A REAL-TIME YANGTZE FINLESS PORPOISE TARGETS DETECTION METHOD BASED ON IMPROVED YOLO V5

Guangjun WANG[1], Hao ZHAO[2], Bowen NIE[3], Ziyao ZHAO[4], Houying HE[5], Honglai HAN[6], Kui LIU[7], Xiaohui CHEN[8,*]

Yangtze Finless Porpoise (YFP) automatic recognition is an important basis for sample statistics, biological habit studies, and real-time alerting of YFP. This study proposes an improved Yolo v5 targets detection method, which realizes a highly accurate and real-time automatic recognition of YFP in the monitoring area. First, we constructed the image dataset by collecting many YFP images and performed image pre-processing. Then, we propose an improved the Yolo v5 network architecture, for the recognition, where the merits include: (1) Combine the Spatial Pyramid Pooling (SPP) module and the Spatial-Channel Sequeeze & Excitation (seSE) attention mechanism to construct the SPP-scSE module, which performs better than the SPP module used in the original Yolo v5s; (2) Insert the scSE module into the optimal position in the proposed backbone network; (3) Optimize the combination and fusion mode of the feature maps input by the medium-sized target detection layer. The results of the experiments show that the improved Yolo v5 could recognize YFP effectively in real-time. The recall rate, accuracy rate, mAP and F1 are 93.9%, 93.6%, 95.5% and 91.9%, respectively. The average recognition time for each image is 0.013 seconds. Therefore, it is concluded that the improved algorithm can well solve the real-time accurate detection of YFP in video surveillance systems and offer technology supports to the protection and research of YFP.

Keywords: automatic recognition; Yangtze Finless Porpoise; improved Yolo v5; attentional mechanism; convolutional neural network

[1] The University Key Laboratory of Intelligent Perception and Computing of Anhui Province, Anqing Normal University, Anqing 246013, China; University of Science and Technology of China, Hefei 230026, China/P. R. China

[2] The University Key Laboratory of Intelligent Perception and Computing of Anhui Province, Anqing Normal University, Anqing 246013, China

[3] The University Key Laboratory of Intelligent Perception and Computing of Anhui Province, Anqing Normal University, Anqing 246013, China

[4] The University Key Laboratory of Intelligent Perception and Computing of Anhui Province, Anqing Normal University, Anqing 246013, China

[5] The University Key Laboratory of Intelligent Perception and Computing of Anhui Province, Anqing Normal University, Anqing 246013, China

[6] The University Key Laboratory of Intelligent Perception and Computing of Anhui Province, Anqing Normal University, Anqing 246013, China

[7] The University Key Laboratory of Intelligent Perception and Computing of Anhui Province, Anqing Normal University, Anqing 246013, China

[8] The University Key Laboratory of Intelligent Perception and Computing of Anhui Province, Anqing Normal University, Anqing 246013, China; School of Information Technology, Anqing Vocational and Technical College, Anqing 246013, China, e-mai: 420236663@qq.com

1. Introduction

The Yangtze finless porpoise, scientifically known as Neophocaena asiaeorientalis, primarily inhabits the lower and middle sections of the Changjiang River in China [1]. In this article, we refer to YFP. In recent years, because of unreasonable fishing, vessel pollution, industrial pollution, and construction works, the population of YFP has been drastically reduced and listed as a key aquatic mammal for national-level protection, which requires human protection [2]. Analyzing the population size and biological habits of YFP is the basis for conservation, while timely discovery and alerting of YFP is a real need to avoid inadvertent harm [3-4]. The key issue of these tasks is the timely recognition of YFP.

YFP has the characteristic of breathing out of the water, and the breathing interval is usually about one minute. Sometimes, the YFP only exposes its head when it breathes, and sometimes, its whole body leaps out of the water. Traditional YFP scientific expedition mainly relies on visual identification by biologists' eyes, which requires biologists to observe for a long time on the banks or boats of the Yangtze River. This recognition way is labor-intensive and time-intensive work, and it may be prone to errors and omissions in YFP counting due to experience, fatigue, and other factors [5-6]. Therefore, to ensure efficient detection and achieve timely detection and alarm of YFP, further research on image-based automatic recognition technology for YFP is crucial [7-8].

With the evolution of artificial intelligence technology, automatic recognition of target detection technology has gained wide application in many fields [9]. Suk-Ju Hong et al. combined Faster R-CNN, Yolo and other deep learning detection techniques with other algorithms to detect birds in various environments [10]. A Jalal et al. automatic classification and counting of fish using a Yolo-based object detection system to study species-related optical features for fish [11]. DH Jung et al. constructed a cow sound classification model based on deep learning, which enables farmers to determine the state of cows [12].

Due to the Yangtze River's complex environment, factors such as garbage, wind and waves, weather, and birds cause large interference to the target detection of YFP. Meanwhile, the water surface of the Yangtze River is wide, and the YFP may be far from the camera, forming a small target detection problem. On the other hand, the monitoring of YFP requires a large deployment of surveillance cameras and timely recognition alarms in the future. Therefore, a lightweight YFP target detection algorithm needs to address the importance of ensuring the accuracy of YFP detection in real time.

In summary, we finally chose the Yolo v5 algorithm for our study and improve the algorithm for the actual problem. First, we constructed the image dataset by collecting many YFP images and performed image pre-processing. Then, aiming at the problem of YFP identification, the improved Yolo v5 methods are as follows:

(1) Combine the scSE attention mechanism and the SPP module to construct the SPP-scSE module to substitute the SPP module in the former Yolo v5 network backbone architecture; (2) Through experimental testing, insert the scSE module into the optimal position in the proposed backbone network; (3) Optimising the combination and fusion patterns for the input feature maps of the medium-sized object detecting layer. Finally, using experiments to analyze and compare the efficiency of the detection algorithm, explore the best automatic detection method of the YFP.

2. Materials and Methods

2.1 Yangtze finless Porpoise images capture

In this study, the wild Yangtze finless porpoise was the research object. The YFP images were mainly obtained through a self-built video surveillance system, with the original image resolution of 1280 pixels *720 pixels and the JPEG format. For some high-resolution images, we performed patch-based processing on the images. We acquired video images from September 2020 to August 2021 to extract images of the presence of YFP. We also used the internet to collect images of YFP. Finally, we obtained 2000 valid images as the primary image dataset.

Firstly, we created a test set of 200 images and a training set of 1800 images by randomly selecting images from 2000 available images. Subsequently, to improve the efficiency of the training process of the YFP object detection model, we compressed the 1800 images in the training set. After that, we use "LabelImg" to label the YFP targets manually by drawing the outside rectangular box of the YFP targets on the compressed river surface images. Finally, we employed dataset enhancement techniques on the training set. These techniques include adjusting the image brightness, flipping the image vertically or horizontally, and rotating the image in multiple angles (90°, 180°, 270°). The final training set consists of 4,200 images, including 1,800 original and 2,400 enhanced images. It is worth noting that there are no duplicate images between the training and test sets.

2.2 Improving the network structure of Yolo v5

The Yolo v5 network model features faster processing speeds and higher detection accuracy. Because the YFP targets detection model in this study has high requirements for small target detection and speed and considers the merits such as accuracy, efficiency, and size. We designed a YFP targets detection network which improved based on Yolo v5 architecture.

2.2.1 The scSE attention mechanism

The attentional mechanism is a method for the network to automatically learn a set of weighting factors, which emphasizes important regions through "dynamic weighting" and reduces the impact of irrelevant background regions [13-14]. The scSE attention mechanism is an enhancement of the squeeze & excitation (SE) mechanism, created by combining the spatial squeeze and channel excitation block (cSE) module with the channel squeeze and spatial excitation block (sSE) module in parallel [15].

The cSE module [16], operates by compressing the global spatial features of the feature map into channel-specific descriptors. This adaptation allows the network to capture the varying relationships between channels and improve its capability to represent significant channel features. By employing the cSE module, the network model can focus on learning important channel features while disregarding less relevant ones. On the other hand, the sSE module aims to compress the channel features of the feature map and enhance the network's ability to learn crucial spatial features by emphasizing the channel features [17]. Fig. 1 provides a visual representation of the network architecture of the sSE module.

The scSE module effectively integrates the cSE and sSE modules [18]. By extracting the input feature maps' channel and spatial importance degrees and adding them together, the resulting feature maps with higher importance (important channel and spatial features) will be more robust. The network is encouraged to study additional significant feature information. This structure enables the network to pay attention to fine-grained features, thus improving the extraction capability. Fig. 2 illustrates the architecture of the network for the scSE module.

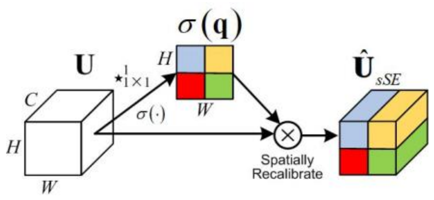


Fig. 1. The architecture of sSE module.

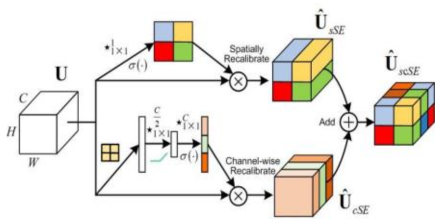
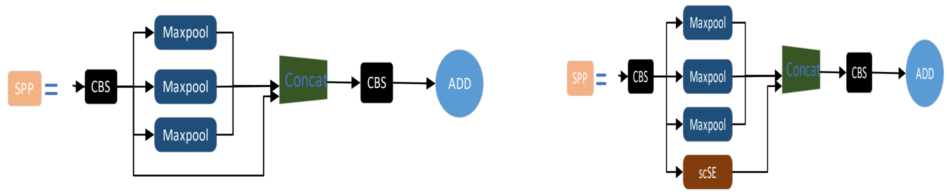


Fig. 2. The architecture of scSE module.

2.2.2 Backbone network improvements

To deal with the problem that the smaller targets of Yangtze finless porpoise cannot be correctly identified due to insufficient extraction of subtle features and effective features, in the backbone network, this paper adds scSE module to the 6th and 8th layers. The scSE module is a lightweight attention module, which will obtain feature maps with higher importance (i.e., both important channel features and important spatial features) for stronger stimulation, prompting the network to get more useful feature information. This allows the network to attend more to subtle features, focusing on learning effective information, enhancing the extraction of effective features and suppressing ineffective information.

The SPP module incorporates three Maxpool layers with different-sized kernels to capture local area receptive fields and near-global field information from feature maps. It leverages these multiple scales of perceptual fields to conduct feature fusion to enhance the expressiveness of the feature map by combining information from different scales. By utilizing the SPP module, the perceptual coverage of the output features from the backbone network is improved. The module enables the extraction of both local and global information, allowing for a more comprehensive understanding of the input data. Moreover, this fusion of different scales helps separate important background information from the foreground features, leading to a more focused and accurate detection performance. To optimize the feature extraction process and achieve the best detection performance, we introduce the scSE module to the shortcut connection of the SPP module. The scSE module enhances the learning and extraction of effective features, thereby improving the feature extraction capabilities of the backbone network. Fig. 3 in the study presents a comparison of diagrams, illustrating the structural improvements made to the SPP module before and after the proposed modifications.



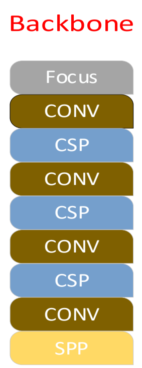
(a) The original SPP structure. **(b)** The improved SPP structure.
Fig. 3. Original SPP structure and Improved SPP structure.

The atrous spatial pyramid pool (ASPP) is a parallel convolutional layer that captures multi-scale information using feature maps through different scaling rates and fuses the outputs together to further improve the perceptual domain of the network [19]. In target detection, a large receptive field has a better ability to extract features and ordinary convolution can increase the receptive field, but the consequence is a decrease in resolution and loss of detailed information. However, the ASPP has solved the above problems by adding the null convolution to the structure, which can obtain a large sensory field without losing too much effective information. Fig. 4a and Fig. 4b show the comparison before and after the backbone network structure improvement. Therefore, we added the ASPP module after the SPP layer of that backbone network to improve the receiving field, reduce the extraction of effective information, and improve the ability to acquire features. Also, it reduces the leakage and misdetection problems of the original network due to insufficient features.

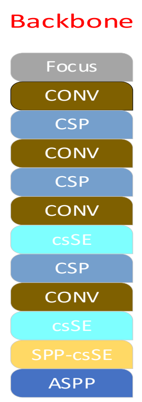
2.2.3 Improved fusion feature layer

Feature fusion is important to improve the effective information extracted from network features in target detection. It is difficult to avoid information loss as the network transfers information across the network from the bottom level to the top level. The bottom level feature graph provides better resolution containing better information regarding a particular object. Nevertheless, the lower level feature maps are less semantic and contain more noise due to fewer features extracted through the convolutional layers. Higher level feature maps contain rich information but lower resolution. Therefore, the merit of feature fusion largely determines the performance of target detection.

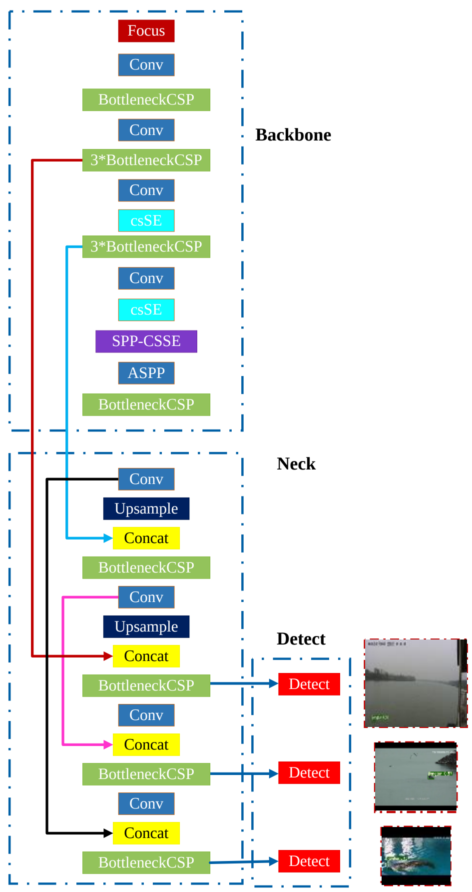
This study has improved the backbone network to add feature extraction layers. Finally, Fig. 4c shows a target detection network model based on the improved Yolo v5s. Effective feature fusion improves the information loss caused by network information transmission to a higher level and improves the network's target detection accuracy. Following experiments, it was shown that the reconstituted hierarchical fusion performed better in terms of detection.



(a) The original architecture of the backbone.



(b) The improved architecture of the backbone.



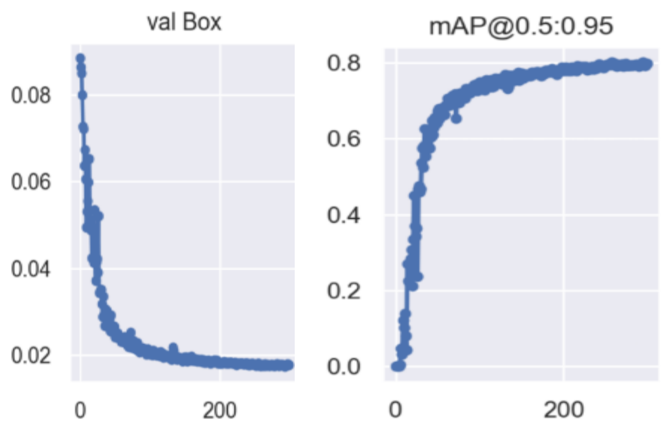
(c) The improved network of the Yolo v5.

Fig. 4. The comparison of the original network and the improved network.

2.3 Network Training

This study was tested in Win10 environment based on pytorch1.7 framework, using python3.8. The hardware conditions are: NVIDIA GeForce RTX 3080 11G GPU, CPU is Intel(R) Xeon(R) CPU E5-2690 v4 dual processor. The experiment was implemented to train and test the Yangtze finless porpoise target detection model of the YFP monitoring platform by writing the program code in Python and calling the required libraries such as CUDA, Cudnn, and OpenCV.

Fig. 5a illustrates the loss curve of the validation set, showcasing a significant reduction in loss value during the initial 100 epochs of network training. As the training progresses, the curve stabilizes after approximately 120 hours. Based on this observation, the output of the model after 120 epochs was chosen as the YFP target identification model for the video surveillance system in this study. Fig. 5b presents the average mean Average Precision (mAP) at different Intersection over Union (IoU) thresholds. These results indicate that the model was trained effectively, as there is no evident overfitting. The mAP values across various IoU thresholds demonstrate the model's ability to accurately identify YFP targets. Furthermore, the mAP results presented in Fig. 5b indicate that the model was successfully trained without overfitting, demonstrating its proficiency in identifying YFP targets.



(a) (b)

Fig. 5. Validation set loss curve and validation set mAP.

In this experiment, four objective indicators of performance, namely precision (P), recall (R), F1 score and mAP, have been used to measure the training target detection model's performance. The mAP metrics are used to evaluate the detection task and are an important measure of the model's overall detection accuracy in single classification target detection. These metrics are represented below:

$$Precision = \frac{TP}{TP+FP} \quad (1) \qquad Recall = \frac{TP}{TP+FN} \quad (2)$$

$$F1 = \frac{2PR}{P+R} \quad (3) \qquad mAP = \frac{1}{C}\sum_{R=i}^{N} P(k)\Delta R(k) \quad (4)$$

In this scenario, several notations are used to represent different metrics and parameters. TP represents the proportion of correctly identified YFP targets, while FP denotes the proportion of background predictions that are incorrectly identified as YFP. FN signifies the proportion of YFP targets that are incorrectly predicted as background. C represents the total number of YFP target categories. P(k) represents the precision for a specific IoU threshold, k, and R(k) signifies the recall for the same IoU threshold. N represents the total number of IoU thresholds considered. The input image size is set at 640 x 640, while the initial learning rate is set to 0.01. The Adam optimizer is utilized during training, with a batch size of 32. The training process consists of a total of 300 rounds. The weight file with the lowest training loss is saved and subsequently used for testing during the experiments.

3. Results and discussion

3.1 Comparison of ASPP and feature fusion layer enhancement

In this study, we performed an ablation experiment to assess the impact of the attention mechanism. The results of this experiment are outlined in Table 1. We introduced the scSENet module into the sixth (C6) and eighth (C8) layers of the original Yolo v5s backbone network. Furthermore, we enhanced the SPP structure by incorporating a shortcut mechanism based on scSENet. To evaluate the effectiveness of these enhancements, we compared the improvement process with the data presented in Table 1.

Table 1.

Ablation experiments of attentional mechanisms.

Model	P	R	F1	mAP
Yolo v5s	90.4	90.2	90.5	91.5
scSE1-Yolo v5s, scSENet added to C6	91.0	90.9	90.9	92.2
scSE2-Yolo v5s, scSENet added at C8	91.4	90.6	91.0	92.4
scSE-Yolo v5s, scSENet added at C6 and C8	91.6	92.2	91.3	93.2
scSE-SPP-Yolo v5s, added scSENet at C6 and C8, and improving SPP	92.4	92.6	91.9	94.4

The results of the ablation experiments demonstrate that incorporating scSENet into both the C6 and C8 levels of the backbone network yields higher detection accuracy compared to adding scSENet to a single location. Furthermore, by integrating scSENet into the shortcut of SPP, the enhanced SPP detection effect is significantly improved.

To enhance the empirical field of the network, the ASPP module is inserted after the SPP layer. This modification reduces the extraction of redundant information and improves ability to extract relevant features, thereby addressing the issue of missed and incorrect detections caused by insufficient features in the original network. Additionally, the feature fusion layer was further optimized to align with the modified network model structure. Based on the experimental results provided in Table 2, comparative tests were conducted to evaluate the effectiveness of these enhancements.

Table 2

Comparative test of ASPP and feature fusion layer improvement

Model	P	R	F1	mAP
YOLO V5s	90.9	90.2	90.5	91.5
scSE-SPP-YOLO V5s	92.4	92.6	92.5	94.4
scSE-SPP-ASPP-YOLO V5s	93.6	93.9	93.7	95.5

By comparing experimental data, the improved network model improved the AP value by 4% compared with the original network. The accuracy recall rate put all had a more obvious improvement, which improved the problem of false detection and low confidence of the original network and improved the problem of poor detection of YFPs with small targets by the network.

3.2 The comparison of various recognition algorithms' detection results

In this study, we conducted a comparison between the improved Yolo v5s network and its original counterparts, namely Yolo v3, v4, and v5s. The objective was to evaluate their respective capabilities in identifying YFP targets using a dataset consisting of 200 test images [20-22]. The performance of each model was assessed based on accuracy, recall, mAP value, and average detection speed. The results of this comprehensive comparison are summarized in Table 3.

Table 3

The performance of different networks compared.

Model	P	R	mAP	FPS
YOLOV3	88.9	89.4	88.3	58
YOLOV4	90.6	90.8	91.3	62
YOLO V5s	90.9	90.2	91.5	85
Improved YOLO V5s	93.6	93.9	95.5	76

As demonstrated in Table 3, it is evident that the original Yolo v5s algorithm has a slightly better detection performance than Yolo v3 and is comparable to Yolo v4. However, the improved Yolo v5s algorithm exhibits a more significant increase in detection speed compared to Yolo v4, making it more suitable for deployment. The mAP value of the enhanced Yolo v5s network reaches 95.5%, which represents a 4% improvement over the original network. This indicates that the improved network can accurately and efficiently detect YFP targets in real-time.

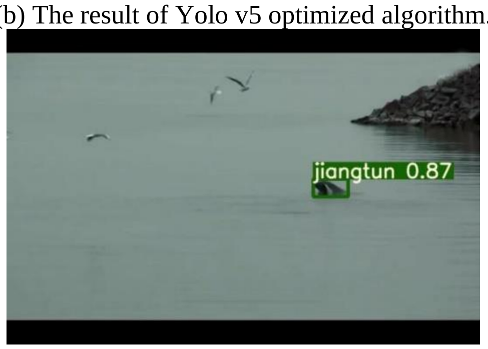
3.3 Comparison of test results

In order to visually demonstrate the capability of the YFP target detection algorithm, we have selected a number of test images for comparison and analysis. Fig. 6 shows the specific detection and comparison results. From Fig. 6a and Fig. 6b, we can see that the original algorithm cannot correctly detect the YFP, which is a small target, and detects a background similar to the target. The optimized algorithm could accurately detect and identify the small target YFP. From Fig. 6c and Fig. 6d it is clear that the original algorithm mistakenly detects the long-distance small target seabird as the YFP. The detection error is caused by the small target and similar shape information. The optimization proposed in this paper can basically solve such problems, and the confidence was high. "jiangtun" in the image is the Chinese pinyin logo of the YFP.

(a)The result of original algorithm.

(b) The result of Yolo v5 optimized algorithm.

(c) The result of original algorithm.



(d) The result of Yolo v5 optimized algorithm.

Fig. 6. The comparison of results of detection algorithms before and after optimization.

4. Conclusions

To enable automatic detection of Yangtze finless porpoises, this study presents an approach for real-time detection of YFP targets using an improved model based on Yolo v5. This method effectively addresses the challenges of misidentification and the identification of small targets encountered in previous algorithms. The proposed model is designed to be lightweight, ensuring its suitability for real-time processing.

The improved network architecture incorporates the scSE attention mechanism module, which enhances the network's focus on subtle characteristics and improves its feature extraction capabilities. Additionally, the backbone network integrates the ASPP structure, which expands the sensory field while mitigating information loss.

The test results demonstrate the effectiveness of our improved model in detecting YFPs in river images. The model achieves a total recall of 93.6%, precision of 93.9%, and a mAP value of 95.5%. With an average detection speed of 0.013 seconds per image, the improved Yolo v5s algorithm outperforms three other algorithms when detecting YFP targets in a test set comprising 200 images. Specifically, the modified Yolo v5s model demonstrates a mAP improvement of 4%, 7.2%, and 4.2% compared to the original Yolo v3, v4, and v5s models. The average detection speed of the improved model is 1.31 and 1.41 times higher than that of the Yolo v3 and v4 networks, respectively, meeting the real-time YFP recognition requirements. Furthermore, the proposed YFP detection algorithm will be applied to a video surveillance system for real-world YFP recognition, statistics, and analysis.

Acknowledgement

Fund Project: This research was funded by the Key Project of Natural Science Research in Higher Education Institutions of Anhui Province “Research on Intelligent Monitoring System for Porpoise Based on Internet of Things Technology” (No. KJ2021A1436), Key Project of Natural Science Research in Higher Education Institutions of Anhui Province “Research on Water Environment Monitoring System Based on Internet of Things Technology” (No. 2022AH052604), Hubei Yangtze River Ecological Protection Foundation” Research on intelligent identification technology of Yangtze finless porpoise”, the Anhui Provincial University Leading Talents Team Fund Grant Project (Anhui Education Secret [2019] No. 16).

R E F E R E N C E S

[1] *Amano M.* Finless Porpoises. Encyclopedia of marine mammals (third edition), 2018, 372-375.

[2] *Tang Bin, Tong Yunyun, Tang Wenqiao, et al.* Population size survery of Yangtze finless porpoise in the Dongfengxisha waters of Yangtze River Estuary. Journal of Shanghai Ocean University, **vol. 27**,no.1, 2018, pp.126-132.

[3] *Mei Z, Chen M, Han Y, et al.* Thresholds of population persistence for the Yangtze finless porpoise: implications for conservation managements. Integrative Zoology, **vol. 6**, no. 4, 2021, pp. 538-547.

[4] *Zhou L, Chen X, Duan P X, et al.* Spatial–temporal variations in biosonar activity of Yangtze finless porpoise in the lower reaches of the Yangtze River and its correlation with underwater noise: Are quieter non-shipping branches the remaining shelters. Aquatic Conservation: Marine and Freshwater Ecosystems, **vol. 31**, no. 4, 2021, pp. 964-978.

[5] *Jiang F.* Saving the Yangtze Finless Porpoise. China Today, **vol. 067**,no. 006, 2018, pp. 62-63.
[6] *Yan Q, Hu Y.* Protecting China's finless Yangtze finless porpoise. Science, **vol. 371**, no. 6533, 2021, pp. 1003-1003.
[7] *Song H, Xu F, Zheng B, et al.* An automatic identification algorithm of Yangtze finless porpoise. 2015 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC). IEEE, 2015, 1-5.
[8] *Li M, Zhang Z, Lei L, et al.* Agricultural greenhouses detection in high-resolution satellite images based on convolutional neural networks: comparison of Faster R-CNN, YOLO v3 and SSD. Sensors, **vol. 20**, no. 17, 2020.
[9] *Chen L, Liu Z, Tong L, et al.* Underwater object detection using Invert Multi-Class Adaboost with deep learning. In 2020 International Joint Conference on Neural Networks (IJCNN), 2020, 9207506.
[10] *Hong S J, Han Y, Kim S Y, et al.* Application of deep-learning methods to bird detection using unmanned aerial vehicle imagery. Sensors, **vol. 19**, no. 7, 2019, pp. 1651.
[11] *Jalal A, Salman A, Mian A, et al.* Fish detection and species classification in underwater environments using deep learning with temporal information. Ecological Informatics, **vol. 57**, 2020, pp. 101088.
[12] J*ung D H, Kim N Y, Moon S H, et al.* Deep learning-based cattle vocal classification model and real-time livestock monitoring system with noise filtering. Animals, **vol. 11**, no. 2, 2021, pp. 357.
[13] *Wang F, Hu H, Shen C.* BAM: A Lightweight and Efficient Balanced Attention Mechanism for Single Image Super Resolution. arXiv preprint arXiv:2104.07566, 2021.
[14] *Hernández A, Amigó J M.* Attention mechanisms and their applications to complex systems. Entropy, **vol. 23**, no. 3, 2021, pp. 283.
[15] *Roy A G, Navab N, Wachinger C.* Concurrent spatial and channel 'squeeze & excitation' in fully convolutional networks. International conference on medical image computing and computer-assisted intervention. Springer, Cham, 2018, 421-429.
[16] *Hu J, Shen L, Sun G.* Squeeze-and-excitation networks. Proceedings of the IEEE conference on computer vision and pattern recognition. 2018, (99): 7132-7141.
[17] *Moustafa M S, Sayed S A.* Satellite Imagery Super-Resolution Using Squeeze-and-Excitation-BasedGAN. International Journal of Aeronautical and Space Sciences, **vol. 22**, 2021, pp. 1481–1492.
[18] *Roy A G, Navab N, Wachinger C.* Recalibrating fully convolutional networks with spatial and channel "squeeze and excitation" blocks. IEEE transactions on medical imaging, **vol. 38**, no. 2, 2018, pp. 540-549.
[19] *Lian X, Pang Y, Han J, et al.* Cascaded hierarchical atrous spatial pyramid pooling module for semantic segmentation. Pattern Recognition, **vol. 110**, 2021, pp. 107622.
[20] *Sang J, Wu Z, Guo P, et al.* An improved yolov2 for vehicle detection. Sensors, **vol. 18**, no. 12, 2018, pp. 4272.
[21] *Liu H, Fan K, Ouyang Q, et al.* Real-Time small drones detection based on pruned YOLOv4. Sensors (Basel, Switzerland), **vol. 21**, no. 10. 2021, pp. 3374.
[22] *Zhou J, Jiang P, Zou A, et al.* Ship Target Detection Algorithm Based on Improved Yolo v5. Journal of Marine Science and Engineering, **vol. 9**, no. 8, 2021, pp. 908