Underwater Sea Cucumber Target Detection Based on Edge-Enhanced Scaling YOLOv4

Ziting Zhang[#], Hang Zhang[#], Yue Wang, Tonghai Liu, Yuxiang He[✉], Yunchen Tian[✉]

Abstract: Sea cucumber detection is widely recognized as the key to automatic culture. The underwater light environment is complex and easily obscured by mud, sand, reefs, and other underwater organisms. To date, research on sea cucumber detection has mostly concentrated on the distinction between prospective objects and the background. However, the key to proper distinction is the effective extraction of sea cucumber feature information. In this study, the edge-enhanced scaling You Only Look Once-v4 (YOLOv4) (ESYv4) was proposed for sea cucumber detection. By emphasizing the target features in a way that reduced the impact of different hues and brightness values underwater on the misjudgment of sea cucumbers, a bidirectional cascade network (BDCN) was used to extract the overall edge greyscale image in the image and add up the original RGB image as the detected input. Meanwhile, the YOLOv4 model for backbone detection is scaled, and the number of parameters is reduced to 48% of the original number of parameters. Validation results of 783 images indicated that the detection precision of positive sea cucumber samples reached 0.941. This improvement reflects that the algorithm is more effective to improve the edge feature information of the target. It thus contributes to the automatic multi-objective detection of underwater sea cucumbers.

Keywords: sea cucumber; edge extraction; feature enhancement; edge-enhanced scaling You Only Look Once-v4 (YOLOv4) (ESYv4); model scaling

1 Introduction

In recent years, China's sea cucumber farming industry has rapidly developed with increasing consumer demand [1]. Most of the fishing methods for sea cucumber culture use manual diving, and the low seafloor temperature and high seawater pressure lead to high labor intensity in underwater operations [2]. It also poses a serious threat to the personal safety of fishing personnel. Automatic harvesting of sea cucumbers by underwater robots can not only solve the safety problems of harvesting personnel but also alleviate the marine environmental problems brought about by the harvesting process. One of the key technologies is the automatic real-time recognition of sea cucumbers. Then, the computer is taught to gain an understanding of the content of the image. Presently, object recognition has served in the recognition of faces [3], license plates [4], fish and underwater creatures [5, 6], ships and military objects [7, 8], and other areas.

The representation of image features plays an important role in the image processing. Due to low transparency and visibility underwater, light transmission in the water will produce strong absorption and scattering effects. The

Ziting Zhang, Hang Zhang, Yue Wang, Tonghai Liu, Yuxiang He and Yunchen Tian are with College of Computer and Information Engineering, Tianjin Agricultural University, Tianjin 300380, China.
✉ Corresponding author. Email: heyuxiang@tjau.edu.cn (Y. He), tianyunchen@tjau.edu.cn (Y. Tian)
#Ziting Zhang and Hang Zhang are co-first authors of the article.

image signal is quickly attenuated [7], which results in impaired image information transmission that is susceptible to problems such as insignificant image contrast, much noise, and single-colour channels. Li et al. [9] improved the defogging effect of sea cucumber images by a priori fusion of Retinex and dark channels, which could effectively identify sea cucumbers. Qiao et al. [10] increased the contrast difference between sea cucumber spines and bodies by contrast-limited adaptive histogram equalization (CLAHE) based on the RGB colour space approach. The accuracy of the algorithm reached 0.965. Using pattern recognition and traditional image processing techniques, the input image can be enhanced to improve the sea cucumber recognition effect. The image processing is complicated, and it needs rich prior knowledge of experts to design the feature extraction process. In traditional feature extraction-based methods, the target sea cucumbers are identified for specific conditions. Complex and variable marine environments describe distinct features, so detection requires higher robustness of algorithms.

The ability of deep learning to process big data can satisfy the urgent requirements of fast and accurate analysis of marine big data and can solve a series of marine problems [11]. Examples of object classification include marine data reconstruction (convolutional neural network, CNN) [12], and marine data classification and recognition (CNN) [13]. Guo et al. [14] compared processed sea cucumber images with unprocessed sea cucumber images to achieve recognition based on a deep residual network (ResNet) and the results showed that the accuracy was 0.972 after image processing. [15] proposed the Shortcut Feature Pyramid Network (SFPN) method, which modified the existing multiscale feature fusion strategy using shortcut connections with an average precision (mAP) of 0.915. CNN started with the LeNet-5 [16], which contained five layers, and developed to ResNet152 [17] with a depth of up to 152 layers. With the deepening of the network structure, redundant feature layers relative to the target are generated, which affects the

model learning efficiency. To effectively extract the image features, some scholars proposed feature detection methods developed with deep learning, such as edge detection. The commonly used edge feature detection methods included three main types: traditional edge operators, learning-based methods, and deep learning [18]. For example, Canny detector [19] represents the brightness step edges by gradient and detects the edges after nonextreme suppression and double thresholding. However, Canny detectors appear to approach unbounded regions in natural images due to high-contrast texture features. As a result, Martin et al. [20] combined brightness, colour, and texture features to obtain a natural boundary. In learning-based methods, custom integrated learning and training models are usually used after manual feature selection. Dollár and Zitnick [21] proposed a structured forest method applied in edge detection. With the application of CNNs in edge detection, deep learning-based models such as DeepEdge [22], holistically nested edge detection (HED) [23], richer convolutional features (RCF) [24], and bidirectional cascade network (BDCN) [18] have been gradually proposed. Representative features better highlight the characteristics of the object. However, detection of distinctive features with multiple targets can be unfair to some objects. On the basis of learning general image features, the enhancement of target features is helpful to improve the image effect.

In recent years, deep learning has shown great performance in the field of object detection, which has improved the marine organism perception ability. Liu et al. [25] used combined nearest neighbour upsampling and deconvolution to improve the precision from 0.932 to 0.963. However, the two-stage detector takes more time than the one-stage detector. Xia et al. [26] redesigned the network structure based on You Only Look Once v2 (YOLOv2) to rapidly detect sea cucumbers underwater. The You Only Look Once (YOLO) series achieves multi-target detection with higher detection speed than the two-

stage target detection algorithm.

In this study, the speed advantage of You Only Look Once v4 (YOLOv4) [27] was used to further enhance the speed to quickly detect sea cucumbers. Meanwhile, the edge features are enhanced in feature extraction for the problems that occur during underwater sea cucumber harvesting, such as muddy sand, complex light environment, and obscuration. Through the edge-enhanced scaling YOLOv4 (edge-enhanced scaling YOLOv4, ESYv4) method, the detailed features of sea cucumbers were highlighted while obtaining the general features of the image.

2 Materials and Methods

2.1 Materials

In this work, a collection of videos in a natural environment was selected as the dataset, which was captured by the C-Watch in a sea cucumber farm at Haiyang in Shandong Province, China. The C-Watch dived into approximately 1.5 m underwater to shoot images of sea cucumbers from the top view and transmit the images to a laptop for storage.

The dataset contains various illumination conditions, background, overlap or occluded cucumber. To reflect the multiscale fusion of the model, the image resolution consists of three sizes: 320×288, 1 280×720, and 1 280×1 080. Some sample images are shown in Fig. 1. The set selected 2 612 images from the video clip and was divided into a training set and a validation set at a ratio of 7:3.

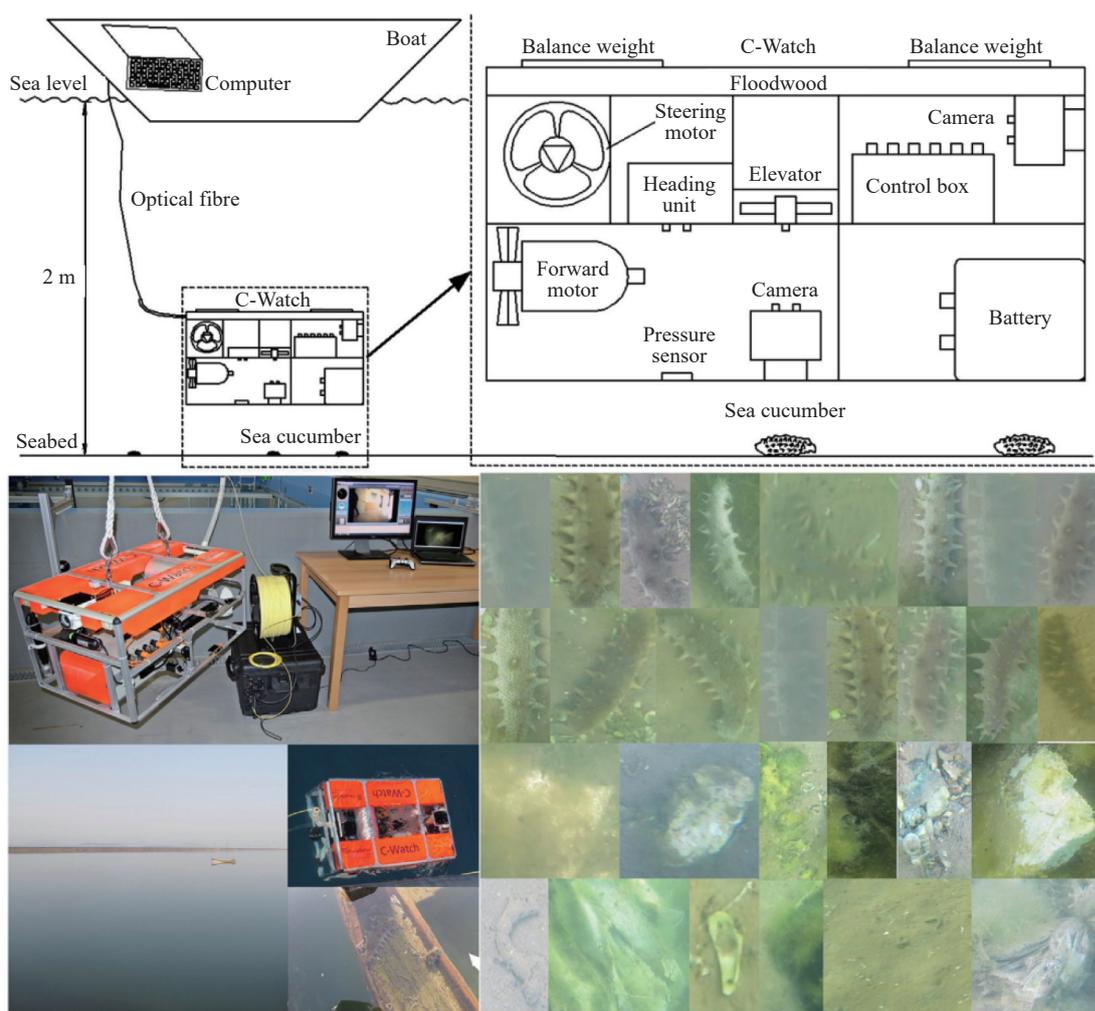This study was implemented on the

Fig. 1　Underwater acquisition equipment and data display [27]

PyTorch-3.8 framework. The cloud computing platform was configured as an Intel Core (Trade-Mark, TM) i7-9750 h Central Processing Unit (CPU) @2.60 GHz, with a 2.59 GHz processor, 16 GB of operating memory, and a 16 GB NVIDIA P100 graphics processing unit (GPU); the system is Ubuntu 18.04.

2.2 Overall Technical Method

To detect sea cucumbers quickly and accurately, an edge-enhanced scaling YOLOv4 (ESYv4) is proposed in this paper, and the overall technical roadmap is shown in Fig. 2. First, this paper uses BDCN as the main network of the feature enhancement module to extract the edge fea-

tures in the image. Then, the scaling YOLOv4 is applied as the backbone network for sea cucumber detection for feature extraction and prediction. Due to the unstructured factors in the living environment of sea cucumbers, there are differences between shallow features and deep features. To make full use of the feature differences among different layers, a bottom-up feature fusion is used to retain the influence of shallow features, especially edge features on the result prediction, as shown in Fig. 3.
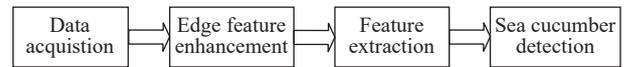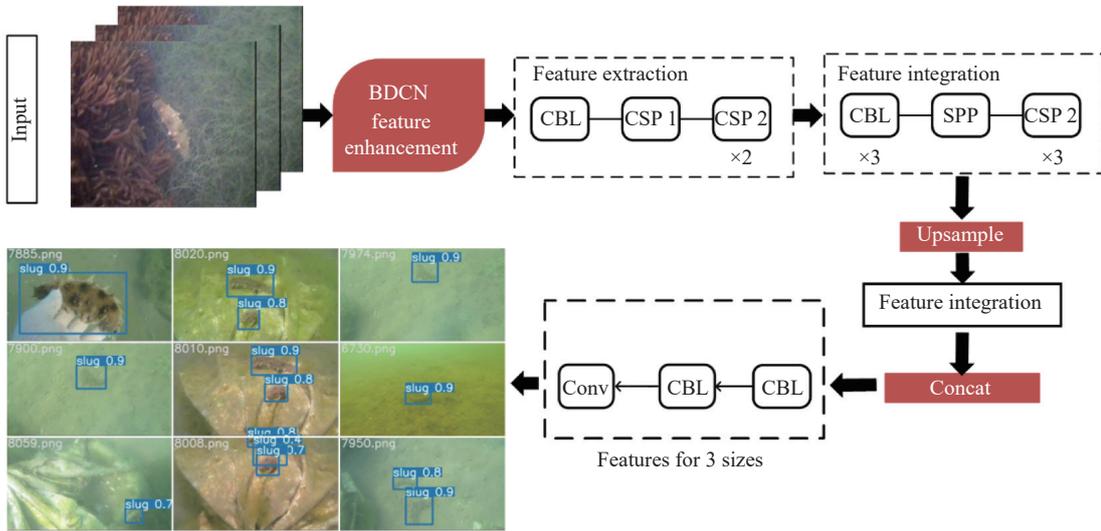

Fig. 2   Overall technical method

Notes: Convolution, batch normalization and ReLU (CBL); convolution (Conv)

Fig. 3   The model structures

2.3 Feature Enhancement by the BDCN Block

Image information is reflected in the feature extraction, which is performed by texture features, statistical features, and colour features. Recently, BDCN [18] was proposed to extract edges at different scales, which introduced a scale enhancement module (SEM) to generate multi-scale features instead of deeper CNNs. Through deep learning, an image-to-image feature extraction approach can be implemented to restore the original features in image data in a data-driven manner and minimize human intervention. Simultaneously, multi-scale vision tasks can be mined for deep features and combined with shal-

low feature information to be conveyed to the object detector, so that the object detector can make full use of the features to detect the target category. To enrich the description of the detailed features of sea cucumbers, the extraction of contour features is improved.

The pretrained BDCN model was inputted to the overall image to obtain the contour information, which enhances the contour features of sea cucumbers. In this study, the object-level boundaries and contour details are of more importance. A CNN-based edge detection model was used to extract the contours of the objects (target and nontarget objects) in the image, and

after a series of convolutions, the features extracted by YOLOv4 degradation were merged, which enhanced the fine-grained recognition of target and nontarget objects using the edge feature enhancement. To enrich the description of the detailed features of sea cucumbers, the extraction of contour features is enhanced.

In this study, the BDCN model based on the Berkeley Segmentation DataSet and Benchmarks 500 (BSDS 500) [28] is used as the edge feature enhancement module. The architecture is based on Visual Geometry Group (VGG)-16 [29] without its three fully connected layers and last pooling layer. The identity (ID) blocks accompany the SEM to form the basic part of the network, where each ID block outputs two sets of edge predictions, and the structure is shown in Fig. 4. SEM was also introduced to achieve multi-scale information extraction, and a func-

tion was proposed by Chen et al. [30], as shown in

$$Y_{i,j} = \sum_{m,n}^{h,w} X\left[(i+rm),(j+rn)\right] W(m,n) \quad （1）$$

where $Y_{i,j}$ is the output after processing by convolution, m, n and h, w represent the height and width of the input image, X is the matrix block to be processed, i, j describes the matrix position of the processing, r is the dilation rate of a dilation convolution, W is a convolutional filter of size m, n. Standard convolution can be treated as a special case with $r = 1$. Different numbers of layers at which the dilation convolution is performed correspond to different dilation convolutions with different dilation rates.spond to different dilation convolutions with different dilation rates.
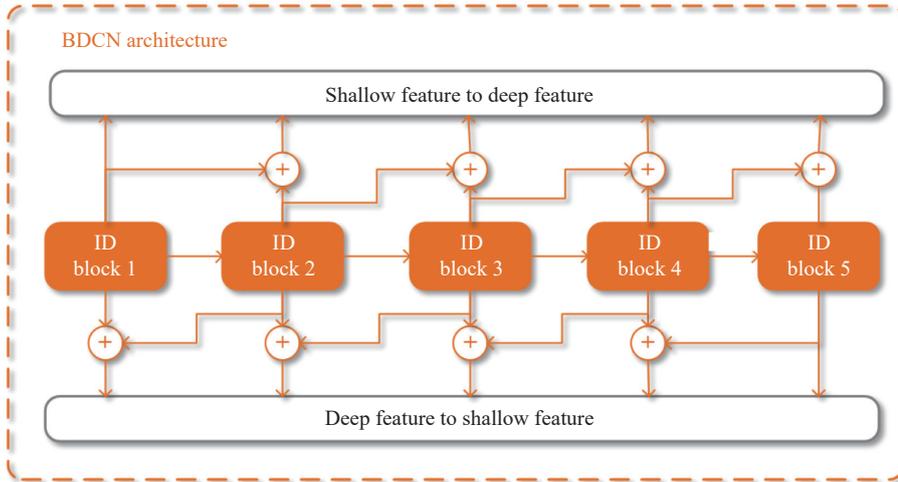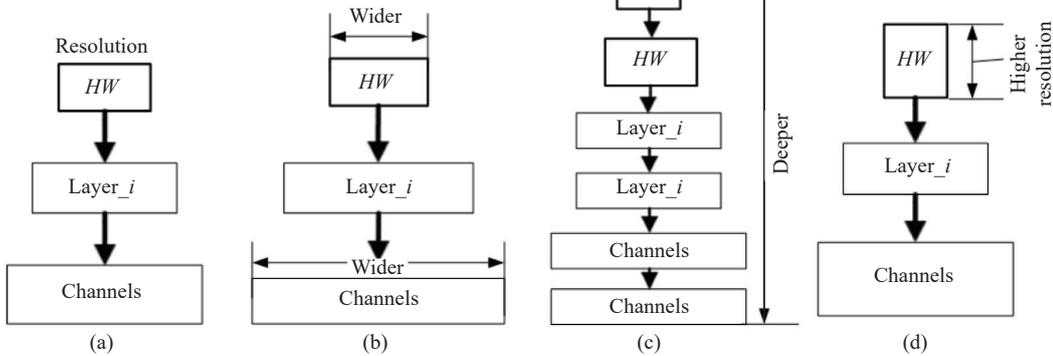

Fig. 4　The BDCN structure

2.4　Sea Cucumber Detection Model Based on Scaling YOLOv4

An object detection model contains a backbone part that extracts image features from datasets to pretrain the model [31– 36] and a neck part that integrates the features [37– 39]. The model also contains a head part which is used to predict the target category and object borders [40, 41]. In this study, the CBL module is used in the cross stage partial-connections (CSP) module,

which extracts features in the backbone. Due to its cross-stage part, CSPNet [42] has been applied to various CNN architectures and has sufficient versatility. Especially in feature extraction, the cross-stage connection solves the problem of duplicate gradient information appearing in the network optimization process, which enables the variability of the gradients to be respected in the integrated feature map stage. Therefore, the CSP module is one of the main

reasons to speed up the inference in this model.

In the neck part, the spatial pyramid pooling (SPP) module is used to increase the receptive field of the backbone features. He et al. [43] improved the SPP module to the concatenation of max-pooling outputs with kernel size $k \times k$, where $k = \{1, 5, 9, 13\}$, and stride is equal to 1. The issue of substantial differences in feature map changes before and after feature fusion is avoided, which optimizes the case of missed detection of sea cucumbers at image edges and false detection of other underwater objects.

2.5 Model Scaling Based on Edge-Enhancement Scaling YOLOv4

Scaling often appears in previous studies, as shown in Fig. 5, which contains three dimensions: depth, width, and image size. The most common model scaling method is to change the depth of a model, i.e. to add more convolutional layers [17, 29]. However, both network depth and width [33, 34, 44] are important for model expressive power, and it remains an open question of how to effectively scale a model to achieve better efficiency and accuracy.

Note: *H*, *W* represent the size of the resolution.

Fig. 5　Model scaling: (a) baseline; (b) width scaling; (c) deeper scaling; (d) resolution scaling

To detect sea cucumber on the device side, the original YOLOv4 had been partially changed. The structure of CSP1, CSP2, CSP8, CSP8, and CSP4 (1-2-8-8-4) in the original backbone network is changed to CSP1, CSP2, CSP2 (1-2-2). Scaling is performed by reducing the number of layers to reduce the computational effort caused by redundant features. In addition to pruning redundant features, the convergence speed of the model was increased. One CSP module was introduced in the neck part of YOLOv4 to replace the part where five of the CBL modules are connected. As shown in Fig. 6, reducing the amount of computation by connecting across stages leads to faster inference.

2.6 Evaluation Indicators

In this study, sea cucumbers were considered positive samples, and others were considered negative samples. The detection network was evalu-

ated by four indices of precision and recall. The P_r represents the probability that the predicted sea cucumber in the actual image is a sea cucumber. The R_c indicates the probability that the sea cucumber is detected in the actual sample.

$$P_r = \frac{T_P}{T_P + F_P} \qquad (2)$$

$$R_c = \frac{T_P}{T_P + F_N} \qquad (3)$$

where $(T_P + F_P)$ is the number of predicted positive samples, and $(T_P + F_N)$ is the actual number of positive samples.

The edge probability map obtained from edge detection is not a matrix of binary values. The values in the matrix are the probabilities of being an edge point in the image. Threshold η must be set for the subsequent evaluation of the results. To assess the fairness, the average precision (AP), optimal dataset scale (ODS), and
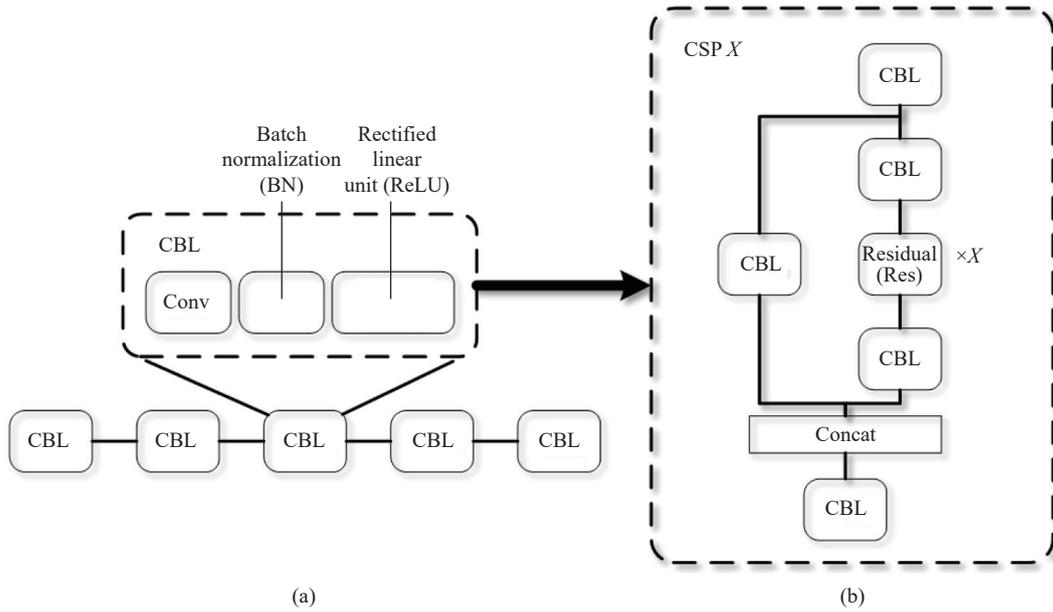
Fig. 6   The replacement of CSP module in the original network: (a) five of the CBL in the original network; (b) the CSP module after replacement

optimal image scale (OIS) were used to evaluate the edge extraction results. The ODS is the threshold that corresponds to the maximum F_{m} for the entire dataset. The OIS is the threshold value that corresponds to the maximum F_{m} of a single image. The F_{m} formula is shown as

$$F_{\mathrm{m}} = \frac{2P_r R_c}{P_r + R_c} \qquad (4)$$

3   Results and Discussion

3.1   Comparison Experiments of Different Edge Detection Modules

Aiming at edge feature enhancement, three of the more commonly used edge detection models, HED, RCF and BDCN, were compared. The results are presented on our sea cucumber dataset after an experimental comparison using the publicly available dataset BSDS 500. The experimental results are shown in Tab. 1.

Tab. 1   Comparison of edge detection experiments

Method	ODS	OIS	AP
Human	0.803	0.803	–
HED	0.788	0.808	0.840
RCF	0.798	0.815	–
BDCN	0.806	0.826	0.847
BDCN multi-scale	0.828	0.844	0.890

In the comparative evaluation of ODS, both single-scale 0.806 and multi-scale detection 0.828 of BDCN were superior to HED and RCF. Additionally, the ODS was 0.025 higher than that of human observations.

The BDCN model also performs well in multi-scale dataset. This characteristic BDCN is better adapted than other methods to the complex underwater light environment with variable hues. BDCN outperforms HED and RCF by 0.04 and 0.034 respectively, and the comparison of experimental results in Tab. 1 and Tab. 2 shows that the BDCN approach is more stable in the face of different tasks.

Tab. 2   Comparison experiments on multi-scale dataset

Method	ODS	OIS	AP
Human	0.750	–	–
HED	0.851	0.864	–
RCF	0.857	0.862	–
BDCN	0.891	0.898	0.935
BDCN multi-scale	0.894	0.901	0.941

Moreover, if the model requires compression, the number of ID blocks can be reduced from 5 to 2. The number of parameters can be reduced from 8.69 MB to 0.28 MB, after which it per-

forms better than HED and RCF. The results are shown in Tab. 3.

Tab. 3   Experimental comparison of different numbers of ID blocks

ID block	HED	RCF	BDCN
1	0.595	0.595	0.727
2	0.697	0.71	0.762
3	0.75	0.766	0.771
4	0.748	0.761	0.802
5	0.637	0.758	0.815
Fuse	0.79	0.805	0.82

After being pretrained on the BSDS 500, the edge extraction result is visualized in Fig. 7. The greyscale image is the edge extracted from the ordinary image which is on the right. Although the sea cucumber and background colour are relatively similar, the edge details of the sea cucumber are still clearly extracted. Simultaneously, the influence of background is gradually weakened during the extraction of contours by BDCN. In Fig. 7, the folds and algae that appeared in the background were partially extracted, but distinctly from the contours of the sea cucumber, which indicates that the BDCN can enhance edge features.

(a)   (b)

Fig. 7   Results of sea cucumber contour extraction: (a) the sea cucumber edge extraction with complex background; (b) the sea cucumber edge extraction with simple background

3.2   Scaling the YOLOv4 Network Model

When training the sea cucumber dataset with the original YOLOv4 model, the inference time of the validation set does not meet our expectation. Therefore, model scaling is deemed to improve the detection time to migrate to mobile devices.

First, in the original network, the activation function in the convolution, batch normalization and mish (CBM) module, which is Mish, introduces nonlinear factors for the model. However, in this experiment, its characteristics are not completely exploited by the influence of computational resources and the dataset. Instead, it prolongs the inference time. Consequently, all the CBMs were replaced with CBLs to accelerate the model convergence. The original CSP in the backbone network uses a 1-2-8-8-4 structure, where the number indicates the number of residual blocks. This dataset is influenced by the image quality, such as low or high overall image brightness and blurred object details in the image, so there is limited feature information. The redundant residual block structure increases the computational burden and makes similar image information easily repeatedly learned which results in overfitting. Therefore, the backbone structure was adjusted from 1-2-8-8-4 to 1-2-2. The number of parameters was reduced by 0.48, while the improved network had 0.008 higher precision than the original network and only half of the original inference time. The experimental results are shown in Tab. 4. The inference speed of the obtained scaling YOLOv4 was improved by half, with no concomitant decrease in precision after refinement for few deep layers. This indicates that the shallow feature information is more critical, while the decrease in recall indicates that the feature discrepancy between foreground targets and the backgrounds needs further improvement. Therefore, enhancement of shallow edge features was chosen to facilitate the model to learn the feature differences between foreground targets and backgrounds. The precision and recall of the ESYv4 with the introduction of the edge enhancement module were 0.941 and 0.936 respectively, with an improvement of 0.037 and 0.028 over YOLOv4.

Tab. 4　Experimental results comparing the number of parameters with the inference time

Configuration	Parameter (MB)	Inference time (ms)	Precision	Recall
YOLOv4	245.52	18.5	0.904	0.908
Scaling YOLOv4	127.57	9.6	0.912	0.874
ESYv4	190.1	19.0	0.941	0.936

3.3　Comparative Analysis of Feature Enhancement Modules

In this study, the BDCN was introduced for the edge feature enhancement. Three different edge detection structures are also compared as feature enhancement modules: RCF, HED and BDCN. The precision of different edge feature enhancements is 0.908 for HED, 0.912 for RCF, and 0.941 for BDCN, respectively. The enhancement of the results shows the effectiveness of the edge feature enhancement module. The introduction of the BDCN model improves the precision by 0.029 and 0.033 compared to HED and RCF, respectively. Thus, BDCN more sufficiently expresses the edge information of sea cucumbers and better enhances the features. In the HED and the RCF structures, it is the output of the edge prediction information from the intermediate layers that enable multi-scale detection, i.e. the lower layers by acquiring local information and the higher layers by acquiring object-level information from a larger domain of interest. However, this process ignores the fact that a particular intermediate layer can only predict edge information at a particular scale. In contrast, the SEM introduced in the BDCN structure fuses multi-layer information features while expanding the perceptual domain of the current information layer. The fusion of shallow edge information with deep edge information for multi-scale sea reference targets presented in the dataset facilitates the differentiation of feature differences between foreground targets and backgrounds in complex environments. The robustness of the model is thus enhanced during the multi-target detection of sea cucumbers.

3.4　Analysis of Misjudgment of Sea Cucumber Detection

Objective extraction of characteristics is the key principle of sea cucumber detection in the natural environment in this experiment. Therefore, the feature selection and annotation process are as objective as possible. After the image is imported, no preprocessing is conducted, such as colour correction and foreground extraction, which inevitably affect the detection results. The detection results are shown in Fig. 8, when there is a similarity between the background profile feature and the sea cucumber profile feature. A false detection issue occurs, such as a red box callout in an image with the ordinal number "7748.png". Images are influenced by murky water bodies and light environments. Algae and sea cucumbers attached to the reef create a similarity in features, which leads to misidentification, as indicated by the red marker in the image label "6274.png". Although there is a misidentification problem, the confidence level of the detection box is not high. Choosing a more appropriate confidence level can avoid false detection problems.

In the case of IoU=0.5, the confidence levels of 0.1–0.9 were compared. The result is shown in Fig. 9. As can be seen from Fig. 9, precision and mAP are positively correlated with confidence, while recall decreases when confidence increases. Thus, the confidence boost can filter out results which does not satisfy the criteria and effectively improve accuracy. Especially when the confidence level=0.6, the difference between the recall and the previous one exceeds 0.10 for the first time, but the precision increases by 0.02, which also proves that confidence is effective for improving the precision.

4　Conclusion

Accurate and fast identification of sea cucumbers is crucial to automate the management and collection of underwater sea cucumbers. In this

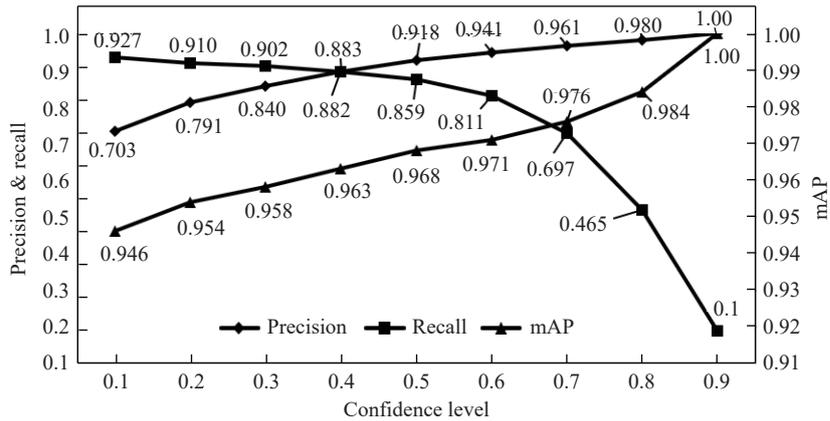Fig. 8   Result of the sea cucumber detection

Fig. 9   Precision, recall, and mAP of different confidences of the sea cucumber detection

paper, model fusion is used to achieve feature enhancement to improve the accuracy. The feature enhancement module can realize model scaling, which can reduce the number of fusion layers from 5 to 2. The number of parameters will be reduced from 8.69 MB to 0.28 MB to reduce the computation. For scaling network the CSP module is used to replace the convolutional blocks in the neck part. The number of parameters is reduced to 0.48 of the original, the inference time is reduced to half, and the accuracy rate is improved from 0.904 to 0.912. The ESYv4 model proposed in this study combined edge feature enhancement and model scaling with a precision of 0.941, which helps achieve automatic underwater sea cucumber acquisition.

Meanwhile, the complex underwater optical environment, mud, reefs, underwater organisms, and other background problems affect the method of other feature enhancement, and the stability of the algorithm in complex environments must be improved. The method is less flexible in terms of target detection, despite the addition of an edge enhancement module and model visibility. The lightweight and high-precision model is more suitable for mobile equipment, which is conducive to improving the utilization of underwater equipment. Future improvements will be made to the flexibility of the model and the speed of detection, as well as to the application

of the proposed method in practical systems and to the needs of sea cucumber farming management.

References：

[1] X. Qiao, J. Bao, H. Zhang, L. Zeng, and D. Li, "Underwater image quality enhancement of sea cucumbers based on improved histogram equalization and wavelet transform," *Information Processing in Agriculture*, vol. 4, no. 3, pp. 206-213, 2017.

[2] G. U. O. Xiangyun, H. U. Min, W. Wensheng, and L. I. U. Yahui, "Sea cucumber automatic identification based on deep learningin unstructured environment", *Journal of Beijing Information Science & Technology University*, vol.34, no.3, pp. 1-5, 2019.

[3] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song, "SphereFace: Deep hypersphere embedding for face recognition," in*2017 IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Honolulu, HI, USA, pp. 6738-6746, 2017.

[4] C. N. E. Anagnostopoulos, I. E. Anagnostopoulos, I. D. Psoroulas, V. Loumos, and E. Kayafas, "License plate recognition from still images and video sequences: A survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 9, no. 3, pp. 377-391, 2008.

[5] L. Jin and H. Liang, "Deep learning for underwater image recognition in small sample size situations," *OCEANS 2017 – Aberdeen*, vol. 2017, no. 61379007, pp. 1-4, 2017.

[6] L. Meng, T. Hirayama, and S. Oyanagi, "Underwater-drone with panoramic camera for automatic fish recognition based on deep learning," *IEEE Access*, vol. 6, pp. 17880-17886, 2018.

[7] W. Yang, Y. Xu, X. Qiao, W. Rao, D. Li, and Z. Li, "Method for image intensification of underwater sea cucumber based on contrast-limited adaptive histogram equalization," *Transactions of the Chinese Society of Agricultural Engineering*, vol. 32, no. 6, pp. 197-203, 2016.

[8] M. M. A. Zabidi, J. Mustapa, M. M. Mokji, M. N. Marsono, and A. Z. Sha'ameri, "Embedded vision systems for ship recognition," in *IEEE Region 10 Annual International Conference, Proceedings/ TENCON*, Singapore, pp. 1-5, 2009.

[9] Z. Li, G. Li, B. Niu, and F. Peng, "Sea cucumber image dehazing method by fusion of Retinex and dark channel," *IFAC–PapersOnLine*, vol. 51, no. 17, pp. 796-801, 2018.

[10] X. Qiao, J. Bao, L. Zeng, J. Zou, and D. Li, "An automatic active contour method for sea cucumber segmentation in natural underwater environments," *Comput Electron Agric*, vol. 135, no. C, pp. 134-142, 2017.

[11] N. Wang, Y. Wang, and M. J. Er, "Review on deep learning techniques for marine object recognition: Architectures and algorithms," *Control Eng Pract*, vol. 118, pp. 104458, 2020.

[12] A. Ducournau and R. Fablet, "Deep learning for ocean remote sensing: An application of convolutional neural networks for super-resolution on satellite-derived SST data," in *2016 9th IAPR Workshop on Pattern Recogniton in Remote Sensing* (*PRRS*), Cancun, Mexico, pp. 1-6, 2016.

[13] Z. Duo, W. Wang, and H. Wang, "Oceanic mesoscale eddy detection method based on deep learning," *Remote Sens* (*Basel*), vol. 11, no. 16, pp. 1921, 2019.

[14] X. Y. Guo, W. S. Wang, Y. H. Liu, and H. Zhang, "Automatic recognition of underwater sea cucumber based on resnet," *Mathematics in Practice and Theory*, vol. 49, no. 13, pp. 203-212, 2019.

[15] F. Peng, Z. Miao, F. Li, and Z. Li, "S-FPN: A shortcut feature pyramid network for sea cucumber detection in underwater images," *Expert Systems with Application*, vol. 182, pp. 115306.1-115306.13, 2021.

[16] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278-2324, 1998.

[17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, pp. 770-778, 2016.

[18] J. He, S. Zhang, M. Yang, Y. Shan, and T. Huang, "Bi-directional cascade network for perceptual edge detection," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Long Beach, CA, USA, pp. 3823-3832, 2019.

[19] J. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern*

Analysis and Machine Intelligence, vol. 8, no. 6, pp. 679-698, 1986.

[20] D. R. Martin, C. C. Fowlkes, and J. Malik, "Learning to detect natural image boundaries using local brightness, color, and texture cues, " *IEEE Computer Society*, vol. 26, no. 5, pp. 530-549, 2004.

[21] B. Nair, "Fast edge detection using structured forests, " *International Journal of Emerging Trends & Technology in Computer Science*, vol. 37, no. 8, pp. 1558-1570, 2015.

[22] G. Bertasius, J. Shi, and L. Torresani, "DeepEdge: A multi-scale bifurcated deep network for top-down contour detection, " in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, pp. 4380-4389, 2015.

[23] S. Xie and Z. Tu, "Holistically-nested edge detection, " in *Proceedings of the IEEE International Conference on Computer Vision*, Santiago, Chile, pp. 1395-1403, 2015.

[24] Y. Liu, M. M. Cheng, X. Hu, K. Wang, and X. Bai, "Richer convolutional features for edge detection, " *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 8, pp. 1939-1946, 2019.

[25] Y. Liu and S. Wang, "A quantitative detection algorithm based on improved faster R-CNN for marine benthos, " *Ecological Informatics*, vol. 61, pp. 101228, 2021.

[26] C. Xia, L. Fu, H. Liu, and L. Chen, "In situ sea cucumber detection based on deep learning approach, " in *2018 OCEANS – MTS/IEEE Kobe Techno-Oceans, OCEANS - Kobe 2018*, Kobe, Japan, pp. 1-4, 2018.

[27] X. Qiao, J. Bao, H. Zhang, F. Wan, and D. Li, "FvUnderwater sea cucumber identification based on principal component analysis and support vector machine, " *Measurement (Lond)*, vol. 133, pp. 444-455, 2019.

[28] P. Arbeláez, M. Maire, C. Fowlkes, and J. Malik, "Contour detection and hierarchical image segmentation, " *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 5, pp. 898-916, 2011.

[29] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition, " in *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, San Diego, CA, USA, pp. 1-14, 2015.

[30] C. Chen, W. Zhu, Y. Guo, C. Ma, W. Huang, and C. Ruan, "A kinetic energy model based on machine vision for recognition of aggressive behaviours among group-housed pigs, " *Livestock Science*, vol. 218, no. 2018, pp. 70-78, 2018.

[31] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks, " *Communications of the ACM*, vol. 60, no. 6, pp. 84-90, 2017.

[32] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks, " in *Proceedings – 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, Honolulu, HI, USA, pp. 5987-5995, 2017.

[33] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications, " *ArXiv*, arXiv.1704.04861, 2017. Available: http://arxiv.org/abs/1704.04861

[34] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks, " in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, pp. 4510-4520, 2018.

[35] X. Zhang, X. Zhou, M. Lin, and J. Sun, "ShuffleNet: An extremely efficient convolutional neural network for mobile devices, " in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, pp. 6848-6856, 2018.

[36] N. Ma, X. Zhang, H. T. Zheng, and J. Sun, "Shufflenet V2: Practical guidelines for efficient CNN architecture design, " *Computer Vision – ECCV 2018, Lecture Notes in Computer Science*, vol. 11218 LNCS, pp. 122-138, 2018.

[37] T. Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature Pyramid networks for object detection, " in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, pp. 1500-1504, 2019.

[38] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path Aggregation Network for Instance Segmentation, " in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, pp. 8759-

8768, 2018.

[39] M. Tan, R. Pang, and Q. V. Le, "EfficientDet: Scalable and efficient object detection," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, pp. 10778-10787, 2020.

[40] M. Tan, R. Pang, and Q. V. Le, "EfficientDet: Scalable and efficient object detection," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Seattle, WA, USA, pp. 10778-10787, 2020.

[41] T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," in *2017 IEEE International Conference on Computer Vision* (*ICCV*), Venice, Italy, pp. 318-327, 2020.

[42] C. Y. Wang, A. Bochkovskiy, and, H. Y. M. Liao, "CSPNet: A new backbone that can enhance learning capability of CNN," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* (*CVPRW*), Seattle, WA, USA, pp. 1571-1580, 2020.

[43] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial Pyramid pooling in deep convolutional networks for visual recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1904-1916, 2015.

[44] S. Zagoruyko and N. Komodakis, "Wide Residual Networks," in *British Machine Vision Conference 2016, BMVC 2016*, York, UK, pp. 87.1-87.12, 2016.

Ziting Zhang is pursuing the Master degree in electronic information at Tianjin Agricultural University, Tianjin, China. Her current research interests include computer vision and deep learning.

Hang Zhang received the Ph.D. degree in agricultural information technology from China Agricultural University, Beijing, China, in 2016. He is a Lecturer with the College of Computer and Information, Tianjin Agricultural University, Tianjin, China. His research interests include image processing and pattern recognition.

Yue Wang is pursuing the Master degree in electronic information at Tianjin Agricultural University, Tianjin, China. Her current research interests include computer vision and image processing.

Tonghai Liu received the Ph.D. degree in agricultural engineering from China Agricultural University, Beijing, China, in 2014. He is a Professor at the College of Computer and Information, Tianjin Agricultural University, Tianjin, China. His current research interests include computer vision and image processing.

Yuxiang He is a Professor at the College of Computer and Information, Tianjin Agricultural University, Tianjin, China. Her research interests include application of information technology in fine regulation and control of agricultural production environment.

Yunchen Tian is a Professor at the College of Computer and Information, Tianjin Agricultural University, Tianjin, China. His research area includes computer application and agricultural internet of things technology.