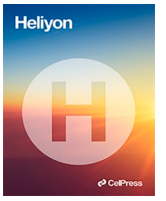
Heliyon 10 (2024) e31029

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

Shallow mud detection algorithm for submarine channels based on improved YOLOv5s

Jiankang Hou, Cunyong Zhang *

School of Marine Technology and Geomatics, Jiangsu Ocean University, Lianyungang, 222005, China

ARTICLE INFO

Keywords:
Submarine channel
Shallow mud
Sub-bottom profiler
EMA
NWD loss
YOLOv5s-EF

ABSTRACT

Submarine mud poses a risk to channel navigation safety. Traditional detection methods lack efficiency and accuracy. As a result, this paper proposed an enhanced shallow submarine mud detection algorithm, leveraging an improved YOLOv5s model to increase accuracy and effectiveness in identifying such hazards in marine channels. Firstly, the sub-bottom profiler was employed to assess the submarine channel of Lianyungang Port to acquire the image data of the shallow mud sound print. Concurrently, the analysis incorporated the characteristics of changes in sound intensity peaks to precisely identify the shallow mud's location. Furthermore, the incorporation of C2F feature module into the backbone module enhances the gradient flow of the algorithm, augments the feature extraction information, and improves the algorithm's detection performance. Subsequently, Efficient Multi-Scale Attention (EMA) mechanism is incorporated into the neck module, aiming to optimize the algorithm's channel dimensions, minimize computational overhead, and enhance its detection efficiency. Finally, the study introduced Normalized Wasserstein Distance (NWD) loss function into bounding box regression loss function. This integration effectively addresses the issue of multi-scale defects, emphasizes the transformation of target planar position deviation, and improves the accuracy of the algorithm's detection capabilities. The results indicate that the improved YOLOv5s-EF algorithm outperforms the original YOLOv5s algorithm and other widely used detection algorithms. It achieved a validation set precision rate of 97.8%, recall rate of 97.6%, F1 value of 97.7%, mean Average Precision (mAP)@0.5 of 98.2%, mAP@0.95 of 69.6%, and Frames Per Second (FPS) of 51.8. YOLOv5s-EF algorithm proposed in this study offers a novel technical approach for detecting mud in submarine channels, which is importance for ensuring the safe operation and maintenance of dredging in such channels.

1. Introduction

In recent years, the safety of submarine channels has become a matter of significant concern due to the ongoing development and utilization of marine resources. Sedimentary accretions of mud, comprising organic materials, minerals, and various particulates, frequently impede navigation within submarine channels [1]. These mud formations result from both natural processes, such as tidal movements and wave action, and anthropogenic influences, including maritime traffic and coastal activities. Such interactions facilitate the gradual deposition of mud on coastlines and seabeds, culminating in significant siltation phenomena [2,3]. The

* Corresponding author.
E-mail address: ouczhangcunyong@163.com (C. Zhang).

https://doi.org/10.1016/j.heliyon.2024.e31029
Received 25 August 2023; Received in revised form 9 May 2024; Accepted 9 May 2024
Available online 10 May 2024

accumulation of mud can result in a reduction of water depth within the submarine channel, thereby constraining the safe navigation of offshore ports and channels [4]. It may also lead to the inability of large ships to enter or depart from the port, thus causing traffic congestion, delays, and higher costs [5]. Furthermore, mud can also pose challenges to ship navigation, thereby elevating the potential for collisions and grounding [6]. Consequently, the investigation of algorithms for detecting mud in submarine channels holds significant practical importance and application value. This research not only has the potential to enhance the safety and dependability of harbor channels, but also to offer crucial insights and a foundation for decision-making in the realms of marine engineering, ship navigation, and channel planning. Addressing these navigational challenges effectively can play a pivotal role in bolstering coastal economies and promoting the sustainable utilization of marine resources.

You Only Look Once (YOLO) detection algorithm has seen extensive application in underwater target detection [7,8], underwater object tracking [9], and various other domains, owing to the rapid advancements in computer vision technology and artificial intelligence. The emergence of this technology enhances the accuracy and efficiency of underwater target detection and tracking, thereby introducing novel avenues for innovation and application across diverse industries. Zhang et al. [10] proposed a method for detecting underwater targets using a combination of MobileNetv2, YOLOv4 algorithm, and attentional feature fusion. This approach achieved both accuracy and speed in target detection within the marine environment. Kim et al. [11] employed a detection algorithm that relies on Darknet-53 and YOLOv3. The algorithm integrated the improved YOLOv3 with sub-bottom profiling to achieve intelligent classification detection of submarine sediments. Li et al. [12] utilized an improved YOLO-SC algorithm for submarine cable detection, demonstrating its effectiveness in accurately locating the submarine cable. Yang et al. [13] conducted a cascade algorithm that is based on UGC-YOLO network architecture. This approach utilized YOLOv3 convolutional neural network (CNN) as the foundational structure, leading to enhanced detection accuracy of underwater targets. Zhang et al. [14] proposed an improved algorithm for detecting underwater targets in YOLOv5. The algorithm incorporated a global attention mechanism and a multi-branch re-parameterization module, resulting in high accuracy in detecting underwater targets. These studies share certain similarities. Each employed YOLO algorithm suite for target detection, developed suitable datasets for algorithm training, and utilized CNNs for image feature extraction. Additionally, they enhancde the original algorithms using a variety of optimization methods. These studies illustrate the potential of employing the improved YOLO algorithm for detecting underwater targets, thereby enhancing the performance and accuracy of such detection and creating new opportunities for innovation and development in various underwater target detection applications.

YOLO detection algorithm is based on deep learning techniques and neural network modeling [15], necessitating a specific quantity of data for its functionality. However, with regard to the identification of shallow mud in submarine channels, the unique and complexity of the seabed environment has resulted in a scarcity of effective detection techniques and a deficiency of pertinent training data. Conventional techniques like precision level [16] and total station electronic rangefinder [17] are employed for fixed-point positioning measurements, offering high accuracy and convenience. However, these methods are primarily suitable for land slopes and are not suitable for submarine channel detection. Optical techniques, such as infrared rangefinders [18] and 3D laser scanners [19], have the capability to detect objects from a distance without physical contact. However, these methods are not appropriate for detecting submarine channels due to the rapid attenuation of light in seawater, resulting in diminished light energy and reduced accuracy. The rapid advancement of acoustic technology has led to the widespread utilization of sub-bottom profilers in the detection of submarine sediment. These profilers are employed in various applications including submarine geological classification [20–22], research on submarine disasters [23–25], and exploration of submarine pipelines [26,27]. This is due to their high accuracy, non-immersibility, real-time data display, portability, ease of operation, strong penetration, and relatively minimal attenuation in liquid [28,29].

In this study, the sub-bottom profiler is utilized to identify the submarine channel on the seabed and to gather shallow mud sound print image data from various locations and depths. The research focuses on the detection algorithm of shallow mud in submarine channels using sound print image data and the improved YOLOv5s-EF. The paper's main contributions are outlined as follows: (1) To overcome the limitation of training data on shallow submarine channel mud, the study utilizes a sub-bottom profiler to collect sound print image data through field probing. It accurately identifies the specific location of shallow mud by analyzing the peaks of acoustic intensity oscillation, thereby establishing the data foundation for the subsequent investigation of the improved algorithm. (2) C2F feature extraction module is introduced to enhance the gradient flow of the algorithm by incorporating additional cross-branching layer links. This enhancement allows the algorithm to obtain more comprehensive feature information while maintaining a lightweight structure. (3) Efficient Multi-Scale Attention (EMA) is proposed to decrease the computational overhead by preserving channel information, restructuring certain channels into batch dimensions, and organizing the channel dimensions into multiple sub-features. This approach ensures that spatial semantic features are uniformly distributed within each feature group. (4) Using Normalized Wasserstein Distance (NWD) loss function and normalized Wasserstein to measure the distribution similarity can improve the target localization accuracy. This approach is effective for measuring small targets, significantly improving the detection performance of the algorithm.

This study seeks to address the issue of identifying shallow mud in submarine channels. Conventional approaches are constrained by the submarine setting and the absence of pertinent training data, thereby impeding the effective detection of submarine mud. Consequently, a sub-bottom profiler is proposed to acquire acoustic image data with the improved YOLOv5s-EF algorithm for the purpose of mud detection. The algorithm offers several advantages and practical applications. Firstly, it performs traditional methods that rely on manual interpretation by detecting targets in images, thus improving detection efficiency. Secondly, unlike conventional methods prone to misjudgment or omission in complex image recognition, the improved YOLOv5s-EF algorithm, featuring automatic extraction and learning functions, has the potential to significantly enhance target identification and detection. Thirdly, the improved YOLOv5s-EF algorithm demonstrates broad applicability in practical scenarios. With additional training and algorithm optimization, it

can be utilized for tasks such as seabed target detection and seabed engineering anomaly detection. The remainder of the paper is structured as follows: Section 2 provides an overview of the field data detection process, YOLOv5s algorithm, and the structural system of the improved YOLOv5s-EF algorithm. Section 3 is dedicated to the evaluation and analysis of the performance of the improved YOLOv5s-EF algorithm, as well as each of the points of improvement. Section 4 discusses the detection effectiveness of the improved YOLOv5s-EF algorithm, along with its limitations and potential future developments. The study's conclusions are presented in Section 5.

2. Materials and methods

2.1. Detection data

The study area focuses on the inner section in the outer channel of Lianyungang Port, Jiangsu Province, within the geographical coordinates of 34°44′–34°45′N latitude and 119°27′–119°32′E longitude (Fig. 1). The navigation survey of the area was conducted using SES-2000 parametric array sub-bottom profiler (Fig. 2). The detection process involved mounting the transducer on the side of the ship and connecting it to the host computer, which was positioned in the secure area of the waterproof cabin. Additionally, the attitude sensor was connected to the interface on the host computer panel, while GPS was affixed to the mounting bar. The arrangement placed the transducer and GPS at a distance of approximately 1.5 m apart. Simultaneously, the instrument parameters were configured according to the specifications (Table 1). Upon activation of the host computer, the signal transmitting unit initiated the transmission of acoustic signals to the seafloor through the transducer for the purpose of seabed geological structure detection. The acoustic signal will propagate at a specific velocity along the seabed and reflect upon encountering interfaces of various sediments. The signal receiving unit is responsible for receiving the reflected acoustic signals and storing them in a computer chip via the control system. Subsequently, the post-processing ISE software in the computer was opened and the data in the chip was imported into ISE. Finally, the acquired acoustic data are processed by ISE to form point, line and block sound print images with certain textures. A total of 150 images were obtained in this study.

The submarine channel environment exhibits specific characteristics and a high level of complexity. The acoustic signals detected and reflected at various depths and times exhibit significant variations, displaying different intensity characteristics and generating sound print images with diverse textures. Assessing the depth of the channel's shallow mud solely with the naked eye can be challenging at times. The sound print image captured by the sub-bottom profiler is derived from the sound intensity signal, which is crucial for interpreting the boundaries of submarine geological layers. The sound intensity varies at each interface of the submarine geological layers, contributing to their differentiation. The shallow layer of the seabed, which serves as the interface between seawater and submarine mud, exhibits a significant acoustic wave impedance effect. This phenomenon influences the sound intensity at the interface. To determine the location of the shallow mud in the submarine channel, the peak sound intensity of the vertical sequence was chosen as the reference point. This decision was informed by the results of previous simulation experiments in the area and field exploration studies [30–32]. Additionally, it involved integrating the sound intensity associated with the sound print image obtained from ISE, the post-processing software of the sub-bottom profiler. Fig. 3 shows the variation of the peak sound intensity of the shallow mud and the corresponding sound print location. As can be seen from the figure, the areas with large variations in peak sound intensity correspond to the locations where the shallow mud is located. The results indicate that the peak sound intensity can provide a more intuitive and efficient reflection of the coarseness and uniformity of the shallow mud in submarine channels.

Fig. 1. Survey region.

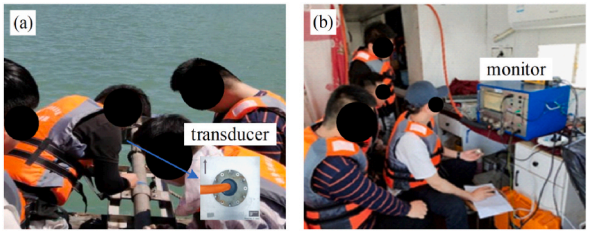


Fig. 2. Field Detection, (a) transducer, (b) monitor.

Table 1
SES-2000 instrument technology and setup parameters.

Item	Content	Main indicators
Performance index	water depth range	1–400 m
	sediment penetration	the deepest 40 m
	attitude compensation	rise and sink
	wave beam angle	4°
	resolution	5 cm
Signal transmitting unit	basic frequency	100 kHz
	low frequency	15 kHz
	pulse width	0.07–1 ms
	pulse type	CW, Ricker
	bandwidth	2–22 kHz
Signal receiver unit	basic frequency	bottom tracking
	low frequency	sub-bottom profile data
	adopt frequency	96 kHz
Hardware construction	host unit size	52 cm × 40 cm × 38 cm
	transducer unit size	30 cm × 26 cm × 7 cm
	control system	internal computer
Power supply	voltage	100–240 V
	power	<500W

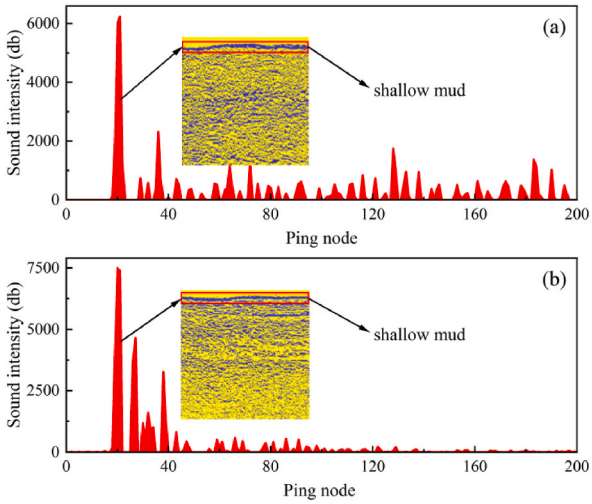


Fig. 3. Shallow mud vertical sequence of sound intensity and sound print.

2.2. YOLOv5s

Fig. 4 illustrates the architecture of YOLOv5s algorithm, which comprises four components: Input, Backbone, Neck, and Head [33, 34]. The first segment represents Input, with the training image input size set at 640 × 640. The second component, Backbone, primarily handles the extraction of features from the input image. The third component, known as Neck, plays a key role in integrating multi-scale features from the feature map and transmitting these features to the detection layer. The fourth component is Head, primarily accountable for the ultimate regression detection.

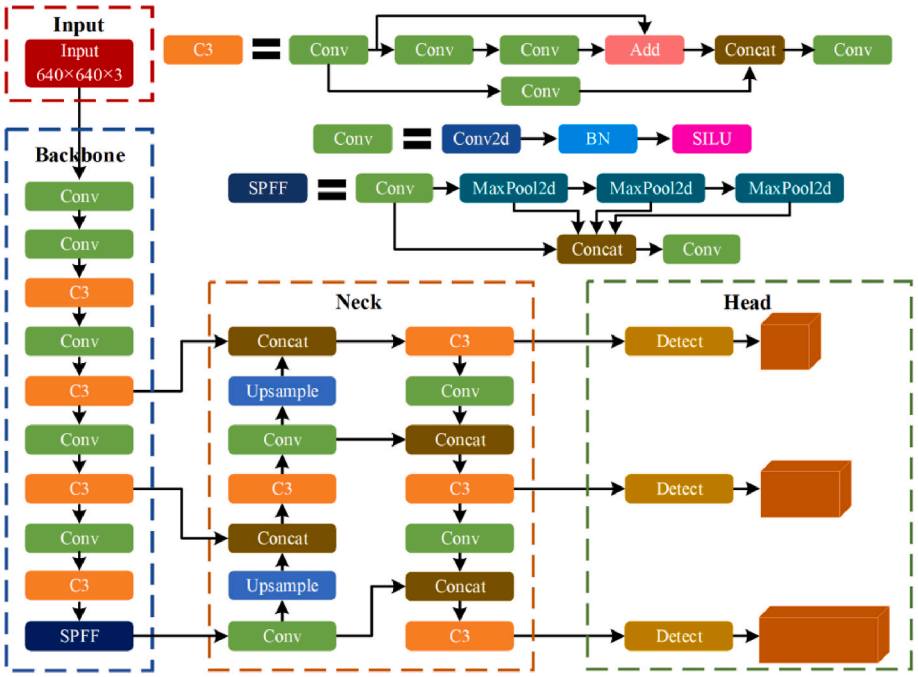


Fig. 4. YOLOv5s algorithm structure.

YOLOv5s algorithm primarily utilizes 1 × 1 and 3 × 3 convolution kernels, and its convolutional structure comprises convolutional, Batch Normalization (BN), and Sigmoid Linear Unit (SILU) layers. Spatial Pyramid Pooling-Fast (SPPF) module utilizes three 5 × 5 small-scale pooling kernels to achieve multi-scale fusion. Additionally, YOLOv5s employs upsampling to adjust the scale of the feature map from low resolution to high resolution, thereby enhancing the algorithm's detection efficiency and generalization capability. Meanwhile, YOLOv5s utilizes adaptive anchor frame computation, allowing for better adaptation to targets with varying scales and aspect ratios through dynamic adjustments to the size and position of the anchor frame for different training samples. In addition, the

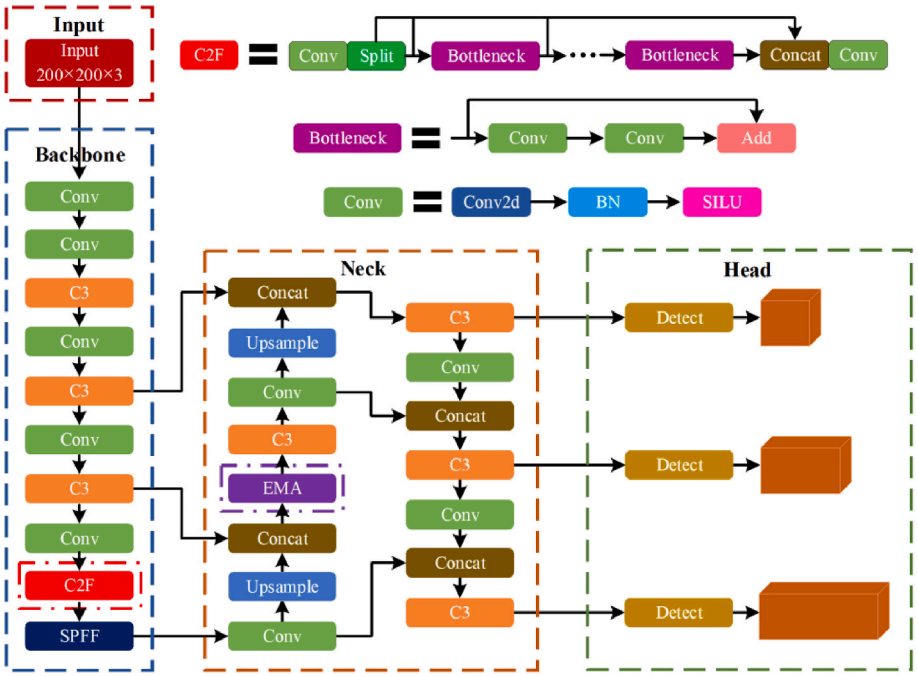


Fig. 5. Improved YOLOv5s-EF algorithm structure.

YOLOv5 algorithm includes data enhancement techniques such as random rotation, random cropping, random scaling, etc., which are added by diversifying the data during the training process to reduce the overfitting problem.

2.3. Improved YOLOv5s-EF

Fig. 5 depicts the structure of the improved YOLOv5s-EF algorithm. In this paper, the algorithm's three components, namely Backbone module, Neck module, and bounding box regression loss function, were primarily enhanced and optimized. C2F module was initially introduced in the backbone module to substitute a portion of C3 module in the original algorithm. This modification extends the algorithm's capability to acquire higher-resolution information, thereby enhancing the detection accuracy and generalization ability of the algorithm. Secondly, Neck module introduces EMA attention mechanism, which enables automatic learning and selective focus on important features of various scales within the image. This effectively enhances the accuracy and robustness of the detection algorithm. Finally, NWD loss function was incorporated into bounding box regression loss function, resulting in a reduction in the bias transformation sensitivity of the weak target position. Consequently, this integration improved the detection performance and efficiency of the algorithm.

2.3.1. Backbone module improvements

Given that C3 module in YOLOv5s has a tendency to capture larger targets and more global contextual information, it might face limitations in effectively capturing subtle features of smaller shallow mud targets within submarine channels. This challenge arises from the module's relatively large receptive field, which may result in the loss of target details and a subsequent decrease in the accuracy of target detection. Therefore, this paper introduces the C2F module to the YOLOv5s algorithm [35,36]. The design of C2F module is derived from C3 module and integrates Efficient Long-Range Attention Network (ELAN) concept. This module has the capability to improve the algorithm for capturing detailed gradient flow information, thus improving the capacity to detect small targets. Furthermore, C2F module integrates fine-grained features from lower levels with semantic knowledge from higher levels via cross-branch connections, enabling the algorithm to possess feature representation capabilities encompassing high resolution and abundant semantic information. The comparison of C3 and C2F modules reveals that C2F module incorporates a greater number of skip layer connections in relation to DenseNet (Figs. 4 and 5). It eliminates the convolution operation in the branch and introduces an additional Split operation, thereby enhancing the richness of feature information and leading to a further reduction in the number of parameters and computations, which can somewhat reduce the overfitting problem.

2.3.2. Neck module improvements

The limited coverage of the shallow mud within the input submarine channel, occupying only a small portion, coupled with the predominance of background information, leads to the convolution process iterative accumulating background information. This accumulation results in the generation of a significant amount of redundant data, which might overshadow certain targets, thereby leading to diminished detection accuracy. Consequently, this paper proposes EMA multiscale attention [37,38]. The structure of EMA (Fig. 6) comprises three parallel paths designed to extract image feature information. One of the pathways involves a 3×3 branch that captures local cross-channel interactions by utilizing a 3×3 convolution to extend the feature space. The remaining two pathways consist of 1×1 branches that combine two channel-level attention mappings within each group through a basic multiplication operation, to generate different cross-channel interaction features between the two parallel routes. Therefore, EMA structure can be utilized to improved pixel-level attention during convolution operations without diminishing the channel dimensions. This is achieved by maintaining the batch dimensions, preserving information on each channel, and ensuring uniform distribution of spatial semantic features within each feature group.

Furthermore, EMA structure utilizes a cross-spatial learning strategy with the goal of encoding global information and capturing remote dependencies to generate multi-scale feature representations for more comprehensive feature aggregation. This approach effectively facilitates the extraction and preservation of crucial feature information. In Fig. 6, the cross-space learning strategy employs two outputs: a 3×3 branch output and a 1×1 branch output. The global spatial information is encoded through a 2D global mean pool (Avg Pool), which converts the joint activation mechanism in the channel features into the corresponding dimensional shapes, thereby enhancing computational efficiency. Simultaneously, Softmax function is employed to apply a linear transformation to Avg Pool in the branch. The output feature maps within each branch are subsequently calculated as a pair of spatial attention weight values. The

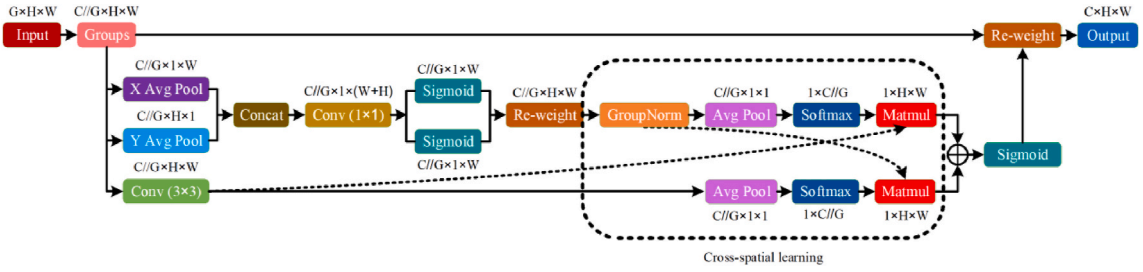


Fig. 6. EMA network structure.

calculation for 2D global pooling operation is presented in Eq. (1), where H, W, and C represent the height, width, and dimensions of the input feature maps, respectively.

$$z_c = \frac{1}{H \times W} \sum_{j}^{H} \sum_{i}^{W} x_c(i,j) \tag{1}$$

2.3.3. Boundary regression loss function improvements

CIOU loss utilized in YOLOv5s focuses exclusively on the overlap area, center point distance, and aspect ratio of the boundary regression. However, it disregards the actual disparity between the width and height, their respective confidences, and the differential sensitivity to smaller targets. As a result, this study introduced NWD loss fucntion for bounding box regression [39–41]. NWD loss model represents bounding box as 2D Gaussian distributions, with the similarity between these distributions assessed using NWD. The benefit of using Wasserstein metric is its ability to gauge distribution similarity effectively, even in scenarios where there is minimal or no overlap between the distributions. Furthermore, for the purpose of providing a more detailed representation of the pixel weights within bounding box, NWD loss function characterizes bounding box as a 2D Gaussian distribution. In this model, the central pixel of bounding box carries the greatest weight, with the significance of the pixel diminishing towards the periphery. Specifically, for the horizontal bounding box $R=(c_x, c_y, w, h)$, (c_x, c_y), w, and h represent the center coordinates, width, and height, respectively. The calculation of the internal ellipse is shown in Eq. (2), where (μ_x, μ_y) is the center coordinate of the ellipse while σ_x and σ_y are the length of the half axis along the x axis and the y axis. In this case, $\mu_x = c_x$, $\mu_y = c_y$, $\sigma_x = w/2$, and $\sigma_y = h/2$.

$$\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{\left(y-\mu_y\right)^2}{\sigma_y^2} = 1 \tag{2}$$

The probability density determination of the 2D Gaussian distribution is shown in Eq. (3), where x, μ, and Σ denote the coordinates (x, y), mean vector, and covariance matrix of Gaussian distribution, respectively.

$$f(x|\mu,\Sigma) = \frac{\exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)}{2\pi|\Sigma|^{\frac{1}{2}}} \tag{3}$$

The second-order Wasserstein distance determination is shown in Eq. (4), where N_a and N_b are Gaussian distributions modeled by $A=(cx_a, cy_a, \omega_a, h_a)$ and $B=(cx_b, cy_b, \omega_b, h_b)$, respectively.

$$W_2^2(N_a,N_b) = \left\| \left(\left[cx_a, cy_a, \frac{\omega_a}{2}, \frac{h_a}{2} \right]^T, \left[cx_b, cy_b, \frac{\omega_b}{2}, \frac{h_b}{2} \right]^T \right) \right\|_2^2 \tag{4}$$

NWD determination is expressed in Eq. (5), where C is closely related to the data set.

$$NWD\ (N_a,N_b) = \exp\left(-\frac{\sqrt{W_2^2(N_a,N_b)}}{C} \right) \tag{5}$$

The loss function expression is given in Eq. (6), where N is the number of detection frames.

$$Loss = \frac{1}{2N}\sum\nolimits_{i=1}^{N}(1-NWD_i) + \frac{1}{2N}\sum\nolimits_{i=1}^{N}(1-IoU_i) \tag{6}$$

3. Experiments

3.1. Experimental conditions

The study utilizes an Intel® Core™ i5-8265U CPU @ 1.80 GHz with 8 GB of RAM. The graphics card employed is an Intel® UHD Graphics 620, and the operating system utilized is Windows 10 Professional, 64-bit. The entire experiment relies on the pyCharm 2020.1 x64 development environment and the encoder Python 3.8. During the experiment, the image data of the shallow mud sound print of the submarine channel was divided into a training set and a validation set at a ratio of 7:3. The training set consists of 105 samples, while the validation set comprises 45 samples. The input image scale is 200 × 200, the batch size is 20, the training momentum is 0.94, the initial learning rate is set to 0.01, the weight decay is 0.0005, the number of epochs is 200, and the model is trained using Stochastic Gradient Descent (SGD) as the optimization function for model training.

3.2. Evaluation indicators

To evaluate the performance of the improved YOLOv5s-EF algorithm, this study primarily employs Precision, Recall, F1 value, and mean Average Precision (mAP) as the evaluation metrics [42,43]. Precision is defined as the ratio of the true positive rate to the sum of the true positive rate and the false positive rate (Eq. (7)). Meanwhile, recall is defined as the ratio of the true positive rate to the sum of

the true positive rate and the false negative rate (Eq. (8)). In this equation, TP represents True Positive, FN represents False Negative, and FP represents False Positive. F1 value represents the harmonic mean of precision and recall (Eq. (9)). mAP is calculated by averaging average precision values for all categories (Eq. (10)).

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{7}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{8}$$

$$\text{F1} = 2 \times \frac{\text{Precision} \times \text{Recall}}{(\text{Precision} + \text{Recall})} \tag{9}$$

$$mAP = \frac{1}{c} \sum_{j=1}^{c} AP_j \tag{10}$$

3.3. Experimental results

3.3.1. Training results

Fig. 7 depicts Precision-Confidence Curve (PCC) and Recall-Confidence Curve (RCC) during the training of the improved YOLOv5s-EF and YOLOv5s algorithms. The horizontal axis represents the confidence level, while the vertical axis represents the precision or recall rate. The shape and position of the curve may indicate the performance and stability of the detection algorithm. In Fig. 7(a), PCC curve generated by YOLOv5s algorithm exhibits an upward and leftward bend when the confidence level is below 0.18. When the confidence level exceeds 0.18, PCC curve derived from the improved YOLOv5s-EF demonstrates an upward and leftward bend. This scenario suggests that the original algorithm exhibits relatively high performance when the number of training epochs is limited. As the number of training epochs increases gradually, the improved algorithm demonstrates improved performance, resulting in higher detection accuracy compared to the original algorithm. In Fig. 7(b), Receiver Operating Characteristic (ROC) curve derived from the improved YOLOv5s-EF model demonstrates closer proximity to the upper right corner in comparison to YOLOv5s algorithm. This situation suggests that the improved algorithm outperforms the original algorithm and has the capacity to further enhance the algorithm's detection accuracy while upholding a high recall rate.

Fig. 8 illustrates bounding box regression loss curves for the training set and validation set during the training of the improved YOLOv5s-EF and YOLOv5s algorithms. Bounding box regression loss value of the improved YOLOv5s-EF algorithm is consistently lower than that of YOLOv5s algorithm throughout the training process. Specifically, the loss value of the training set stabilizes at approximately 0.004, while the loss value of the validation set stabilizes at around 0.01. Fig. 9 shows the precision and recall curves for both the training and validation sets during the training of the improved YOLOv5s-EF and YOLOv5s algorithms. At approximately 140 training epochs (Fig. 9(a)), the precision rates of the improved YOLOv5s-EF and YOLOv5s algorithms are comparable. However, in the subsequent training epochs, the precision rate of the improved YOLOv5s-EF algorithm gradually surpasses that of YOLOv5s algorithm. In Fig. 9(b), YOLOv5s algorithm demonstrates a relatively high recall when the number of training epochs ranges from 1 to 40. Similarly, the improved YOLOv5s-EF and YOLOv5s algorithms yield comparable recall values when the number of training epochs falls between 50 and 140. Furthermore, in the subsequent training epochs, the recall achieved by the improved YOLOv5s-EF algorithm consistently surpasses that of YOLOv5s algorithm. The results suggest that the improved YOLOv5s-EF algorithm exhibits superior overall detection performance compared to YOLOv5s algorithm.

3.3.2. Feature extraction visualization

To explore the evolutionary process of depth feature extraction methods, the feature maps from the middle layer of both the improved YOLOv5s-EF and the original YOLOv5s algorithms were extracted and compared (Fig. 10). The first row presents the intermediate feature map of the improved YOLOv5s-EF algorithm, while the second row presents the intermediate feature map of the

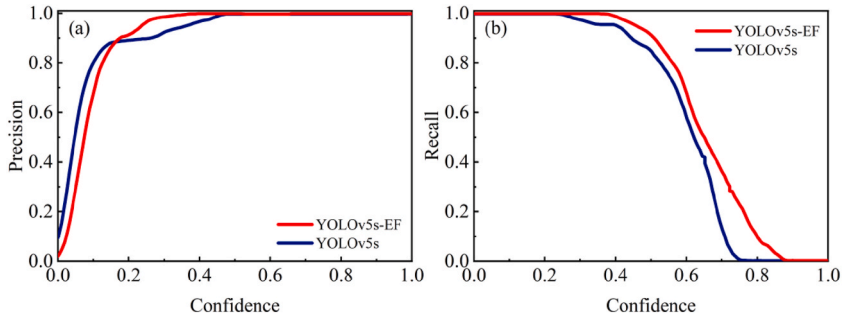


Fig. 7. PCC and RCC training results, (a) PCC, (b) RCC.

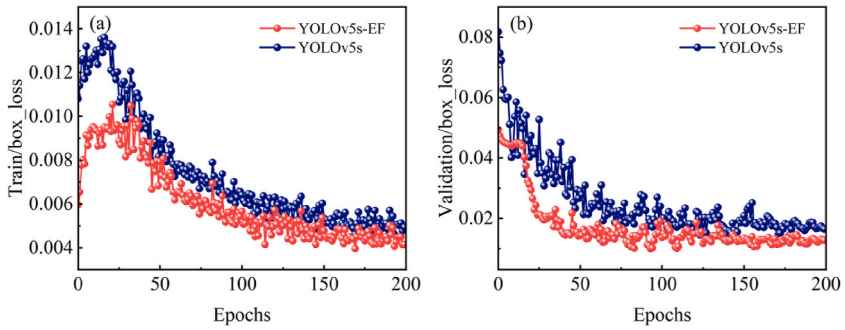


Fig. 8. Improved YOLOv5s-EF algorithm and YOLOv5s algorithm training bounding box regression loss curve, (a) training, (b) validation.

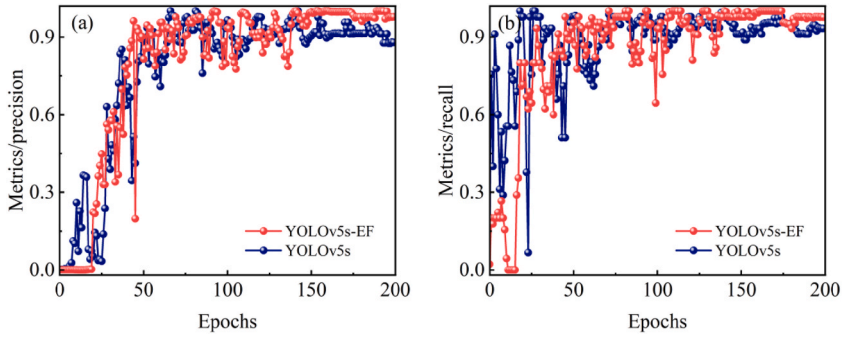


Fig. 9. Improved YOLOv5s-EF algorithm and YOLOv5s algorithm training precision and recall curve, (a) precision, (b) recall.

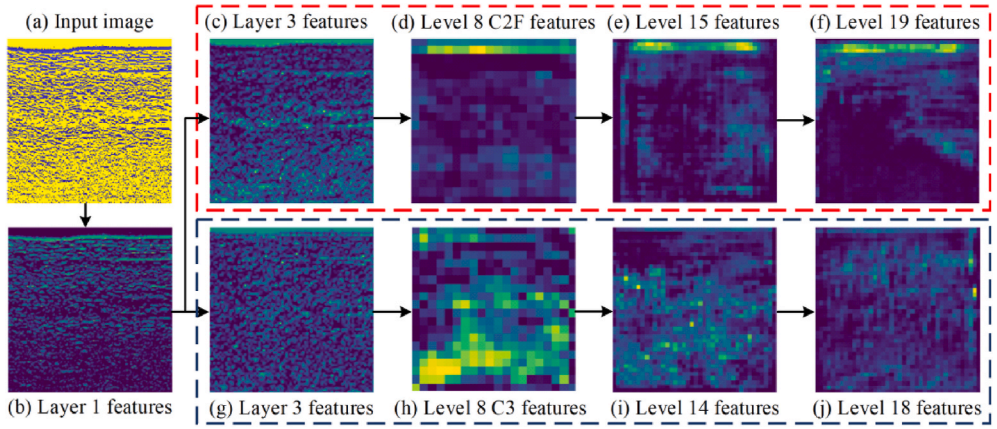


Fig. 10. Improved YOLOv5s-EF and YOLOv5s feature extraction visualization.

original YOLOv5s algorithm. Fig. 10(b), (c), and 10(g) reveals that the shallow network of the two algorithms extracts features that closely resemble the input and encompass a greater amount of pixel information, including color, texture, edges, and other geometric features. This is primarily due to the fact that the shallow network places greater emphasis on detailed information, resulting in a relatively small general receptive field, rich extracted local information, and a relatively high resolution of the obtained feature map. The improved YOLOv5s-EF algorithm demonstrates a greater extraction of shallow feature information, as evidenced in Fig. 10(c) and (g). This results in a clearer depiction of the contour of shallow mud. Fig. 10(d)-10(f) and 10(h)-10(j) illustrate that withi an increase in the number of layers, the features and outputs extracted by both algorithms exhibit relatively close similarities, indicating a higher level of semantic information. This is attributed to the deep network's enhanced capability to prioritize and extract semantic information. As the quantity of convolutions rises, the receptive field expands gradually, leading to a continuous increase in the overlapping area. Consequently, the extracted global information becomes more abundant, albeit at the cost of a relatively lower resolution in the resulting feature image. The semantic information extracted by the improved YOLOv5s-EF algorithm is more pronounced (Fig. 10(d)-10(f) and 10(h)-10(j)). This enhancement facilitates more accurate localization of shallow mud, thereby increasing the detection

confidence of the improved algorithm compared to the original algorithm.

3.4. Comparison of experiments

3.4.1. Ablation experiment

To assess the effect of each improvement on YOLOv5s-EF algorithm, this study conducted ablation experiments. The experimental results (Table 2) illustrates that the improved algorithm 1 incorporates C2F module in Backbone to enhance precision, recall, F1 value, and mAP@0.5. This can be attributed to the fact that C2F module's focus on the diversity of gradient flow, which surpasses that of C3 module, hence enabling more effective extraction of image features. Furthermore, C2F module adapts the channel count for varying algorithm proportions, thereby enhancing the algorithm's detection performance. The improved algorithm 2 incorporates EMA attention mechanism in Head to enhance precision, recall rate, F1 value, and mAP@0.5. This phenomenon can be attributed to the establishment of a multi-scale parallel sub-network by EMA module, which allows for the capture of both long and short dependencies. This enables each parallel sub-network to facilitate cross-channel interaction without reducing the number of channels, consequently enhancing the diversity and quality of feature fusion. As a result, this improves the capability of shallow mud detection and positioning. The improved algorithm 3 incorporates NWD loss function into the boundary regression loss function to enhance precision, recall rate, F1 value, mAP@0.5, mAP@0.95, and Frames Per Second (FPS). This can be attributed to the capability of NWD loss to represent bounding box as a Gaussian distribution and utilize Wasserstein distance for a more precise assessment of the disparity between the anticipated and real bounding boxes. Therefore, this approach can diminish the sensitivity to minor adjustments in the target position, subsequently enhancing the computational speed and efficiency of the algorithm. Upon implementing the aforementioned three improvements to YOLOv5s algorithm simultaneously, the precision, recall, F1 value, mAP, and FPS of the trained YOLOv5s-EF algorithm have undergone significant improvement and optimization. This can be attributed to C2F module's ability to acquire more comprehensive gradient flow information while maintaining a lightweight design. Additionally, EMA module can facilitate multi-scale fusion of the acquired gradient flow information, and NWD loss function is effective in reducing the sensitivity of weak target position offset transformations. The results indicate that the concurrent utilization of C2F, EMA, and NWD loss modules significantly enhances the overall performance of YOLOv5s, making them integral to the effective detection of shallow mud in submarine channels.

3.4.2. Algorithm comparison

The effectiveness of the improved YOLO5s-EF algorithm in the shallow mud of the submarine channel was verified by comparing it with existing Faster R–CNN [44], Mask R–CNN [45], YOLOv3 [46], YOLOv3-SPP [47], YOLOv7 [48], YOLOv7x [49], and other target detection algorithms. The results (Table 3) indicates that YOLO5s-EF algorithm achieved a detection precision of 97.8%, a recall of 97.6%, an F1 value of 97.7%, an mAP@0.5 of 98.2%, and an FPS of 51.8 on the same validation set. In comparison to Faster R–CNN, Mask R–CNN, YOLOv3, YOLOv3-SPP, YOLOv7, YOLOv7x, and YOLOv5s algorithms, the precision has been improved by 9.5%, 6.6%, 14.6%, 16.9%, 6.5%, 5.0%, and 9.9%, respectively. The FPS was 2.4 times, 2.9 times, 4.1 times, 4.3 times, 2.6 times, and 4.5 times faster than Faster R–CNN, Mask R–CNN, YOLOv3, YOLOv3-SPP, YOLOv7, and YOLOv7x algorithms, respectively. Compared to YOLOv5s, the FPS was only 1.5 slower, while the remaining recall, F1 value, and mAP@0.5 showed significant enhancements. This is mainly because compared with Faster R–CNN, and Mask R–CNN, the improved YOLO5s-EF algorithm adopts a single-phase detection method, which has a higher inference speed, and is able to complete the target detection and localization at one time, and the overall architecture is relatively simple, easy to understand and implement, and has a much better adaptability to smaller-sized targets. The improved YOLO5s-EF algorithm is distinguished by its adoption of a novel network structure featuring smaller convolution kernels and more efficient network modules, in contrast to YOLOv3 and YOLOv3-SPP. This modification serves to augment the algorithm's ability to perceive targets of varying scales, while simultaneously reducing computational load and the number of algorithm parameters. Consequently, this leads to improved detection performance and inference speed. Compared to YOLOv7 and YOLOv7x, the improved YOLO5s-EF algorithm features a lightweight network structure, enabling swift iteration and deployment. The model is capable of maintaining high accuracy and exhibits extremely fast inference speed, making it well-suited for real-time application scenarios. The results indicate that the improved YOLO5s-EF algorithm demonstrates superior capability in detecting shallow mud within the submarine channel compared to other algorithms.

3.4.3. Backbone modules

To verify the effect of C2F module on the performance of YOLOv5s algorithm, this study compared C2F module with other algorithmic modules including C3, SAC [50], DCNv2 [51], DSC [52], and RFEM [53] to determine its improvement effect. Table 4

Table 2
Ablation experiment.

Model	C2F	EMA	NWD	Precision/%	Recall/%	F1/%	mAP@0.5/%	mAP@0.95/%	FPS
YOLOv5s	×	×	×	87.9	93.3	90.5	94.8	67.2	53.3
	✓	×	×	90.7 (+2.8)	93.4 (+0.1)	92.1 (+1.6)	94.9 (+0.1)	66.2 (−1.0)	52.6
	×	✓	×	95.7 (+7.8)	99.6 (+6.3)	97.6 (+7.1)	97.4 (+2.6)	64.7 (−2.5)	51.5
	×	×	✓	89.8 (+1.9)	95.5 (+2.2)	92.5 (+2.0)	97.6 (+2.8)	69.5 (+2.3)	55.2
	✓	✓	✓	97.8 (+9.9)	97.6 (+4.3)	97.7 (+7.2)	98.2 (+3.4)	69.6 (+2.4)	51.8

Table 3
Comparison of different algorithms.

Model	Precision/%	Recall/%	F1/%	mAP@0.5/%	FPS
Faster R–CNN	88.3	91.1	89.6	89.9	21.7
Mask R–CNN	91.2	92.4	91.8	93.5	17.6
YOLOv3	83.2	97.8	89.9	95.1	12.5
YOLOv3-SPP	80.9	95.6	87.7	88.6	12.1
YOLOv7	91.3	88.9	90.1	95.8	19.8
YOLOv7x	92.8	95.6	94.2	90.9	11.6
YOLOv5s	87.9	93.3	90.5	94.8	53.3
YOLO5s-EF	97.8	97.6	97.7	98.2	51.8

Table 4
Different backbone modules.

Model	Backbone modules	Precision/%	Recall/%	F1/%	mAP@0.5/%	FPS
YOLOv5s	C3	87.9	93.3	90.5	94.8	53.3
	SAC	87.4	92.9	90.1	90.4	37.4
	DCNv2	84.4	86.7	85.5	90.7	49.5
	DSC	87.2	90.6	88.9	91.2	52.4
	RFEM	89.2	92.1	90.6	95.4	46.1
	C2F	90.7	93.4	92.1	94.9	52.6

demonstrates that YOLOv5s algorithm, improved by the introduction of C2F module, exhibits superior performance. The precision was 90.7%, with a recall of 93.4%, an F1 value of 92.1%, a mAP@0.5 of 94.9%, and an FPS of 52.6. In comparison to C3, SAC, DCNv2, DSC, and RFEM, the detection precision shows an increase of 2.8%, 3.3%, 6.3%, 3.5%, and 1.5%, respectively. Additionally, the remaining recall, F1 value, mAP@0.5, and FPS exhibit a certain degree of improvement. This can be attributed to C2F module's multi-scale fusion of low-level and high-level features through a cross-stage connection method, which enables the algorithm to better capture subtle changes and edge feature information of the target. This results yields more detailed semantic information, enhanced target perception, and improved recognition and positioning capabilities, leading to higher detection accuracy. This performance surpasses that of C3, SAC, DCN, DSCv2, and RFEM modules. The results indicate that, when considering the shallow data of the submarine channel, the incorporation of C2F module is better suited for enhancing YOLOv5s algorithm compared to other modules.

3.4.4. Attention mechanisms

To assess the efficacy of EMA attention mechanism in enhancing YOLOv5s algorithm, this study compares EMA with various other widely used attention mechanisms, including SGE [54], SE [55], ECA [56], GAM [57], and SimAM [58]. Table 5 demonstrates that YOLOv5s algorithm, when augmented with EMA, exhibits superior detection capabilities. The precision was 95.7%, with a recall of 99.6%, an F1 value of 97.6%, a mAP@0.5 of 97.4%, and an FPS of 51.5. In comparison to SGE, SE, ECA, GAM, and SimAM, the detection precision shows an increase of 8.5%, 0.6%, 1.6%, 3.1%, and 0.4%, respectively. Additionally, there are significant improvements in the remaining recall, F1 value, mAP@0.5, and FPS. This can be attributed to the fact that EMA attention mechanism differs from SGE, SE, ECA, GAM, and SimAM by incorporating a cross-channel interaction mechanism. The method considers the dependencies among different channels, extracts crucial information, and enhances the expressive capacity and discriminative power of the feature information. Simultaneously, multi-scale features are employed in modeling to improve the algorithm's ability to perceive and understand the target, thereby further enhancing the algorithm's detection performance. The results indicate that incorporating EMA attention mechanism to improve YOLOv5s enables the algorithm to effectively concentrate on the multi-scale features of the target, leading to improved detection accuracy and robustness.

3.4.5. Bounding box regression loss function

To evaluate the effectiveness of NWD loss function in improving YOLOv5s algorithm's performance, this study undertakes a detailed comparative analysis. It compares NWD loss function against several leading bounding box regression loss functions, namely CIoU loss, SIoU loss [59], EIoU loss [60], WIoU loss [61], and Focal loss [62]. In Table 6, YOLOv5s algorithm, when augmented with NWD loss, demonstrates the most favorable outcome. The precision was 89.8%, with a recall of 95.5%, an F1 value of 92.5%, a mAP@0.5 of 97.6%, and an FPS of 55.2. Compared to CIoU loss, SIoU loss, EIoU loss, WIoU loss, and Focal loss, the detection precision shows an increase of 1.9%, 5.6%, 2.1%, 0.2%, and 5.2%, respectively. Additionally, there is a certain degree of improvement in the remaining recall, F1 value, mAP@0.5, and FPS. This can be attributed to the fact that, in comparison with CIoU loss, SIoU loss, EIoU loss, WIoU loss, and Focal loss, NWD loss function has the capability to mitigate the scale differences between various targets. It achieves this by employing NWD measurement method, thereby enhancing the robustness and stability of the algorithm. Meanwhile, NWD loss function exhibits insensitivity to variations in target scales, making it more appropriate for detection algorithms with high similarity among small targets. This function also contributes to enhancing the accuracy and rate of detection algorithms. The results indicate that the application of the proposed NWD loss function for bounding box regression improves the performance YOLOv5s algorithm for detecting shallow mud data from submarine channels. This enhancement is significant when compared to the outcomes achieved using

Table 5
Different attention mechanisms.

Model	Attention mechanisms	Precision/%	Recall/%	F1/%	mAP@0.5/%	FPS
YOLOv5s	SGE	87.2	90.9	89.1	93.1	47.4
	SE	95.1	95.6	95.3	96.0	49.1
	ECA	94.1	93.3	93.7	98.0	50.2
	GAM	92.6	97.8	95.1	93.5	30.2
	SimAM	95.3	97.8	96.5	94.0	49.8
	EMA	95.7	99.6	97.6	97.4	51.5

Table 6
Loss functions for different bounding box regressions.

Model	Bounding Box loss function	Precision/%	Recall/%	F1/%	mAP@0.5/%	FPS
YOLOv5s	CIoU Loss	87.9	93.3	90.5	94.8	53.3
	SIoU Loss	84.2	94.9	89.2	89.3	52.4
	EIoU Loss	87.7	85.2	86.4	95.6	45.8
	WIoU Loss	89.6	95.8	92.6	95.3	49.2
	Focal Loss	84.6	85.3	84.9	85.8	51.8
	NWD Loss	89.8	95.5	92.5	97.6	55.2

other advanced bounding box regression loss functions.

4. Discussion

Fig. 11 shows the shallow mud truth box, the improved YOLOv5s-EF detection results and the original YOLOv5s detection results. The first row presents the shallow mud truth box, the second row presents the improved YOLOv5-EF algorithm detection results, and the third row presents the original YOLOv5s algorithm detection results. From Fig. 11(g), (h), 11(m), and 11(n) it can be seen that the improved YOLOv5s-EF algorithm has a significant improvement in detection confidence and accuracy compared to the original YOLOv5s algorithm. Fig. 11(i), (j), 11(o), and 11(p) show that the original YOLO5s algorithm has omissions in detecting relatively complex shallow mud. Conversely, the improved YOLOv5s-EF algorithm improves the overall detection ability of shallow mud. In Fig. 11(k), (l), 11(q), and 11(r), the original YOLOv5s algorithm produces multiple detection results at the same position, whereas the improved YOLOv5s-EF algorithm yields only one detection result at this position. This modification mitigates detection confusion and enhances detection confidence. Additionally, the improved YOLOv5s-EF algorithm indicates exceptional overall confidence and effectiveness in detecting shallow mud within submarine channels compared to the original YOLOv5s algorithm. YOLOv5s-EF algorithm has been enhanced with the introduction of C2F, EMA, and NWD loss, resulting in improved efficiency in training and feature

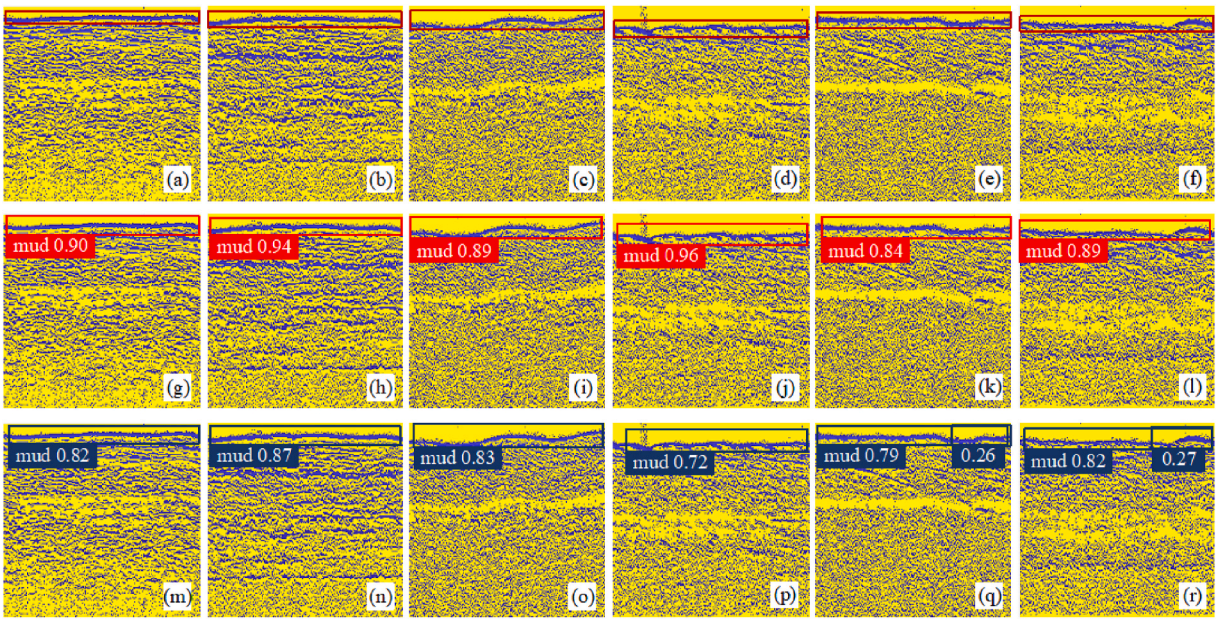


Fig. 11. Shallow mud truth box (first row) and detection results (second row YOLOv5s-EF, third row YOLOv5s).

extraction. The C3 module is replaced with C2F module, allowing the algorithm to effectively utilize correlation information between channels. This is achieved through the integration of channel contexts from multiple branches, which improves target differentiation and identification, consequently enhancing both detection speed and accuracy. Furthermore, the incorporation of EMA multi-scale attention significantly mitigate training oscillations and fluctuations, thereby enhancing the robustness of the algorithm. This approach simultaneously improves the algorithm's resistance to noisy data interference and mitigates the addresses overfitting, further enhancing detection accuracy and stability. NWD loss function for bounding box regression has been incorporated into CIOU loss function for bounding box regression. This approach has the potential to decrease the error in bounding box regression, enhance the algorithm's detection performance for small targets and targets with high aspect ratios, and improve the robustness and accuracy of the algorithm. The proposed approaches in this paper demonstrate the effectiveness of YOLOv5s-EF algorithm as a submarine mud detection method. This algorithm not only sustains high efficiency but also improves algorithmic accuracy and stability. Furthermore, it offers new insights and avenues for intelligent detection in the domain of submarine channel mud.

However, the shallow mud detection algorithm still exhibits certain limitations. The study has identified the following issues based on data analysis: Firstly, the utilization of peak sound intensity values from the sub-bottom profiler for identifying shallow mud locations, while intuitive and efficient, is compromised by unstable submarine geological signals that can lead to data loss, thereby obstructing precise localization. In the future, a combination of side scan sonar, multi-beam, drilling sampling, and other methods should be employed to enhance the submarine mud measurement to enhance the accuracy of submarine mud location delineation. Secondly, the sound print data acquired from the sub-bottom profiler can effectively indicate the presence of shallow mud within the submarine channel. The submarine channel is situated in a complex environment, making it vulnerable to noise interference caused by natural and human factors such as waves, tides, and passing ships. This interference can lead to a decrease in image quality and impact the recognition process. Future advancements are expected to leverage filters and image improvement algorithms to reduce noise, consequently enhancing image quality and precision. Furthermore, while the improved YOLOv5s-EF algorithm demonstrates enhanced detection capabilities, it also exhibits certain limitations. The algorithm is susceptible to both missed detection and false detection when it encounters complex scenes, dense targets, or a large number of overlapping targets. For future study, the algorithm should focus on optimizing and upgrading the algorithm by adopting cutting-edge enhancement strategies to improve the detection performance and efficiency.

5. Conclusion

This paper introduced YOLOv5s-EF algorithm, an advanced version of YOLOv5s, designed specifically to detect shallow mud within submarine channels. A sub-bottom profiler was initially employed to identify the submarine channel in the study area through surveying. The resulting sound print image data was subsequently utilized to visually depict the shallow mud of the submarine channel. Furthermore, substituting the local C3 module in YOLOv5s with C2F module could enable the extraction of more complex and abstract feature information. Simultaneously, the incorporation of EMA multi-scale attention into Neck module improved the algorithm's ability to detect targets across different scales. Additionally, integrating NWD loss function enhanced the optimization of bounding box positioning and elevated the performance of the detection algorithm. The shallow mud sound print image data of the submarine channel was ultimately input into the improved YOLOv5s-EF algorithm for training. The results indicated that the proposed YOLOv5s-EF algorithm improved upon YOLOv5s in terms of in precision, recall, F1 value, and mAP evaluation indices. Moreover, it exhibits promising effectiveness and feasibility for the detection of shallow mud in submarine channels. In summary, this research holds importance for achieving real-time dynamic detection, dredging, and maintenance activities related to submarine channel mud, contributing significantly to the advancement of maritime navigation safety and operational efficiency.

Funding statement

This work was supported by the Key Research and Development Program of Jiangsu Province (grant number BE2018676) and the Graduate Research and Practice Innovation Program of Jiangsu Ocean University (grant number KYCX2021-026).

Data availability statement

The authors declare that the raw data supporting the conclusions of this paper are available from the authors, upon reasonable request. Some of the numbered figure data in the text is stored at https://pan.baidu.com/s/1hBGmi2ptXPPlBgLuGtpwag?pwd=8686.

CRediT authorship contribution statement

Jiankang Hou: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Project administration, Methodology, Investigation, Formal analysis, Conceptualization. **Cunyong Zhang:** Writing – review & editing, Validation, Supervision, Project administration, Investigation, Funding acquisition, Formal analysis, Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Y. Wan, D. Roelvink, W. Li, D. Qi, F. Gu, Observation and modeling of the storm-induced fluid mud dynamics in a muddy-estuarine navigational channel, Geomorphology 217 (2014) 23–36, https://doi.org/10.1016/j.geomorph.2014.03.050.

[2] M.S. Hossain, Biological aspects of the coastal and marine environment of Bangladesh, Ocean Coast Manag. 44 (3) (2001) 261–282, https://doi.org/10.1016/S0964-5691(01)00049-7.

[3] A. Bianchini, F. Cento, A. Guzzini, M. Pellegrini, C. Saccani, Sediment management in coastal infrastructures: Techno-economic and environmental impact assessment of alternative technologies to dredging, J. Environ. Manag. 248 (2019) 109332, https://doi.org/10.1016/j.jenvman.2019.109332.

[4] S. Kanga, G. Meraj, B. Das, M. Farooq, S. Chaudhuri, S.K. Singh, Modeling the spatial pattern of sediment flow in lower Hugli estuary, West Bengal, India by quantifying suspended sediment concentration (SSC) and depth conditions using geoinformatics, Applied Computing and Geosciences 8 (2020) 100043, https://doi.org/10.1016/j.acags.2020.100043.

[5] M.A. Fattah, S.R. Morshed, A.A. Kafy, Insights into the socio-economic impacts of traffic congestion in the port and industrial areas of Chittagong city, Bangladesh, Transport Eng. 9 (2022) 100122, 10.1016/j.tr eng.2022.100122.

[6] K. Schrottke, M. Becker, A. Bartholomä, B.W. Flemming, D. Hebbeln, Fluid mud dynamics in the Weser estuary turbidity zone tracked by high-resolution side-scan sonar and parametric sub-bottom profiler, Geo Mar. Lett. 26 (2006) 185–198, https://doi.org/10.1007/s00367-006-0027-1.

[7] A. Jalal, A. Salman, A. Mian, M. Shortis, F. Shafait, Fish detection and species classification in underwater environments using deep learning with temporal information, Ecol. Inf. 57 (C) (2020) 101088, 10.1016/j.ec oinf.2020.101088.

[8] J. Li, L. Chen, J. Shen, X. Xiao, X. Liu, X. Sun, X. Wang, D. Li, Improved neural network with spatial Pyramid pooling and online datasets Preprocessing for underwater target detection based on side scan sonar Imagery, Rem. Sens. 15 (2) (2023) 440, https://doi.org/10.3390/rs15020440.

[9] A. Mathias, S. Dhanalakshmi, R. Kumar, Occlusion aware underwater object tracking using hybrid adaptive deep SORT-YOLOv3 approach, Multimed. Tool. Appl. 81 (2022) 44109–44121, https://doi.org/10.1007/s11042-022-13281-13285.

[10] M. Zhang, S. Xu, W. Song, Q. He, Q. Wei, Lightweight underwater object detection based on YOLO v4 and multi-scale attentional feature fusion, Rem. Sens. 13 (22) (2021) 4706, https://doi.org/10.3390/rs13224706.

[11] H.D. Kim, S. Aoki, H. Oh, K.H. Kim, J. Oh, Seabed sub-bottom sediment classification using artificial intelligence, J. Coast Res. 114 (sp1) (2021) 305–309, https://doi.org/10.2112/JCR-SI114-062.1.

[12] Y. Li, X. Zhang, Z. Shen, YOLO-submarine cable: an improved YOLO-V3 network for object detection on submarine cable images, J. Mar. Sci. Eng. 10 (2022) 1143, https://doi.org/10.3390/jmse10081143.

[13] Y. Yang, L. Chen, J. Zhang, L. Long, Z. Wang, UGC-YOLO: underwater environment object detection based on YOLO with a global context block, J. Ocean Univ. China 22 (2023) 665–674, https://doi.org/10.1007/s11802-023-5296-z.

[14] J. Zhang, H. Chen, X. Yan, K. Zhou, J. Zhang, Y. Zhang, H. Jiang, B. Shao, An improved YOLOv5 underwater detector based on an attention mechanism and multi-branch Reparameterization module, Electronics 12 (12) (2023) 2597, https://doi.org/10.3390/electronics12122597.

[15] T. Diwan, G. Anirudh, J.V. Tembhurne, Object detection using YOLO: challenges, architectural successors, datasets and applications, Multimed. Tool. Appl. 82 (2022) 9243–9275, https://doi.org/10.1007/s11042-022-13644-y.

[16] P. Lin, X. Liu, S. Hu, P. Li, Large deformation analysis of a high steep slope relating to the Laxiwa Reservoir, China, Rock Mech. Rock Eng. 49 (2016) 2253–2276, https://doi.org/10.1007/s00603-016-0925-0.

[17] K.S. Osasan, T.B. Afeni, Review of surface mine slope monitoring techniques, J. Min. Sci. 46 (2010) 177–186, https://doi.org/10.1007/s10913-010-0023-8.

[18] A.H. Sallenger, P.C. Howard, C.H. Fletcher, P.A. Howd, A system for measuring bottom profile, waves and currents in the high-energy nearshore environment, Mar. Geol. 51 (1–2) (1983) 63–76, https://doi.org/10.1016/0025-3227(83)90089-0.

[19] A.K. Turner, J. Kemeny, S. Slob, R. Hack, Evaluation, and management of unstable rock slopes by 3-D laser scanning, IAEG 404 (2006) 1–11. https://api.semanticscholar.org/CorpusID:230101109.

[20] M. Saleh, M. Rabah, Seabed sub-bottom sediment classification using parametric sub-bottom profiler, NRIAG Journal of Astronomy and Geophysics 5 (1) (2016) 87–95, https://doi.org/10.1016/j.nrjag.2016.01.004.

[21] P. Sitkiewicz, S. Rudowski, R. Wróblewski, J. Dworniczak, New insights into the nearshore bar internal structure using high-resolution sub-bottom profiling: the Vistula Spit case study, Mar. Geol. 419 (C) (2020) 106078, https://doi.org/10.1016/j.margeo.2019.106078.

[22] L. He, J. Zhao, J. Lu, Z. Qiu, High-accuracy acoustic sediment classification using sub-bottom profile data, Estuarine, Coastal and Shelf Science 265 (2022) 107701, https://doi.org/10.1016/j.ecss.2021.107701.

[23] F. Wang, Y. Song, S. Liu, C. Tao, X. Lin, Characteristics and sedimentological significance of acoustic anomalies in silty seabed in the Yellow River subaqueous delta, Continent. Shelf Res. 248 (2022) 104844, https://doi.org/10.1016/j.csr.2022.104844.

[24] H. Zhang, Y. Lu, X. Liu, X. Li, Z. Wang, C. Ji, C. Zhang, Z. Wang, S. Jing, Y. Jia, Morphology and origin of liquefaction-related sediment failures on the Yellow River subaqueous delta, Mar. Petrol. Geol. 153 (2023) 106262, https://doi.org/10.1016/j.marpetgeo.2023.106262.

[25] Z. Wang, Y. Jia, X. Liu, D. Wang, H. Shan, L. Guo, W. Wei, In situ observation of storm-wave-induced seabed deformation with a submarine landslide monitoring system, Bull. Eng. Geol. Environ. 77 (2018) 1091–1102, https://doi.org/10.1007/s10064-017-1130-4.

[26] J.B. Cunha, A.A. Neto, Ultrahigh-resolution seismic enhancement. The use of colored inversion and seismic attributes on sub-bottom profiler data, J. Appl. Geophys. 184 (2021) 104184, https://doi.org/10.1016/j.jappgeo.2020.104184.

[27] G. Zheng, J. Zhao, S. Li, J. Feng, Zero-shot pipeline detection for sub-bottom profiler data based on imaging principles, Rem. Sens. 13 (21) (2021) 4401, https://doi.org/10.3390/rs13214401.

[28] C. Zhu, Y. Pu, K. Yang, Q. Yang, C.L.P. Chen, Distributed optical fiber intrusion detection by image encoding and SwinT in multi-interference environment of long-distance pipeline, IEEE Trans. Instrum. Meas. 72 (2023) 1–12, https://doi.org/10.1109/TIM.2023.3277937.

[29] U. Morelia, V. Heinrich, Combining in situ monitoring using seabed instruments and numerical modelling to assess the transient stability of underwater slopes, Geological Society, London, Special Publications 477 (1) (2019) 511–521, 10.1144/SP477.8.

[30] C. Zhang, Fractal analysis of muddy submarine channel slope instability from sub-bottom profile images, Mar. Georesour. Geotechnol. 40 (6) (2022) 701–711, https://doi.org/10.1080/1064119X.2021.1933278.

[31] C. Zhang, J. Hou, Creep characteristics of muddy submarine channel slope instability, Front. Mar. Sci. 9 (2022) 999151, https://doi.org/10.3389/fmars.2022.999151.

[32] J. Hou, C. Zhang, Stability prediction of muddy submarine channel slope based on sub-bottom profile acoustic images and transfer learning, Front. Mar. Sci. 10 (2024) 1333038, https://doi.org/10.3389/fmars.2023.1333038.

[33] H. Yar, Z.A. Khan, F.U.M. Ullah, W. Ullah, S.W. Baik, A modified YOLOv5 architecture for efficient fire detection in smart cities, Expert Syst. Appl. 231 (2023) 120465, https://doi.org/10.1016/j.eswa.2023.120465.

[34] C. Hou, Z. Guan, Z. Guo, S. Zhou, M. Lin, An improved YOLOv5s-based scheme for target detection in a complex underwater environment, J. Mar. Sci. Eng. 11 (5) (2023) 1041, 10.3390/jm se11051041.

[35] A.A. Adegun, J.V. Fonou Dombeu, S. Viriri, J. Odindi, State-of-the-Art deep learning methods for objects detection in remote sensing satellite images, Sensors 23 (13) (2023) 5849, https://doi.org/10.3390/s23135849.

[36] L. Zhang, G. Ding, C. Li, D. Li, DCF-Yolov8: an improved algorithm for aggregating low-level features to detect agricultural pests and diseases, Agronomy 13 (8) (2023) 2012, https://doi.org/10.3390/agronomy13082012.

[37] D. Ouyang, S. He, G. Zhang, M. Luo, H. Guo, J. Zhan, Z. Huang, Efficient Multi-Scale Attention Module with Cross-Spatial Learning, ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Rhodes Island, Greece, 4-10 June 2023, pp. 1–5, https://doi.org/10.1109/ICASSP49357.2023.10096516.

[38] H. Liao, W. Zhu, YOLO-DRS: a bioinspired object detection algorithm for remote sensing images incorporating a multi-scale efficient lightweight attention mechanism, Biomimetics 8 (6) (2023) 458, 10.3390/biomime tics8060458.

[39] J. Wang, C. Xu, W. Yang, L. Yu, A normalized Gaussian Wasserstein distance for tiny object detection, ArXiv 2110 (2021) 13389, https://doi.org/10.48550/arXiv.2110.13389.
[40] Z. Yu, H. Huang, W. Chen, Y. Su, Y. Liu, X. Wang, YOLO-FaceV2: a scale and occlusion aware face detector, ArXiv. 2208 (2022) 02019, https://doi.org/10.48550/arXiv.2208.02019.
[41] H. Xu, L. Wang, F. Chen, Advancements in electric vehicle PCB inspection: application of multi-scale CBAM, partial convolution, and NWD loss in YOLOv5, World Electric Vehicle Journal 15 (1) (2024) 15, https://doi.org/10.3390/wevj15010015.
[42] B. Gašparović, J. Lerga, G. Mauša, M. Ivašić-Kos, Deep learning approach for objects detection in underwater pipeline images, Appl. Artif. Intell. 36 (1) (2022) 2146853, https://doi.org/10.1080/08839514.2022.2146853.
[43] D. Munteanu, D. Moina, C.G. Zamfir, Ș.M. Petrea, D.S. Cristea, N. Munteanu, Sea mine detection framework using YOLO, SSD and EfficientDet deep learning models, Sensors 22 (23) (2022) 9536, https://doi.org/10.3390/s22239536.
[44] L. Zeng, B. Sun, D. Zhu, Underwater target detection based on Faster R-CNN and adversarial occlusion network, Eng. Appl. Artif. Intell. 100 (2021) 104190, https://doi.org/10.1016/j.engappai.2021.104190.
[45] Z. Fan, W. Xia, X. Liu, H. Li, Detection and segmentation of underwater objects from forward-looking sonar based on a modified Mask RCNN, Signal, Image and Video Processing 15 (6) (2021) 1135–1143, https://doi.org/10.1007/s11760-020-01841-x.
[46] H. Yang, P. Liu, Y.Z. Hu, J.N. Fu, Research on underwater object recognition based on YOLOv3, Microsyst. Technol. 27 (2021) 1837–1844, https://doi.org/10.1007/s00542-019-04694-8.
[47] X. Su, J. Hu, L. Chen, H. Gao, Research on real-time dense small target detection algorithm of UAV based on YOLOv3-SPP, J. Braz. Soc. Mech. Sci. Eng. 45 (9) (2023) 488, 10.1007/s40430-023-04343-2.
[48] W. Yi, B. Wang, Research on Underwater small target Detection Algorithm based on improved YOLOv7, IEEE Access 11 (2023) 66818–66827, https://doi.org/10.1109/ACCESS.2023.3290903.
[49] Y. Wang, J. Yang, K. Wu, M. Hou, G. Chen, A submesoscale eddy identification dataset derived from GOCI I chlorophyll–a data based on deep learning, Earth Syst. Sci. Data Discuss. 2023 (2023) 1–23, 10.5194/essd-2023-138.
[50] S. Qiao, L.C. Chen, A. Yuille, DetectoRS: detecting objects with recursive feature pyramid and switchable atrous convolution, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 19-25 June 2021, pp. 10213–10224, https://doi.org/10.1109/CVPR46437.2021.01008. Online.
[51] X. Zhu, H. Hu, S. Lin, J. Dai, Deformable ConvNets V2: more deformable, better results, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 9308–9316, https://doi.org/10.1109/CVPR.2019.00953. Long Beach US.
[52] M.G. Nascimento, R. Fawcett, V.A. Prisacariu, DSConv: efficient convolution operator, in: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 20-26 October 2019, pp. 5148–5157, https://doi.org/10.1109/ICCV.2019.00525. South Korea.
[53] Q. Tong, E. Zhang, S. Wu, K. Xu, C. Sun, A real-time detector of chicken healthy status based on modified YOLO, Signal, Image and Video Processing 17 (8) (2023) 4199–4207, https://doi.org/10.1007/s11760-023-02652-6.
[54] X. Li, X. Hu, J. Yang, Spatial group-wise enhance: improving semantic feature learning in convolutional networks, ArXiv. 1905 (2019) 09646, https://doi.org/10.48550/arXiv.1905.09646.
[55] J. Hu, L. Shen, S. Albanie, G. Sun, E. Wu, Squeeze-and-Excitation networks, IEEE Trans. Pattern Anal. Mach. Intell. 42 (8) (2020) 2011–2023, https://doi.org/10.1109/TPAMI.2019.2913372.
[56] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, Q. Hu, ECA-net: efficient channel attention for deep convolutional neural networks, Online, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 11534–11542, https://doi.org/10.1109/CVPR42600.2020.01155.
[57] Y. Liu, Z. Shao, N. Hoffmann, Global attention mechanism: retain information to enhance channel-spatial interactions, ArXiv. 2112 (2021) 05561, https://doi.org/10.48550/arXiv.2112.05561.
[58] V.C. Mahaadevan, R. Narayanamoorthi, R. Gono, P. Moldrik, Automatic identifier of socket for electrical vehicles using SWIN-transformer and SimAM attention mechanism-based EVS YOLO, IEEE Access 11 (2023) 111238–111254, https://doi.org/10.1109/ACCESS.2023.3321290.
[59] Z. Gevorgyan, SIoU loss: more powerful learning for bounding box regression, ArXiv 2205 (2022) 12740, 10.48550/arXiv.2205.12740.
[60] H. Peng, S. Yu, A systematic IoU-related method: beyond simplified regression for better localization, IEEE Trans. Image Process. 30 (2021) 5032–5044, https://doi.org/10.1109/TIP.2021.3077144.
[61] Z. Tong, Y. Chen, Z. Xu, R. Yu, Wise-IoU: bounding box regression loss with dynamic focusing mechanism, ArXiv. 2301 (2023) 10051, https://doi.org/10.48550/arXiv.2301.10051.
[62] T. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, Focal loss for dense object detection, IEEE Trans. Pattern Anal. Mach. Intell. 42 (2017) 318–327, https://doi.org/10.1109/TPAMI.2018.2858826.