Journal of Systems Engineering and Electronics
Vol. 34, No. 4, August 2023, pp.981 – 991

An accurate detection algorithm for time backtracked projectile-induced water columns based on the improved YOLO network

LUO Yasong [1], XU Jianghu [1,*], FENG Chengxu [1], and ZHANG Kun [2]

1. College of Weapon Engineering, Naval University of Engineering, Wuhan 430033, China;
2. Unit 91115 of the PLA, Zhoushan 316000, China

Abstract: During a sea firing training, the intelligent detection of projectile-induced water column targets in a firing video is the prerequisite for and critical to the automatic calculation of miss distance, while the correct and precise calculation of miss distance is directly affected by the accuracy, false alarm rate and time delay of detection. After analyzing the characteristics of projectile-induced water columns, an accurate detection algorithm for time backtracked projectile-induced water columns based on the improved you only look once (YOLO) network is put forward. The capability and accuracy of detecting projectile-induced water column targets with the conventional YOLO network are improved by optimizing the anchor box through K-means clustering and embedding the squeeze and excitation (SE) attention module. The detection area is limited by adopting a sea-sky line detection algorithm based on gray level co-occurrence matrix (GLCM), so as to effectively eliminate such disturbances as ocean waves and ship wakes, and lower the false alarm rate of projectile-induced water column detection. The improved algorithm increases the mAP_{50} of water column detection by 30.3%. On the basis of correct detection, a time backtracking algorithm is designed with mean shift to track images containing projectile-induced water column in reverse time sequence. It accurately detects a projectile-induced water column at the time of its initial appearance as well as its pixel position in images, and considerably reduces detection delay, so as to provide the support for the automatic, accurate, and real-time calculation of miss distance.

Keywords: object recognition, projectile-induced water column, you only look once (YOLO), K-means, squeeze and excitation (SE), mean shift.

DOI: 10.23919/JSEE.2023.000106

1. Introduction

In a sea firing training, a ship fires its naval gun at a moving target object towed fast at sea. Accurately calculating the miss distance of each projectile is an important work for evaluating its firing results. In order to calculate miss distance online and grade each firing in a timely manner, the projectile-induced water column targets in a surveillance video of firing range should be automatically detected, which involves the automatic recognition of projectile-induced water column targets in the images taken from the video, and the pixel positions of projectile-induced water columns in these images. The miss rate and false alarm rate of water column detection directly affect the omissions and false alarms in projectile firing results. Additionally, because the shot target object moves very fast, the calculation accuracy of miss distance also relates to whether small projectile-induced water column targets can be detected with low delay as soon as projectiles touch the sea surface.

When a projectile fired by a ship gun falls into water, it generates a projectile-induced water column that may last for only 2−5 s. The water column may be small when the projectile just touches water, and it is characteristically similar to ocean waves, wakes of the ship towing the shot target object, clouds and other targets in images. For this reason, its accurate detection with low delay is highly challenging. In recent years, target detection technology has been quickly developing. Zhang et al. focused on the study of detection with Gaussian and non-Gaussian noise [1,2]. In the image detection field, algorithms such as TridentNet and CenterNet feature very high detection accuracy [3,4], but require a high amount of computation and fail to realize the real-time processing of video streams. Moreover, semantic segmentation algorithms can achieve the accurate extraction of target contour and the high accuracy of positioning in images [5], but face a very high false alarm rate in the refinement of highly random

Manuscript received February 21, 2022.
*Corresponding author.
This work was supported by the National Natural Science Foundation of China (51679247).

targets such as projectile-induced water columns. However, the you only look once (YOLO) algorithm can understand images on the whole [6−9], and realize the automatic recognition and positioning of targets in images. It is able to achieve satisfactorily real-time processing, but functions poorly in the detection of small targets such as the projectile-induced water columns that are just out of water because of the restriction of grid mechanism.

For the purpose of high accuracy, low false alarm rate and low time delay in the detection of projectile-induced water column targets, a time backtracking projectile-induced water column detection algorithm based on the improved YOLO network is therefore proposed. It is developed with the following idea: the characteristics of projectile-induced water columns in images are analyzed firstly. Using the advantages of real-time processing of the YOLO algorithm [10−12], the anchor box and the network structure of the traditional YOLO algorithm are optimized to enhance the detection accuracy and lower the miss rate of large projectile-induced water columns that have risen. Besides, considering that a projectile-induced water column target starts from the sea-sky line, the automatic detection of the sea-sky line is employed to eliminate such disturbances as ocean waves and ship wakes, and reduce the false alarm rate. With the water column bounding box obtained from the improved YOLO algorithm as its benchmark, the correlation between adjacent frame images is utilized together with mean shift to realize the tracking of projectile-induced water column targets in images in reverse time sequence. In this way, the time delay of detection can be reduced by finding out the accurate time when water columns come out of water. The position of a bounding box is updated in the process of tracking in reverse order to increase the positioning accuracy, and systematically improve the detection quality of projectile-induced water column targets. This will subsequently provide the support for automatically and accurately calculating the miss distance of projectiles.

2. Characteristic analysis of projectile-induced water columns and selection of data set

After a projectile fired from a ship gun touches the sea surface, it may induce a transient water column with the height of approximately 3−5 m and the diameter of 1−2 m. If observed distantly, its size is not large in an optical image. Projectile-induced water columns have uncertain shapes because of various complicated factors such as the projectile's speed and attitude of touching the sea. Their shapes are noticeably and stochastically different. The color of water columns is slightly white, and greatly similar to that of the sky, clouds, and ocean waves in the background. There is not much color difference among them. If ship guns fire for effect or salvo, multiple projectil touches sea surface at the same time or sequentially with short time intervals. In this case, lots of water columns appear in images, and there may also be some image overlap between different water columns. Moreover, detection is also affected by such factors in marine environment as high ocean wave, dense fog, rain, and strong light. The image detection method has therefore faced very high requirements to ensure the simultaneous, real-time and accurate detection of multiple projectile-induced water column targets featured by transient and stochastic appearance, small color difference and small size. YOLO (version 4), abbreviated as YOLOv4, is a deep learning network [13,14]. As an image detection method, it has been widely applied at present. By utilizing a lightweight network structure, it can understand and extract targets from images on the whole, and is characterized by strong capability of multi-target recognition, resistance to disturbance, and capability of real-time processing [15,16]. Hence, it is an ideal method for the detection of projectile-induced water columns.

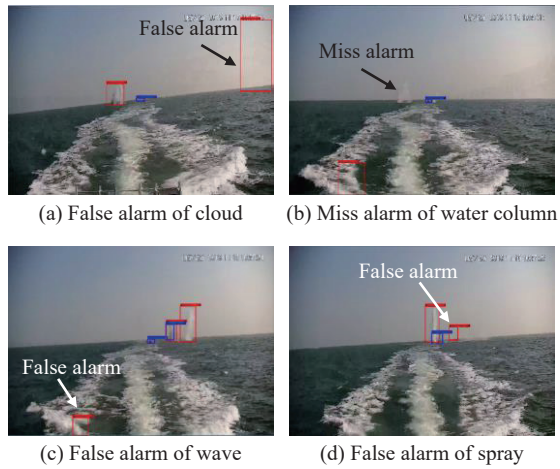
The selection of training dataset has a great influence on the final detection and recognition capability of the YOLO algorithm. Projectile-induced water columns are more complicated than some ordinary objects such as vehicles and humans. After a projectile touches sea surface, the water column produced by the explosion will go through three stages: generation, development and disappearance, as shown in Fig. 1. The entire process may last for approximately 2−5 s. After a projectile-induced water column disappears, the sea surface will restore its smoothness, leaving no trace of firing. In the entire process from generation to disappearance, a projectile-induced water column changes continuously in shape, size and color. When a water column is caused just after the explosion, it is small and not much different from such targets as ship wakes and waves. At this time, it is not suitable for the dataset used in training an algorithm, since it may lead to very high false alarm rate. At the stage of disappearance, a water column is atomized, and becomes too transparent, so that it is not much different

from the sky and clouds in the background. Similarly, it is not suitable for dataset either. Therefore, only a water column at the stage of development is suitable for dataset since its outline is characteristically noticeable. Following this idea, some unsuitable data is removed in this paper to eventually collect and sort out 2 900 images for the water columns induced by the projectiles fired by ship guns to the sea in different scenarios, and take them as the dataset for the detection of water column targets. Additionally, the software LabelImg is employed to mark the water columns in these images. In order to better detect the center of each water column with the trained algorithm, these water columns are marked under the following standards: the marking scope is limited to the highlighted area with concentrated spray, and does not cover the edge of each water column; the water columns are not marked if they have diffused or their color have faded; the sprays are not marked if they are caused by the fall of side water columns; the triangular water columns have only the areas with brighter color and striking shape marked. The marking format is set to YOLO format, that is, [class, x, y, width, height]. In the format, class represents the type of target; x and y indicate the coordinates at the center of a target after normalization; width and height are the width and height of a target after normalization.

(a) Generation (b) Development (c) Disappearance

Fig. 1 Process of water column

Conventional YOLOv4 network is trained with the above dataset. The detection results of projectile-induced water columns with the trained YOLOv4 network are shown in Fig. 2.



(a) False alarm of cloud (b) Miss alarm of water column

(c) False alarm of wave (d) False alarm of spray

Fig. 2 Diagram of recognition with the conventional YOLOv4 algorithm

It is evident that the YOLOv4 network can detect water columns with prominent development, but still easily causes such omissions as given in Fig. 2, and mistakes wakes, ocean waves and clouds as water columns, which will result in the miss rate and false alarm rate in the subsequent calculation of miss distance. In addition, the conventional YOLOv4 algorithm has low accuracy of positioning water column targets, and the size of its bounding box does not match with the shape of water columns, causing some error in the calculation of miss distance. The conventional YOLOv4 algorithm for target detection should be therefore improved to enhance the recognition correctness and positioning accuracy of projectile-induced water column targets, so as to lower the false alarm rate and guarantee the accuracy and credibility in the subsequent calculation of miss distance and the evaluation of training results.

3. An improved YOLOv4 algorithm for projectile-induced water column detection

In this paper, the conventional YOLOv4 network is improved to cope with the problems in the detection of projectile-induced water column targets. With regard to the problem that bounding box is dimensionally unmatched with the shape of targets, the K-means clustering algorithm is utilized to determine the size of the bounding box better fitting with the shape of projectile-induced water columns. In order to reduce miss rate, the squeeze and excitation (SE) attention module is embedded into the YOLOv4 backbone network, which alleviates the disappearance of characteristics of small and weak targets. Moreover, a sea-sky line detection algorithm is employed to preliminarily process the images, which limits the detection area to the vicinity of the sea-sky line in order to prevent such disturbances as ship

wakes, ocean waves and clouds in the background from causing false alarms.

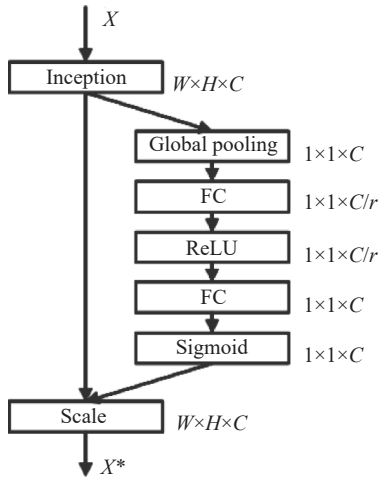
3.1 K-means clustering algorithm

When the YOLOv4 network is employed, nine anchor boxes must be preset to preliminarily determine the length-width ratio of targets to be detected. The targets in images and the areas to which they belong, that is, bounding boxes, are predicted on the basis of these anchor boxes. Projectile-induced water columns are not stationary. When a water column rises out of water, it is thin and long. Its size and shape constantly vary with time. The default anchor boxes of YOLOv4 are not applicable in detecting the moving targets of a single type. For this reason, the K-means clustering algorithm is introduced in this paper to cluster the size of water columns in the dataset and obtain the anchor box values that better match with the water column targets.

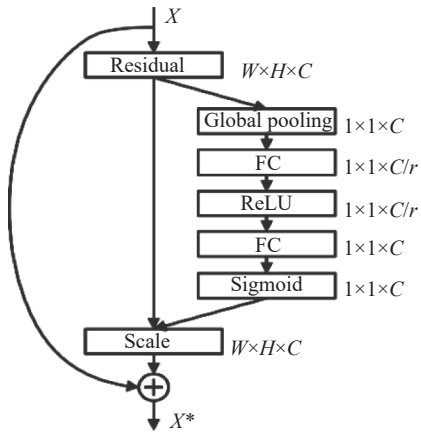
The K-means clustering algorithm is an unsupervised clustering algorithm [17−20]. The lengths and widths of the frames for all marked projectile-induced water columns in the dataset are used to form a data point set, which is input into the clustering algorithm. At first, the algorithm selects nine centroids randomly. The distance from each point (a pair of length and width) in the point set to each centroid is then calculated. On this basis, the point is marked into the same class as the most adjacent centroid. The first clustering is completed. For the clustered point set, the centroids of each class are recalculated. Based on these recalculated centroids, clustering is then carried out again for all the points. After that, centroids are calculated again. The process is repeated until no change happens to centroids. Eventually, nine centroids are obtained from clustering and selected as anchor boxes. With these anchor boxes, the YOLOv4 network could detect and position the targets based on the sizes of actual projectile-induced water columns, so as to enhance the recognition capability and positioning accuracy of the algorithm.

3.2 SE attention mechanism

In a mission of water column detection, the characteristics of targets diminish as the flow of data continuously reaches the deep layer of the network. During detection, some weak and small targets such as projectile-induced water columns may be easily missed. For this reason, an attention module is embedded at an appropriate position in the YOLOv4 network to improve the network's capability of learning key channels, and further enhance the recognition rate of water columns. The SE module proposed in [21−23] is a lightweight attention mechanism module as presented in Fig. 3 where FC means full connected, and ReLU means rectified linear unit. The module can be easily added into a network model, but causes a slight increase in model complexity and computational cost, which does not undermine the real-time nature of processing.



(a) Inception structure



(b) Residual structure

Fig. 3 SE attention mechanism

In order to enhance channel characteristics before the characteristics of small projectile-induced water columns disappeared, an SE attention module is embedded in the middle of the YOLOv4 backbone network in this paper. On the one hand, the SE module is added to the ResNet (Res) module of the original network to generate a new SE-Res module in place of the Res module in the original cross stage partial (CSP) module. On the other hand, the SE module is added behind the concatenate unit of the original CSP module to form a new SE-CSP module in place of all CSP modules in the original YOLOv4 backbone network. The revised network structure is as shown in Fig. 4.

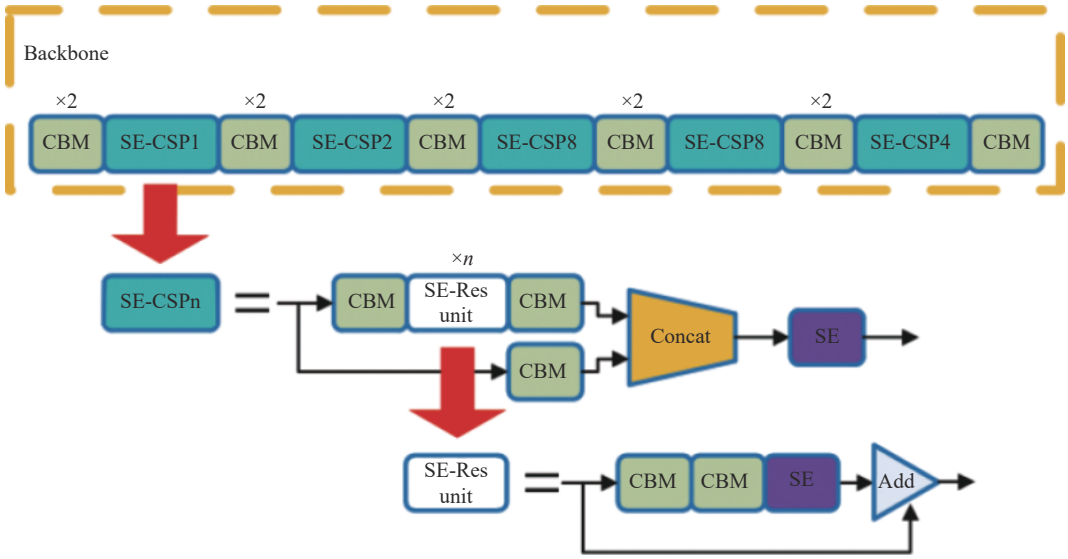


Fig. 4 Schematic diagram of SE-YOLOv4 algorithm modification

3.3 Sea-sky line detection based on gray level co-occurrence matrix (GLCM)

During a firing training, optical observation equipment is normally installed on a surface ship towing a target object, so that the projectile-induced water columns in the images taken are always adjacent to the sea-sky line. For this reason, locating the sea-sky line and detecting water columns only near the sea-sky line can further improve detection rate and lower false alarm rate [24–27]. Based on the GLCM, a fast sea-sky line detection method is devised in this paper to complete the sea-sky line detection for a 1 920×1 080 high definition image within 0.02 s.

As a method for describing the textural characteristics of images, GLCM has been widely applied in the field of image analysis. It finds out the number of gray level pairs (f_1, f_2) formed between any point and multiple points with a fixed step length in a direction on a grayscale image. Normally, it is calculated in four directions, that is, 0°, 45°, 90°, and 135°. The gray level statistics in these four directions form a GLCM, which can satisfactorily display the textural characteristics of images. On this basis, the sea-sky line can be located. The original image is traversed for many times, so that computation is enormous if the image size or its gray value is very large. Enlightened by the study in [28], a $W\times H$ high resolution image is uniformly segmented into 16×16 rectangular areas in this paper before calculating the gray level matrix. The size of each area is represented by $x\times y$. The gray sum in each area is calculated separately, and then proportionally lowered to less than 16 and rounded down to generate a new grayscale image S after gray compression as follows:

$$S\underset{i\leqslant 16, j\leqslant 16}{(i,j)} = \left[\frac{16\sum\limits_{x'\leqslant x, y'\leqslant y} g(x',y')}{255xy}\right] \tag{1}$$

where $g(x',y')$ is the gray value at the point (x',y'); $S(i,j)$ is the gray sum rounded down for the area (i,j).

The size of a new grayscale image is 16×16, and its gray level is also 16. The quality of the image might be undermined to some extent, but the effect of detection is not undermined since sea-sky line detection does not require high resolution. The GLCM P_S is obtained with the compressed grayscale image as follows:

$$\begin{aligned} P_S(i,j,d,\theta) = \#\{&[S(i,j), S(i+m, j+n)] \mid \\ &m = d\cos(\theta), n = d\sin\theta, (i,j) \in 16\times 16\} \end{aligned} \tag{2}$$

where θ = 0°, 45°, 90°, 135°; d is the distance from the point i to the point j; $\#(x)$ indicates the number of elements in the set x.

The GLCMs of four directions are normalized respectively as follows:

$$P(i,j) = \frac{P_S(i,j,d,\theta)}{R} \tag{3}$$

where R is the normalization coefficient, and it could be obtained for $N\times N$ size images as follows:

$$R = \begin{cases} N(N-1), & \theta = 0^\circ \text{ or } 90^\circ \\ (N-1)^2, & \theta = 45^\circ \text{ or } 135^\circ \end{cases}. \tag{4}$$

The GLCM $P(i,j)$ is then utilized to calculate the contrast Con:

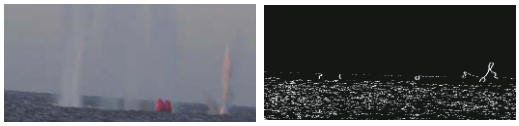
$$\text{Con} = \sum_{i=1}^{16}\sum_{j=1}^{16}(i-j)^2 P(i,j). \tag{5}$$

The position with the largest contrast change in the

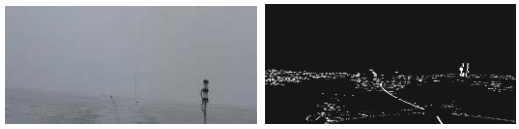
vertical direction is taken as the position of sea-sky line. Moreover, the range of 50 pixels above it and 25 pixels below it is cut out as the detection zone of the algorithm. The sea-sky line detection is highly sensitive to blue color, so that the blue channel of the image is taken as the input. The detection effect is as given in Fig. 5. It is evident that this algorithm could clearly reveal the texture of the sea-sky line, while greatly detecting the sea-sky line with the inclination of up to 45° and the sea-sky line in the background of dense fog.

(a) Situation under small shake

(b) Situation under big shake



(c) Situation under multi water columns



(d) Situation under lighthouse

Fig. 5 Detection effect of sea-sky line based on GLCM

3.4 Detection capability test of the improved algorithm

This paper presents a test with the dataset of target object detection to verify the effectiveness of the improved YOLOv4 network in the detection of projectile-induced water columns. The test environment is conducted with Tensorflow2.3.0-GPU deep learning framework, AMD Ryzen 7 4800H with Radeon Graphics CPU, 32 GB memory, NVIDIA GeForce RTX 2 060 graphics card and 6 GB video memory. Considering that the COCO dataset contains a variety of data, and the weights obtained from its trained network are widely applicable. For this reason, transfer learning is adopted, and the weights of COCO network are taken as the initial weights to train the original and improved YOLOv4 networks. This has considerably shortened the time of training while ensuring the effect of detection. The input images are sized 512×512, and the number of anchor boxes is nine. After K-means clustering on the dataset of target object detection, the size of these anchor boxes is (5, 8; 10, 36; 11, 16; 14, 55; 17, 76; 19, 53; 27, 24; 28, 71; 45, 140). The enlarging scale of anchor boxes is set to [1.2, 1.2, 1.05], and the learning rate is set to 0.001.

Intersection-over-union (IoU) is taken as the basis for judgment. When IoU $\geqslant$ 0.5, it is judged that a water column target is correctly recognized. The network performance is assessed with four indicators including precision P, recall R, mean average precision mAP, and frame per second (FPS). The improvement mechanism for each model and their effect in the detection of projectile-induced water columns are presented in Table 1. Among them, transfer learning is adopted to train the models YOLOv4_0, YOLOv4_1, and YOLOv4_3, and the training epochs are set to 180. The SE attention mechanism is introduced to the models YOLOv4_2, YOLOv4_4, and YOLOv4_5. Transfer learning could not be used since the algorithm structure is changed, so the training must be implemented from scratch and the training epochs are set to 1 000.

Table 1 Comparison of detection effect among models

Mode	K-means	SE	Sea-sky line	P/%	R/%	mAP_{50}/%	FPS
YOLOv4_0				76.2	56.2	53.5	42
YOLOv4_1	√			88.3	66.0	66.3	42
YOLOv4_2		√		88.7	78.7	75.6	38
YOLOv4_3			√	90.0	67.7	67.0	26
YOLOv4_4	√	√		90.3	80.0	79.4	28
YOLOv4_5	√	√	√	**94.5**	**83.0**	**83.8**	21

As shown in the comparison, K-means clustering has barely affected the operation speed of network, but effectively enhances the precision of detection and increases mAP_{50} by 12.8%. The SE attention mechanism is embedded into the backbone network, so that it increases the amount of computation, and causes a slight decrease in FPS. However, mAP_{50} is increased by 22.1% compared with the original network, as well as the increase of the

recall R, which greatly improves the network's capability of recognizing and detecting small projectile-induced water column targets. The sea-sky line detection algorithm affects FPS greatly, but it significantly enhances the network's accuracy of detection. The reason is that the algorithm concentrates on the detection areas nearby the sea-sky line to effectively prevent the detection of water columns from being disturbed by ocean waves, ship wakes, and clouds. After K-means clustering, SE attention mechanism, and sea-sky line detection are simultaneously applied in the conventional YOLOv4 network, mAP_{50} finally reaches 83.8%, a 30.3% increase compared with the original network. Therefore, the precision of water column detection is remarkably enhanced. Meanwhile, the improved algorithm achieves 21 frames per second which also satisfies the requirements for real-time detection. The detection effect with the improved network for the example in Fig. 2 is given in Fig. 6. It is obvious that the improved algorithm can effectively enhance the correct recognition of projectile-induced water columns, reduce the false alarms in the detection, and significantly improve the effect of water column detection.

(a) Avoid alarm of cloud (b) Valid alarm of water column

(c) Avoid alarm of wave (d) Avoid alarm of spray

Fig. 6 Detection effect of the improved YOLOv4 network

4. A time backtracking algorithm for projectile-induced water columns

As detailed in Section 2, the improved YOLOv4 network uses the images of the water columns in the development stage that has risen as its training data set. Therefore, it processes more noticeable and larger water columns to reduce false alarms in the detection to some extent. However, when the algorithm correctly detects a large water column, there is a random time delay of approximately 0.5−1 s from the moment the water column just generated. If the shot target object is towed with a speed of more than 20 knots, such time delay might result in a large difference (more than 5−10 m) between the positions of the shot target object at the detection time and that at the just generated time of the projectile-induced water column, which will cause an intolerable error in the calculation of miss distance. For this reason, a time backtracking algorithm is devised to accurately detect the produced moment and the pixel positions of projectile-induced water columns just coming out of water at the time of its generation. The algorithm is implemented in the following process: the time when the improved YOLOv4 network detected water column targets is taken as the start of backtracking N frames of images. The recognition box of the water column detected by the improved YOLOv4 network is taken as the reference position to cut out the small area screenshots for all these N backtracked frames nearby the position. Within the detection time delay of 0.5−1 s, the position of a projectile-induced water column changes continuously and slightly in the adjacent frames. Hence, these small area screenshots from backtracking could not only demonstrate in reverse order the entire process from a water column rising out of water to its just generation, but also contain only the single water column in the current detection, which reduce the difficulty of subsequent processing and the complexity of computation. Based on the position of the water column detected for the first time by the above improved YOLO algorithm, the water columns in these time reversed screenshots is tracked and detected with mean shift to eventually determine the just generated time of the projectile-induced water column as well as its initial pixel position. The time backtracking algorithm is conducted in the following procedure:

(i) The frames to be detected from a firing video are input into the improved YOLOv4 algorithm proposed in Section 3. When the algorithm detects a water column target in any frame, the position of the bounding box is recorded as $\boldsymbol{B}_0 = [x_c, y_c, w, h]_0$, where x_c, y_c are the coordinates of the central point in the bounding box, and w, h are the height and width of the bounding box.

(ii) The bounding box B_0 is slightly expand by $(\Delta w, \Delta h)$ and obtain the screenshot area $\boldsymbol{C}_0 = [x_c, y_c, w+\Delta w, h+\Delta h]_0$. After backtracking N frames, the screenshot areas of current frame and the N backtracked frames are taken based on $\boldsymbol{C}_0$, arranged in reverse time sequence, and numbered from the 0th to the Nth frame, so as to obtain the reverse development diagram of the projectile-induced water column.

(iii) The height of a reversely developing water column decreases gradually. For the purpose of highlighting the water column and reducing the disturbances in the background, only the bottom one fifth of $\boldsymbol{B}_0$ is taken as

the position of the initial bounding box for the first backtracked image, that is, $\boldsymbol{B}_0'=[x_c,y_c/5,w,h/5]_0$. The mean shift [29,30] algorithm is utilized to track the water column in the first backtracked image and the position of the bounding box is updated to obtain $\boldsymbol{B}_1'=[x_c,y_c/5,w,h/5]_1$, whose statistical diagram of color probability distribution S_1 is calculated.

(iv) Similarly, B_1' is taken as the position of the initial bounding box for the second backtracked image. The mean shift algorithm is used again to obtain the position of the bounding box on the second backtracked image as well as the statistical diagram of color probability distribution S_2. This process continues until the water column disappears abruptly in the nth backtracked image who only has clear background. At this time, S_n suddenly changes from dispersed distribution to concentrated distribution compared with the 0th to the $(n-1)$th frames in which the water column coexists with the background. The target tracking algorithm is halted then and the $(n-1)$th frame is the image in which the water column just appears, and $\boldsymbol{B}_{n-1}'=[x_c,y_c/5,w,h/5]_{n-1}$ represents the pixel position of the water column in the image when it just appears.

(v) The next frame is further input. If the improved YOLOv4 algorithm detects the water column target, IoU is taken to compare the bounding box in the current frame with $\boldsymbol{B}_0$ in the previous frame for position shift. If the two bounding box positions are adjacent, it is believed that the water column has been backtracked and the detection is proceeded with the next frame. If the position shift is large, it is believed that a new water column is detected and has not been backtracked. In this case, $\boldsymbol{B}_0=[x_c,y_c,w,h]_0$ is updated, and returned to step (ii) to backtrack the new water column until all the frames of the firing video are backtracked.

In the implementation of the time backtracking algorithm, it is crucial to use the mean shift algorithm to constantly adjust the position of the bounding box for a water column, so as to track the water column in adjacent frames. The mean shift algorithm is a kernel density estimation method [3,4]. It depends only on the sample points in the characteristic space to compute its density function value. If samples are sufficiently taken, density estimation can be performed for the data following any distribution. The mean shift tracking algorithm aims to move the centroid of bounding box using the mean shift vector in the detection area. The moved centroid will be taken as the new central point of the next bounding box and the mean shift vector is computed again. Through continuous iteration, the center of the bounding box keeps moving to track the targets in images.

In d-dimension space, the mean shift vector for given n sample points $\boldsymbol{x}_i(i=1,2,\cdots,n)$ is basically in the following form:

$$\boldsymbol{M}_h=\frac{1}{k}\sum_{\boldsymbol{x}_i\in\boldsymbol{S}_h}(\boldsymbol{x}_i-\boldsymbol{x}) \tag{6}$$

where $\boldsymbol{S}_h$ is a high-dimensional spherical area with the radius of h; k indicates that k points fall into the area $\boldsymbol{S}_h$ among n sample points.

As a matter of fact, each sample point in the area is differently important to the central point $\boldsymbol{x}$. Normally, a sample point more adjacent to the center has a higher reference value and a larger weight than the sample points at the edge. Hence, the Epanechnikov kernel function showed in (7) is introduced into the basic form of the mean shift vector.

$$K_E(\boldsymbol{x})=\begin{cases}c\left(1-\|\boldsymbol{x}\|^2\right), & \|\boldsymbol{x}\|\leqslant 1\\ 0, & \text{others}\end{cases} \tag{7}$$

The improved mean shift vector is in the following form:

$$M_{h,K_E}=\frac{C_{K_E,d}}{nh^d}\sum_{i=1}^{n}K_E\left(\left\|\frac{\boldsymbol{x}-\boldsymbol{x}_i}{h}\right\|^2\right) \tag{8}$$

where $\dfrac{C_{K_E,d}}{nh^d}$ is the unit density. In order to obtain the maximum of M_{h,K_E}, the derivative of M_{h,K_E} can be obtained as follows:

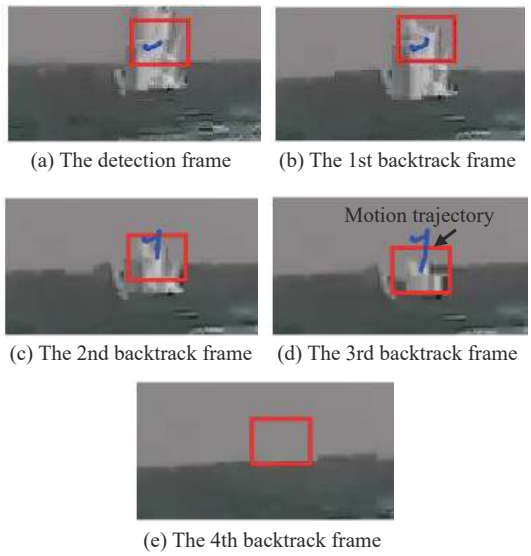
$$M'_{h,K_E}=\frac{2C_{K_E,d}}{nh^{d+2}}\sum_{i=1}^{n}(\boldsymbol{x}-\boldsymbol{x}_i)K_E'\left(\left\|\frac{\boldsymbol{x}-\boldsymbol{x}_i}{h}\right\|^2\right)=$$
$$-\frac{2C_{K_E,d}}{nh^{d+2}}\left[\sum_{i=1}^{n}K_E'\left(\left\|\frac{\boldsymbol{x}-\boldsymbol{x}_i}{h}\right\|^2\right)\right]\left[\frac{\sum_{i=1}^{n}\boldsymbol{x}_iK_E'\left(\left\|\frac{\boldsymbol{x}-\boldsymbol{x}_i}{h}\right\|^2\right)}{\sum_{i=1}^{n}K_E'\left(\left\|\frac{\boldsymbol{x}-\boldsymbol{x}_i}{h}\right\|^2\right)}-\boldsymbol{x}\right]. \tag{9}$$

In order to make $M'_{h,K_E}=0$, the second term must be 0. Thus the coordinates of the center in the new bounding box are

$$\boldsymbol{x}=\frac{\sum_{i=1}^{n}\boldsymbol{x}_iK_E'\left(\left\|\frac{\boldsymbol{x}-\boldsymbol{x}_i}{h}\right\|^2\right)}{\sum_{i=1}^{n}K_E'\left(\left\|\frac{\boldsymbol{x}-\boldsymbol{x}_i}{h}\right\|^2\right)}. \tag{10}$$

The effect of the time backtracking algorithm based on mean shift in tracking a projectile-induced water column target is presented in Fig. 7. The backtracked images in Fig. 7 are the time reversed screenshot areas near the water column recognition box detected by the improved YOLOv4 network. As revealed in the motion trajectory of the center of the bounding box in Fig. 7, the algorithm

could automatically track the reverse development of the water column, and adaptively complete the adjustment of the bounding box including translation and descent, so as to achieve the stable reverse tracking of the water column.



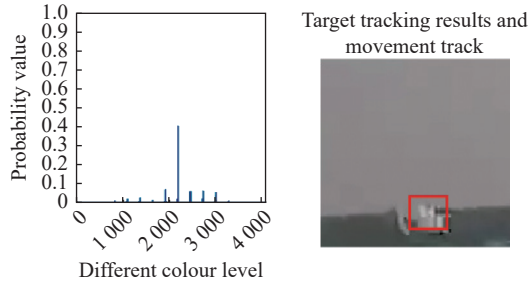
(a) The detection frame (b) The 1st backtrack frame

(c) The 2nd backtrack frame (d) The 3rd backtrack frame

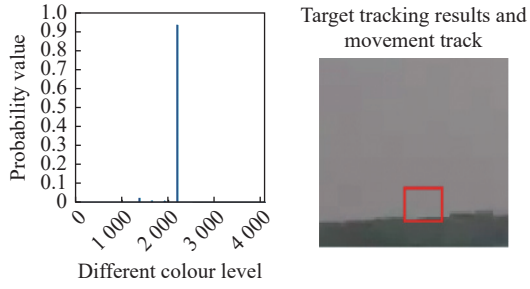
(e) The 4th backtrack frame

Fig. 7 Effect of target tracking with mean shift

On the basis of backtracking, the statistics of color probability distribution S_i in the bounding box for the water column in each frame is calculated in a real-time manner. Fig. 8 shows the S_i of the frames just before and after the water column is generated. At the time of its generation, the water column exists with the background in the bounding box, and the color distribution in the bounding box is relatively scattered with color distribution in both high-value and low-value areas. Before its generation, there is only the background in the bounding box, and color distribution is concentrated. Therefore, the variation of S_i in the bounding box for tracking can be used as the basis for accurately judging when a water column is generated, and its pixel position in images. Hence, it can effectively shorten the delay of detection, so as to guarantee the precision in the subsequent calculation of miss distance.



(a) After a water column appears



(b) Before a water column appears

Fig. 8 Comparison of statistical diagrams for color probability distribution between the frames before and after a water column appears

The variance value of S_i could reflect the differences and volatility of colors in the bounding box. If the bounding box contains all colors, S_i would be zero and conversely if the bounding box contains only one or few colors, S_i would be large. The variance values of S_i for each backtracked image of Fig. 7 are shown in Fig. 9. It shows that the variance value increases significantly in the fifth frame of Fig. 7, which means that the fifth frame contains relatively few kinds of colors because only background color is left. Thus the fifth frame is the image just before the water column is generated and the previous frame, that is the fourth image of Fig. 7, is the image the water column just appears. Accordingly, the pixel position of the bounding box in the fourth image is the pixel position where the water column just appears.

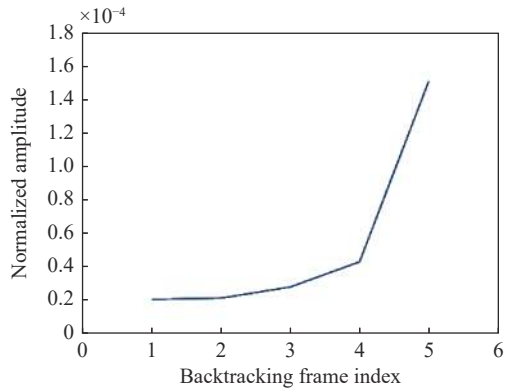


Fig. 9 Variance value of S_i in the backtracked sequence images

The time backtracking algorithm proposed in this section is also successfully verified in practical sea firing training. A typical example to detect and backtrack three successive projectile-induced water columns is shown in Fig. 10. It can be seen that the proposed algorithm can automatically realize the low-delay detection and accurate positioning of projectile-induced water columns even when they are close in time and exist in the same pictures, which provides a beneficial support for the subsequent calculation of miss distance.

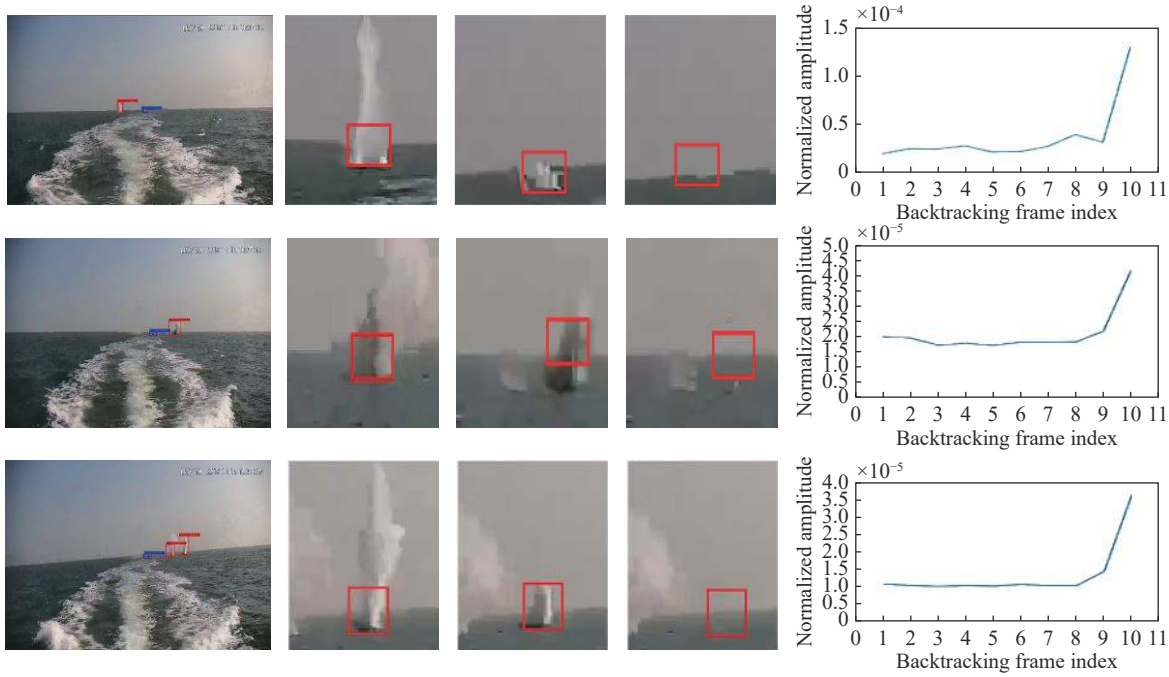


Fig. 10 Detection based on improved YOLOv4 network and the backtracked images for projectile-induced water columns

5. Conclusions

In order to achieve the automatic online calculation of miss distance in sea firing, this paper presents a study on the algorithm for the accurate detection of projectile-induced water columns in the sea firing video. Focusing on better accuracy, lower false alarm rate, and shorter time delay, an improved YOLOv4 network including K-means clustering, SE attention module, and sea-sky line detection is proposed, which considerably improves the capability of detecting projectile-induced water columns. Besides, a time backtracking algorithm with mean shift is also devised in this paper to realize accurate recognition and positioning of small water columns that just appears out of water, which effectively shortens the delay of detection and improves the calculation precision of miss distance caused thereby. The effects of the algorithm are verified successfully in practical sea firing training, which improves the detection quality of projectile-induced water columns, so as to lay a foundation for the subsequent real-time and high-precision detection of targets for miss distance calculation.

References

[1] ZHANG X B, YING W W, YANG P X, et al. Parameter estimation of impulsive noise with Class B model. IET Radar, Sonar and Navigation, 2020, 14(7): 1055–1060.

[2] ZHANG X B, WU H R, SUN H X, et al. Multireceiver SAS imagery based on monostatic conversion. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2021, 14: 10835–10853.

[3] GUO H Y, YANG X, WANG N N, et al. A CenterNet++ model for ship detection in SAR images. Pattern Recognition, 2021, 112: 107787.

[4] WANG Z Q, ZHOU Y, WANG F T, et al. SDGH-Net: ship detection in optical remote sensing images based on Gaussian heatmap regression. Remote Sensing, 2021, 13(3): 499.

[5] ZHONG Y F, HAN X B, ZHANG L P. Multi-class geospatial object detection based on a position-sensitive balancing framework for high spatial resolution remote sensing imagery. ISPRS Journal of Photogrammetry and Remote Sensing, 2018, 138: 281–294.

[6] HUANG F Q, CHEN M, FENG G F. Improved yolo object detection algorithm based on deformable convolution. Computer Engineering, 2021, 47(10): 269–275, 282. (in Chinese)

[7] ZHU M M, HU G P, ZHOU H, et al. Rapid ship detection in SAR images based on YOLOv3. Proc. of the 5th International Conference on Communication, Image and Signal Processing, 2020: 13−15.

[8] CHEN L Q, SHI W X, DENG D X. Improved YOLOv3 based on attention mechanism for fast and accurate ship detection in optical remote sensing images. Remote Sensing, 2021, 13(4): 660.

[9] LI Q Z, XU X Y. Fast detection of surface ship targets based on improved YOLOV3-Tiny. Computer Engineering, 2020, 47(10): 283–297. (in Chinese)

[10] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection. Proc. of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 27−30.

[11] REDMON J, FARHADI A. YOLO9000: better, faster, stronger. Proc. of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 6517−6525.

[12] REDMON J, FARHADI A. YOLOv3: an incremental improvement. Proc. of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 2767−2773.

[13] XIE B H, YUAN S, GONG D L. Detection of blocked pedestrians based on RDB-YOLOv4 in the mine. Computer Engi-

neering and Applications, 2021, 58(5): 200−207. (in Chinese)
[14] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: optimal speed and accuracy of object detection. https://doi.org/10.48550/arXiv.2004.10934.
[15] YUAN M X, ZHANG L M, ZHU Y S, et al. Ship target detection based on deep learning method. Ship Science and Technology, 2019, 41(1): 111–115, 124. (in Chinese)
[16] LI J W, QU C W, SHAO J Q. Ship detection in SAR images based on an improved faster R-CNN. Proc. of the SAR in Big Data Era: Models, Methods and Applications, 2017: 17413068.
[17] YANG J C, ZHAO C. Survey on k-means clustering algorithm. Computer Engineering and Applications, 2019, 55(23): 7–14, 63. (in Chinese)
[18] FAHIM A. K and starting means for k-means algorithm. Journal of Computational Science, 2021, 55: 101445.
[19] LIU S X, LIU X D. Research on density-based k-means clustering algorithm. Proc. of the 5th International Conference on Electrical, Mechanical and Computer Engineering, 2021: 29−31.
[20] HUANG S Y. Research and implementation of convolution k-means algorithm based on unsupervised learning. Proc. of the 2nd International Conference on Computer Information and Big Data Applications, 2021: 26−28.
[21] HU J, SHEN L, SUN G. Squeeze-and-excitation networks. Proc. of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7132−7141.
[22] ZHENG Y Y, XU S W, XU X L. Facial expression recognition method based on squeeze and excitation module. International Core Journal of Engineering, 2021, 7(3): 433–439.
[23] ROY A G, NAVAB N, WACHINGER C. Recalibrating fully convolutional networks with spatial and channel "squeeze and excitation" blocks. IEEE Trans. on Medical Imaging, 2019, 38(2): 540–549.
[24] LIANG D, LIANG Y. Horizon detection from electro-optical sensors under maritime environment. IEEE Trans. on Instrumentation and Measurement, 2020, 69(1): 45−53.
[25] BAI Y M, LEI S D, LIU L C. The ship target detection based on sea-sky-line. Proc. of the 6th International Conference on Automation, Control and Robotics Engineering, 2021: 474−478.
[26] FENG T W, LIU J Q, XIAO J C, et al. Sea-sky line detection methods: an overview. Laser and Optoelectronics Progress, 2020, 57: 160002.
[27] LIANG D, ZHANG W G, HUANG Q Z, et al. Robust sea-sky-line detection for complex sea background. Proc. of the IEEE International Conference on Progress in Informatics and Computing, 2015: 348−352.
[28] JING H L, NIE J. Efficient integral image calculation based on gray level co-occurrence matrix. Journal of Jilin Normal University (Natural Science Edition), 2014, 35(4): 136–138. (in Chinese)
[29] MWHAK N S. A comparison of k-means and mean shift algorithms. International Journal of Theoretical and Applied Mathematics, 2021, 7(5): 76–84.
[30] FENG Z W. High speed moving target tracking algorithm based on mean shift for video human motion. Proc. of the International Conference on Mechanical Automation and Computer Engineering, 2020: 28−30.

Biographies

LUO Yasong was born in 1982. He received his B.S., M.S., and Ph.D. degrees from Naval University of Engineering in 2004, 2007, and 2010. Currently he is an associate professor in Naval University of Engineering. His research interests are intelligent identification and unmanned system.
E-mail: yours_baggio@sina.com

XU Jianghu was born in 1975. He received his B.S. and M.S. degrees from Dalian Naval Academy, and Ph.D. degree from Naval University of Engineering in 2008. Currently he is a lecturer in Naval University of Engineering. His research interests are electronic countermeasure and signal processing.
E-mail: xujianghu123@sina.com

FENG Chengxu was born in 1986. He received his B.S. and M.S. degrees from Huazhong University of Science and Technology, and Ph.D. degree from Naval University of Engineering in 2014. Currently he is a lecturer in Naval University of Engineering. His research interests are system engineering and data fusion.
E-mail: kerryfengcx@126.com

ZHANG Kun was born in 1993. He received his B.S. and M.S. degrees from Naval University of Engineering. Currently he is an engineer in Unit 91115 of the PLA. His research interests include unmanned combat system and adaptive target detection.
E-mail: zhangkun_547@sina.com