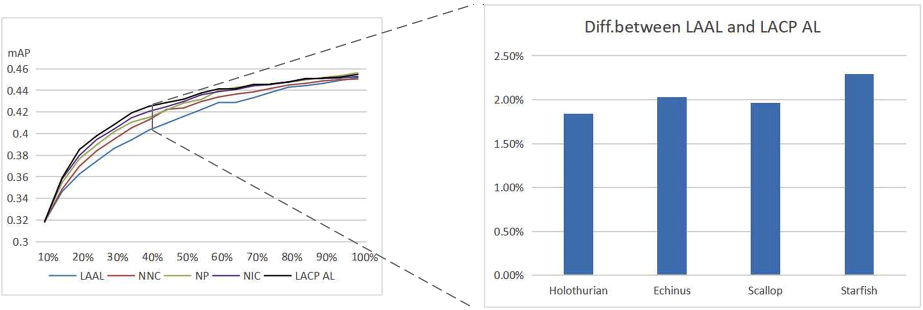
Figure 5. Comparison of the recognition results of LAAL and LACP AL

4.3 Ablation Experiments

To verify the improvement of the model by the three innovations, the following control groups were set up for experiments.

1. Localization-Aware Active Learning (LAAL) pseudo;
2. Sample selection algorithm + noisy sample selection (No Pseudo-label, abbreviated: NP);
3. sample selection algorithm + improved pseudo-label selection mechanism (No Noise Choice, abbreviated: NNC);
4. Noisy sample selection + improved pseudo-label selection mechanism (No Initial Choice, abbreviated: NIC).

The above experiments are used to verify the role of each of the three innovations in the model. First, the maximum number of iterations is set to 200, and the number of iterations is increased by 10 for each model iteration as the number of active learning cycles increases. The variation of model accuracy with the number of labeled data is observed using the CURPC public dataset, as shown in Figure 6.



a)

b)

Figure 6. Results of ablation experiments

As can be seen from the left side of Figure 6, the LACP AL model improves the accuracy by 1.73 %, 1.04 % and 1.17 % compared with the NIC group, NNC group and NP group when the labeled data reaches 40 %, respectively, and the innovations proposed in this paper can effectively improve the training speed of the model and make the model have higher accuracy when training the same number of samples. Among them, improving the active learning noise label selection mechanism plays an important role in improving the model accuracy.

The right side of Figure 6 shows the average accuracy difference between LAAL and LACP AL, when labeling 40 % of the data, on all types. From the figure, LACP AL achieves a 1.84 % improvement in identifying the more difficult sea cucumber class and a 2.29 % improvement in identifying the simpler sea star.

To verify the improvement of the proposed method on recognizing more difficult classes, experiments were designed to observe the improvement of AP for each class and further analyze the performance of each method. Table 3 shows the accuracy of each method on the Zhanjiang competition dataset after three rounds of active learning.

Method	Sea Cucumber	Sea Urchin	Scallop	Sea Star	mAP
LAAL	33.836	36.781	37.293	38.445	36.588
NP	35.165	37.845	38.837	39.546	37.848
NNC	34.043	36.915	37.724	38.851	36.887
NIC	35.360	37.939	39.147	39.812	38.065
LACP AL	35.393	38.415	39.389	39.865	38.266

Table 3. Accuracy in all categories after three rounds of active learning

The experiments showed that the LACP AL model achieved a 1.557 % improvement in identifying the more difficult sea cucumbers, with an average improvement of 1.678 % over the four categories. The method that did not include the improved active learning noise label selection mechanism achieved only 0.21 % improvement on the sea cucumber category, and this difference suggests that the improved active learning noise label selection mechanism can greatly help the model learn the difficult category.

The proposed algorithm improves the accuracy by 2.2 % compared with the LAAL algorithm when labeling 30 % of the data, and the recognition accuracy is equivalent to that achieved by the LAAL algorithm when labeling 45 % of the data, which can effectively save 15 %. The accuracy of the LACP AL algorithm with 50 % data is only 2.29 % lower than that with all data, which shows that this method can be effectively used for data labeling.

5 CONCLUSION

In this paper, an improved active learning algorithm LACP AL for target detection is proposed. Before the start of active learning, LACP AL select representative samples using a clustering algorithm; when selecting unlabeled images for labeling to train a new model, the algorithm combines the classification results and localization results of unlabeled images and adds noise processing for sample selection. Pseudo-labels are added during iteration. Considering that the model cannot accurately detect images at the beginning of the iteration, samples with higher confidence are selected as pseudo-labels in the early stage, and the confidence criterion for selecting pseudo-labels is gradually reduced as the number of iterations increases. Experiments show that by using clustering to select initial samples, screening noisy samples and adding pseudo-labels for intelligent annotation of deep-sea macrobenthos, the active learning algorithm LACP AL proposed in this paper can effectively use fewer resources to achieve higher accuracy.

Acknowledgements

This project is partially supported by the Guanghe Foundation of Dawning Information Industry Co., Ltd. "Deep-sea biological in situ intelligent recognition system and quantitative analysis system development project" (ghfund202107021586).

REFERENCES

[1] Su, H.—Qi, W.—Hu, Y.—Karimi, H. R.—Ferrigno, G.—De Momi, E.: An Incremental Learning Framework for Human-Like Redundancy Optimization of Anthropomorphic Manipulators. IEEE Transactions on Industrial Informatics, Vol. 18, 2022, No. 3, pp. 1864–1872, doi: 10.1109/TII.2020.3036693.

[2] Freund, Y.—Schapire, R. E.: A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. Journal of Computer and System Sciences, Vol. 55, 1997, No. 1, pp. 119–139, doi: 10.1006/jcss.1997.1504.

[3] Girshick, R.—Donahue, J.—Darrell, T.—Malik, J.: Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 580–587, doi: 10.1109/CVPR.2014.81.

[4] Lee, D. H.: Pseudo-Label: The Simple and Efficient Semi-Supervised Learning Method for Deep Neural Networks. International Conference on Machine Learning (ICML 2013), Workshop on Challenges in Representation Learning, Vol. 3, 2013.

[5] Settles, B.: Active Learning Literature Survey. Technical Report 1648, University of Wisconsin-Madison, 2009.

[6] Kapoor, A.—Grauman, K.—Urtasun, R.—Darrell, T.: Active Learning with Gaussian Processes for Object Categorization. 2007 IEEE 11th International Conference on Computer Vision, IEEE, 2007, pp. 1–8, doi: 10.1109/ICCV.2007.4408844.

[7] Freytag, A.—Rodner, E.—Denzler, J.: Selecting Influential Examples: Active Learning with Expected Model Output Changes. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (Eds.): Computer Vision – ECCV 2014. Springer, Cham, Lecture Notes in Computer Science, Vol. 8692, 2014, pp. 562–577, doi: 10.1007/978-3-319-10593-2_37.

[8] Konyushkova, K.—Sznitman, R.—Fua, P.: Introducing Geometry in Active Learning for Image Segmentation. Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), 2015, pp. 2974–2982, doi: 10.1109/ICCV.2015.340.

[9] Jain, S. D.—Grauman, K.: Active Image Segmentation Propagation. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 2864–2873, doi: 10.1109/CVPR.2016.313.

[10] Su, H.—Mariani, A.—Ovur, S. E.—Menciassi, A.—Ferrigno, G.—De Momi, E.: Toward Teaching by Demonstration for Robot-Assisted Minimally Invasive Surgery. IEEE Transactions on Automation Science and Engineering, Vol. 18, 2021, No. 2, pp. 484–494, doi: 10.1109/TASE.2020.3045655.

[11] Li, C.—Fahmy, A.—Li, S.—Sienz, J.: An Enhanced Robot Massage System in Smart Homes Using Force Sensing and a Dynamic Movement Primitive. Frontiers in Neurorobotics, Vol. 14, 2020, Art. No. 30, doi: 10.3389/fnbot.2020.00030.

[12] Hasan, M.—Roy-Chowdhury, A. K.: Continuous Learning of Human Activity Models Using Deep Nets. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (Eds.): Computer Vision – ECCV 2014. Springer, Cham, Lecture Notes in Computer Science, Vol. 8691, 2014, pp. 705–720, doi: 10.1007/978-3-319-10578-9_46.

[13] Hasan, M.—Roy-Chowdhury, A. K.: Context Aware Active Learning of Activity Recognition Models. 2015 IEEE International Conference on Computer Vision (ICCV), 2015, pp. 4543–4551, doi: 10.1109/ICCV.2015.516.

[14] Bietti, A.: Active Learning for Object Detection on Satellite Images. Technical Report, Caltech, 2012.

[15] Sivaraman, S.—Trivedi, M. M.: Active Learning for On-Road Vehicle Detection: A Comparative Study. Machine Vision and Applications, Vol. 25, 2014, No. 3, pp. 599–611, doi: 10.1007/s00138-011-0388-y.

[16] Wang, D.—Shang, Y.: A New Active Labeling Method for Deep Learning. 2014 International Joint Conference on Neural Networks (IJCNN), IEEE, 2014, pp. 112–119, doi: 10.1109/IJCNN.2014.6889457.

[17] Roth, D.—Small, K.: Margin-Based Active Learning for Structured Output Spaces. In: Fürnkranz, J., Scheffer, T., Spiliopoulou, M. (Eds.): Machine Learning: ECML 2006. Springer, Berlin, Heidelberg, Lecture Notes in Computer Science, Vol. 4212, 2006, pp. 413–424, doi: 10.1007/11871842_40.

[18] Brust, C. A.—Käding, C.—Denzler, J.: Active Learning for Deep Object Detection. 2018, doi: 10.48550/arXiv.1809.09875.

[19] Roy, S.—Unmesh, A.—Namboodiri, V. P.: Deep Active Learning for Object Detection. 29th British Machine Vision Conference (BMVC 2018), 2018, `http://bmvc2018.org/contents/papers/0287.pdf`.

[20] Haussmann, E.—Fenzi, M.—Chitta, K.—Ivanecky, J.—Xu, H.—Roy, D.—Mittel, A.—Koumchatzky, N.—Farabet, C.—Alvarez, J. M.: Scalable Active Learning for Object Detection. 2020 IEEE Intelligent Vehicles Symposium (IV), IEEE, 2020, pp. 1430–1435, doi: 10.1109/IV47402.2020.9304793.

[21] Kao, C. C.—Lee, T. Y.—Sen, P.—Liu, M. Y.: Localization-Aware Active Learning for Object Detection. In: Jawahar, C., Li, H., Mori, G., Schindler, K. (Eds.): Computer Vision – ACCV 2018. Springer, Cham, Lecture Notes in Computer Science, Vol. 11366, 2018, pp. 506–522, doi: 10.1007/978-3-030-20876-9_32.

[22] Desai, S. V.—Chandra, A. L.—Guo, W.—Ninomiya, S.—Balasubramanian, V. N.: An Adaptive Supervision Framework for Active Learning in Object Detection. 30th British Machine Vision Conference (BMVC 2019), 2019, doi: 10.48550/arXiv.1908.02454.

[23] Iyer, R.—Khargoankar, N.—Bilmes, J.—Asanani, H.: Submodular Combinatorial Information Measures with Applications in Machine Learning. Proceedings of the 32nd International Conference on Algorithmic Learning Theory, Proceedings of Machine Learning Research (PMLR), Vol. 132, 2021, pp. 722–754.

[24] Kaushal, V.—Kothawade, S.—Ramakrishnan, G.—Bilmes, J.—Iyer, R.:

PRISM: A Unified Framework of Parameterized Submodular Information Measures for Targeted Data Subset Selection and Summarization. 2021, doi: 10.48550/arXiv.2103.00128.

[25] Goldberger, J.—Gordon, S.—Greenspan, H.: Unsupervised Image-Set Clustering Using an Information Theoretic Framework. IEEE Transactions on Image Processing, Vol. 15, 2006, No. 2, pp. 449–458, doi: 10.1109/TIP.2005.860593.

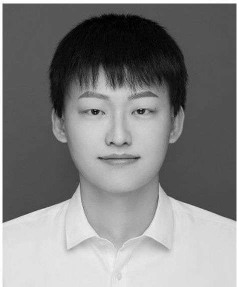
[26] Tishby, N.—Pereira, F. C.—Bialek, W.: The Information Bottleneck Method. 2000, doi: 10.48550/arXiv.physics/0004057.

[27] Ren, S.—He, K.—Girshick, R.—Sun, J.: Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. Advances in Neural Information Processing Systems 28 (NIPS 2015), 2015, pp. 91–99.

[28] Dai, J.—He, K.—Sun, J.: Instance-Aware Semantic Segmentation via Multi-Task Network Cascades. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 3150–3158, doi: 10.1109/CVPR.2016.343.

Qihang Wu is currently pursuing his Master's degree in the School of Information Science and Technology at Qingdao University of Science and Technology, China, with research interests including active learning and object detection.

Yong Liu graduated from the Ocean University of China in 2011, majoring in computer application technology. She is dedicated to the research of intelligent identification and quantitative analysis of marine organisms, medical knowledge graph and intelligent medical big data. So far, she has obtained one national invention patent, seven software copyrights, and published more than 20 research papers (SCI/EI). She has published two textbooks.



Jianyi Zhang is currently pursuing his Master's degree at the School of Information Science and Technology, Qingdao University of Science and Technology. His research interests include object detection, medical image segmentation, and natural language processing.

Yongpan Wang is currently pursuing her Master's degree at the School of Information Science and Technology, Qingdao University of Science and Technology. Her research interests include information extraction and knowledge graphs.