Journal of Systems Engineering and Electronics
Vol. 37, No. 3, June 2026, pp.755 – 766

Cross-domain feature fusion and classification for weak target in sea clutter based on metric learning

CHEN Shichao[1,*], DING Mengke[1], and LUO Feng[2]

1. Department of Computer and Information Engineering, Nanjing Tech University, Nanjing 211816, China;
2. Hangzhou Institute of Technology, Xidian University, Hangzhou 311200, China

Abstract: Cross-domain feature fusion offers an approach to weak target recognition in complex sea environments. This paper proposes a distance metric learning-based method for weak target classification. The method first extracts three time-domain features and three frequency-domain features from radar echo signals. Then, the features are partitioned and mapped to low-dimensional subspaces using linear projection matrices. The squared Euclidean distance is used as a metric function to measure the similarity between samples, and supervised optimization is performed by introducing information from similar and dissimilar sample pairs. Next, the projection matrices of each group are jointly updated iteratively using the gradient descent method to achieve supervised feature fusion. Finally, the fused feature is input into an ensemble one-class support vector machine (EOCSVM) for classification. Verified by IPIX measured data, the proposed method can effectively improve the separability of targets and sea clutter and improve the classification ability of sea clutter and weak targets under short-time observation. The proposed method enhances the features correlation from different domains through metric learning and EOCSVM, which can effectively alleviate the sample imbalance problem between sea clutter and targets.

Keywords: sea clutter, weak target, metric learning, feature fusion, clustering.

DOI: 10.23919/JSEE.2026.000056

1. Introduction

In complex scenarios, surface-target detection on the sea faces encounters problems such as strong sea-clutter perception and suppression difficulties, low signal-to-clutter ratio (SCR) for sea-surface targets, and limited adaptability of detectors. With the miniaturization and stealth development of equipment, robust detection methods for weak targets, such as low-altitude, small radar cross-section, and slow-moving objects, have become a major research topic for radar warning systems [1]. In a sea-clutter background, traditional target detection methods are typically model-driven. They assume sea clutter follows a certain statistical distribution and convert target detection into a binary hypothesis-testing problem via hypothesis testing [2]. However, with the advancement of high-resolution radar, the non-Gaussian, nonlinear, and nonstationary characteristics exhibited by sea clutter pose significant challenges to traditional detection methods in complex scenarios.

To overcome this limitation, machine learning-based feature detection methods offer a new perspective [3–6]. These methods analyze the differences between sea clutter and targets in various domains, such as the time domain, frequency domain, fractal domain, and polarization domain. By extracting features from these different domains, the binary detection problem is transformed into a classification problem within anomaly detection. Therefore, the separability of features and the robustness of classification methods are key to feature-driven weak target detection on the sea surface [7]. In recent years, researchers have combined machine learning with the characteristics of sea clutter, proposing a series of feature-driven methods for weak target detection on the sea surface. Shui et al. [8] proposed the classic three-feature detection method by combining three time-domain features with convex hull learning, providing the earliest model framework for feature-based detection. Fan et al. [9] studied the nonlinearity of sea clutter, extracting the box dimension and auto regressive spectrum intercept of sea surface echoes from the fractal domain as input features and combining them with support vector machine (SVM), which improved the detection performance of weak targets. However, fractal-based methods are generally suit-

Manuscript received October 09, 2025.
*Corresponding author.
This work was supported by the National Natural Science Foundation of China (62201251) and the Open Fund for the Hangzhou Institute of Technology Academician Workstation at Xidian University (XH-KY-202306-0291).

able for long-time radar observations, and their performance is limited under short-time observations. Shui et al. [10] combined time-frequency domain features with convolutional neural networks, proposing an intelligent detection method in sea clutter background through data preprocessing, feature extraction, and network parameter optimization. However, deep learning-based methods usually require many samples for training, and the imbalance of sea surface echo samples can easily lead to overfitting or underfitting of the optimized model, and the real-time requirements of the algorithm cannot be met under short-time observations. Shi et al. [11] extracted polarization domain features of sea surface echoes from a full-polarimetric perspective, introduced *K*-means into SVM, and alleviated the sample imbalance problem between sea clutter and targets by constructing a nonlinear classification interface. However, this method requires extracting full-polarimetric sea surface radar echo data, and the high requirements on radar hardware limit the application scenarios of this method. Different domain features have their own advantages, but the splicing of high-dimensional features easily leads to an increase in computation and a decrease in real-time performance. Therefore, aiming at the above problems, feature fusion and feature dimensionality reduction can provide new ideas for solving the sample imbalance problem of sea clutter and weak targets under short-time observations [12–14].

This paper proposes a sea-surface weak-target classification method based on distance metric learning and an ensembled one-class SVM (DML-EOCSVM). The contribution of this paper is as follows:

(i) Metric learning is employed is to reduce intra-class distances and increase inter-class distances between sea clutter and target samples, performing feature fusion and dimensionality reduction on high-dimensional sea clutter and target samples to improve feature separability.

(ii) ensembled one-class SVM (EOCSVM) is used by combining *K*-means clustering with a one-class SVM to construct a nonlinear classification boundary, enhancing the classification performance between sea clutter and targets.

(iii) Metric learning is combined with the EOCSVM to improve target feature detection from two perspectives: feature space construction and classification boundary partitioning, which can effectively enhance the classification ability between sea clutter and weak targets in complex scenarios.

(iv) Time-domain and frequency-domain features that are suitable for single-polarization radar and have low computational complexity is used. Compared with fully polarized features, they have a broader range of applicability.

The structure of this paper is organized as follows. Section 2 introduces the extraction methods for the three time-domain features and three frequency-domain features used in this paper. Section 3 designs the feature fusion method and classification method based on metric learning. Section 4 uses real-world measured data to verify the proposed method and compares the results with other methods. Section 5 summarizes the entire paper.

2. Feature extraction

Existing methods for extracting features of weak targets on the sea surface mainly start from the time domain and the frequency domain to analyze the characteristics of the received echo signals. These features not only have clear physical meanings but also offer a certain degree of separability for distinguishing targets from sea clutter [15]. This section extracts typical features of targets and sea clutter from both the time-domain and frequency-domain perspectives，and describes their definitions and computation methods in detail.

2.1 Time-domain feature extraction

Time-domain features include relative average feature (RAA), relative peak-to-peak height (RPH), and time domain entropy mean (TEM).

(i) RAA

RAA features reflect the amplitude variations of echo signals in the time domain. Its core concept is to analyze the energy contrast between the cell under test and the reference cell, thereby reflecting the intensity variations between the target and clutter. RAA is defined as the ratio of the average amplitude of the cell under test to the mean of the average amplitudes of the reference cells, and its expression is as follows:

$$F_{\mathrm{RAA}}(x_m, x_e) = \frac{\dfrac{1}{L}\displaystyle\sum_{d=1}^{L}|x_m(d)|}{\dfrac{1}{E}\displaystyle\sum_{e=1}^{E}\left(\dfrac{1}{L}\sum_{d=1}^{L}|x_e(d)|\right)} \tag{1}$$

where x_m and x_e represent the echoes from the test cell and reference cell, respectively, each with a length of L; $m = 1$; E is the number of reference cells.

(ii) RPH

Under high sea-state conditions, the echo signals collected by the receiver often contain sea clutter that manifests as stronger, sharp scattering peaks, whose energy can even exceed that of the target echo. The RPH feature can effectively reflect differences in the fluctuation of signal peaks. It is defined as the ratio between the peak amplitude of the cell under test in the time domain and the average amplitude of its neighboring cells. The calculation formula is as follows:

$$F_{\mathrm{RPH}}(x_m)=\frac{\max(x_m)}{\dfrac{1}{L-1}\displaystyle\sum_{m=1}^{L-1}x_m} \tag{2}$$

where $\max(x_m)$ represents the peak amplitude of the echo from the unit under test.

(iii) TEM

In echo signals, sea clutter typically originates from the random scattering of sea surface waves. Its waveform height is unstable and difficult to predict, resulting in higher entropy values. Conversely, target echoes are relatively stable, and their waveform changes exhibit some regularity, leading to lower entropy values. TEM reflects the uncertainty of the signal. First, let $\boldsymbol{x}=[x_1,x_2,x_3,\cdots,x_L]$ be a time-domain signal of length L, and let W be the length of the rectangular window function. Slide the window across the time-domain signal, and calculate the time-domain entropy value within each window as follows:

$$a_i=-\sum_{n=i}^{i+W-1}p_f(n)\ln p_f(n) \tag{3}$$

where $p_f(n)$ represents the normalized probability distribution of the signal within the ith sliding window. Calculate the mean value of the time-domain entropy values for that window as follows:

$$F_{\mathrm{TEM}}(a_i)=\frac{1}{L+W-1}\sum_{i=1}^{L+W-1}a_i. \tag{4}$$

2.2 Frequency-domain feature extraction

Frequency-domain features include relative Doppler peak height (RDPH), relative Doppler vector entropy (RVE), and second-order moment of frequency domain entropy (SOFE).

(i) RDPH

RDPH essentially measures the extent of peak abruptness in the Doppler domain between a target and sea clutter: sea clutter, due to random scattering, has relatively smooth peaks, whereas target echoes behave oppositely. Denote the Doppler amplitude spectrum by F_{DAS} and the calculation of F_{DAS} can be found in [15]. The RDPH value is calculated as follows:

$$F_{\mathrm{RDPH}}(x_m;x_e)=\frac{\mathrm{DPH}(x_m)}{\dfrac{1}{E}\displaystyle\sum_{e=1}^{E}\mathrm{DPH}(x_e)}, \tag{5}$$

$$\mathrm{DPH}(x_m)=\frac{F_{\mathrm{DAS}}\left(f_d^{\max};x_m\right)}{\dfrac{1}{\#\gamma}\displaystyle\sum_{f_d\in f_d^{\max}(x_m)+\gamma}F_{\mathrm{DAS}}(f_d;x_m)}, \tag{6}$$

$$f_d^{\max}(x_m)=\arg\max\{F_{\mathrm{DAS}}(f_d;x_m)\}, \tag{7}$$

where $\gamma=[-\delta_1,\delta_2]\cup[\delta_2,\delta_1]$, δ_1 is determined by the average Doppler bandwidth of sea clutter; δ_2 is determined by the guard interval of the Doppler peak; $\#\gamma$ represents the number of Doppler cells within interval γ; $f_d^{\max}$ is the Doppler frequency point that maximizes the F_{DAS} amplitude.

(ii) RVE

Transfer the concept of entropy from information theory to Doppler-domain analysis, constructing a new feature dimension centered on the measure of ordering in the Doppler-domain probability distribution. Define RVE as the ratio of the information entropy of the test cell to that of the reference cell; the computation is as follows:

$$F_{\mathrm{RVE}}(x_m;x_e)=\frac{\mathrm{VE}(x_m)}{\dfrac{1}{E}\displaystyle\sum_{e=1}^{E}\mathrm{VE}(x_e)}, \tag{8}$$

$$\mathrm{VE}(x_m)=-\sum_{f_d}\overline{F_{\mathrm{DAS}}}(f_d;x_m)\ln\left[\overline{F_{\mathrm{DAS}}}(f_d;x_m)\right]. \tag{9}$$

(iii) SOFE

SOFE is obtained by computing the variance of the spectral-domain entropy over a sliding window and is used to measure the dispersion of entropy values of target echoes and sea clutter sequences in the Doppler domain. The spectral-domain entropy b_i of a signal f_i is calculated as

$$b_i=-\sum_{n=i}^{i+W-1}p_f(n)\ln p_f(n) \tag{10}$$

where $p_f(n)$ denotes the normalized energy of the nth frequency bin within the window. The second moment of the spectral entropy is computed as

$$F_{\mathrm{SOFE}}(b_i)=\frac{1}{L+W-1}\sum_{i=1}^{L+W-1}\left(b_i-\frac{1}{L+W-1}\sum_{i=1}^{L+W-1}b_i\right)^2. \tag{11}$$

3. Weak target classification based on metric learning

Time-domain features of radar echo signals reflect instantaneous signal changes, making them suitable for capturing the spike characteristics of sea clutter. Frequency-domain features reflect energy distribution patterns and are more sensitive to changes in spectral structure. Fusing the two can help detect targets more accurately. Nevertheless, directly concatenating multiple features may lack clear physical meaning, easily introduce redundant information, and result in high dimensionality and com-

putational cost. Consequently, this section proposes a metric learning-based feature fusion method to integrate the features.

The algorithm flowchart of metric learning-based feature fusion is shown in Fig. 1.

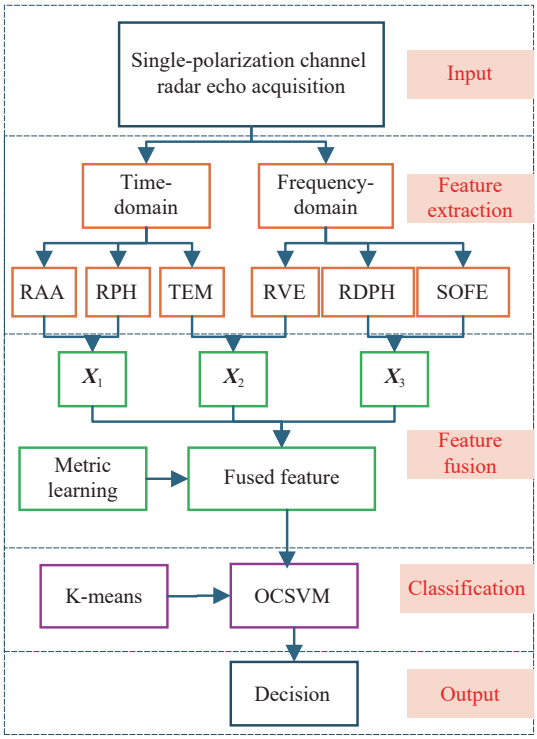


Fig. 1 Flowchart of the DML-EOCSVM

3.1 Feature fusion algorithm based on metric learning

Based on the three extracted time-domain features and three frequency-domain features, the six features are concatenated to form three two-dimensional features $\boldsymbol{X}_1$, $\boldsymbol{X}_2$, and $\boldsymbol{X}_3$.

Pearson correlation coefficient is used to analyze the features [15]. The Pearson correlation coefficient represents the strength and direction of the linear relationship between two continuous variables, with a value ranging from −1 to 1. The closer the absolute value is to 1, the stronger the linear relationship; the closer it is to 0, the weaker the linear relationship. Taking Dataset 26 as an example, the Pearson correlation coefficient between each feature in the target unit is calculated, and the results are shown in Fig. 2. The fuller the filled sector, the larger the absolute value of the correlation coefficient; the fill color indicates the direction of the correlation.

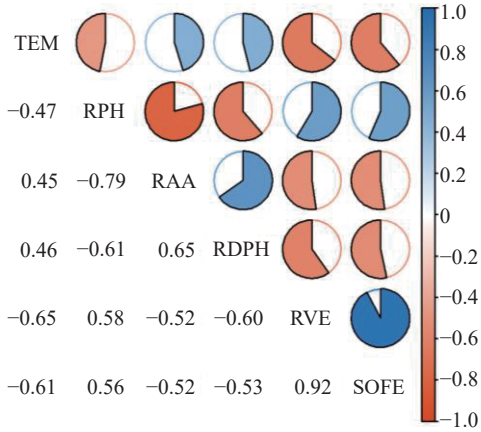


Fig. 2 Pearson correlation coefficient on Dataset 26

As can be seen from Fig. 2, the feature coefficient between RVE and SOFE is 0.92, showing the strongest positive linear relationship. The feature correlation coefficient between RAA and RPH is −0.79, indicating that there is a significant negative linear correlation between the two. The feature correlation coefficient between TEM and RDPH is 0.46, indicating a relatively strong correlation. On the one hand, the strong linear correlation between features within a group enables metric learning to better capture the intrinsic relationships of correlated features when constructing the projection matrix, thereby reducing the interference caused by features without linear relationships. On the other hand, the weak linear correlation between features between groups ensures that different feature groups have complementarity in information expression. Therefore, RAA and RPH, RVE and SOFE, and TEM and RDPH are concatenated. Therefore, the number of features V=3. Feature fusion is used to improve the accuracy of sea clutter classification [16].

Let N be the total number of samples. Let $\boldsymbol{X}_v = \{\boldsymbol{x}_{v,i} \in \mathbf{R}^{q_v}, v = 1,2,\cdots,V; i = 1,2,\cdots,N\}$ represent the sample set of the vth class feature, where $\boldsymbol{x}_{v,i}$ represents the vth feature of sample $\boldsymbol{x}_i$, and q_v represents the dimension of feature $\boldsymbol{x}_{v,i}$. Construct a corresponding linear projection matrix $\boldsymbol{W}_v$ for each feature to map the original feature $\boldsymbol{x}_{v,i}$ into a low-dimensional subspace, resulting in a fused feature representation:

$$\begin{cases} \boldsymbol{h}_{v,i} = \boldsymbol{W}_v \boldsymbol{x}_{v,i} \\ \boldsymbol{W}_v \in \mathbf{R}^{r_v \times q_v} \end{cases}. \tag{12}$$

Concatenating multiple $\boldsymbol{h}_{v,i}$ yields the fused feature vector $\boldsymbol{h}_i$. To measure sample similarity in the fused subspace, the squared Euclidean distance is introduced as the metric, defined as follows:

$$d_\theta^2(x_i,x_j)=\left\|\boldsymbol{h}_i-\boldsymbol{h}_j\right\|_2^2=\sum_{v=1}^{V}(\boldsymbol{x}_{v,i}-\boldsymbol{x}_{v,j})^{\mathrm{T}}\boldsymbol{W}_v^{\mathrm{T}}\boldsymbol{W}_v(\boldsymbol{x}_{v,i}-\boldsymbol{x}_{v,j})=\sum_{v=1}^{V}(\boldsymbol{x}_{v,i}-\boldsymbol{x}_{v,j})^{\mathrm{T}}\boldsymbol{M}_v(\boldsymbol{x}_{v,i}-\boldsymbol{x}_{v,j}) \tag{13}$$

where $\theta=\{\boldsymbol{W}_v,v=1,2,\cdots,V\}$ represents the set of all projection parameters.

The ability of fused features to effectively improve the detection performance of small sea-surface targets hinges on the learning of the projection matrix. Due to the complexity of the marine environment, unsupervised feature mapping alone cannot accurately distinguish small targets from background clutter. Therefore, sample supervision information is introduced: by constraining the distance between samples of the same class to be less than a threshold τ_s, and the distance between samples of different classes to be greater than a threshold τ_d, the optimization of the projection matrix is guided. The specific constraint conditions are defined as follows:

$$d_\theta^2(x_i,x_j)\leqslant\tau_s,\ \ i,j\in S, \tag{14}$$

$$d_\theta^2(x_i,x_j)\geqslant\tau_d,\ \ i,j\in D, \tag{15}$$

where $\tau_s>\tau_d>0$, S represents the set of same-class sample pairs, and D represents the set of different-class sample pairs.

Based on the constraints above, we construct an objective function with hinge loss and employ the L_2 norm as a regularization term to prevent overfitting. The final objective function to be optimized is as follows:

$$\min_\theta J=\frac{1}{|S|}\sum_{(i,j)\in S}g\left(d_\theta^2(x_i,x_j)-\tau_s\right)+\frac{1}{|D|}\sum_{(i,j)\in D}g\left(\tau_d-d_\theta^2(x_i,x_j)\right)+\lambda\sum_{v=1}^{V}\|\boldsymbol{W}_v\|_{\mathrm{F}}^2 \tag{16}$$

where $|S|$ and $|D|$ represent the cardinality of sets S and D, respectively, $g(x)=\max(x,0)$ represents the hinge loss function, and $\|\boldsymbol{W}\|_{\mathrm{F}}^2$ represents the squared Frobenius norm of the matrix $\boldsymbol{W}$. In (16), the first term is the intra-class distance penalty term, which penalizes samples where the intra-class distance is greater than a threshold τ_s. The second term is the inter-class distance penalty term, which penalizes samples where the inter-class distance is less than a threshold τ_d. The last term is the regularization term, and λ is the regularization parameter.

An alternating optimization strategy is used to optimize the objective function, updating the projection matrix for each subspace. The update process is based on the gradient descent method, and the gradient calculation is as follows:

$$\frac{\partial J}{\partial\boldsymbol{W}_v}=\boldsymbol{W}_v\left\{\frac{2}{|S|}\sum_{(i,j)\in S}g'\left(d_\theta^2(x_i,x_j)-\tau_s\right)C_{v,ij}\right\}-\boldsymbol{W}_v\left\{\frac{2}{|D|}\sum_{(i,j)\in D}g'\left(\tau_d-d_\theta^2(x_i,x_j)\right)C_{v,ij}\right\}+2\lambda\boldsymbol{I}_v \tag{17}$$

where $g'(x)$ represents the derivative of the hinge function $g(x)$, and non-differentiable points are defined as $g'(0)=0$. $\boldsymbol{x}_{v,i}$ and $\boldsymbol{x}_{v,j}$ represent the outer product of the sample difference, i.e., $C_{v,ij}=(\boldsymbol{x}_{v,i}-\boldsymbol{x}_{v,j})(\boldsymbol{x}_{v,i}-\boldsymbol{x}_{v,j})^{\mathrm{T}}$. $\boldsymbol{I}_v$ represents an identity matrix of $q_v\times q_v$. Finally, $\boldsymbol{W}_v$ is updated to the following form:

$$\boldsymbol{W}_v=\boldsymbol{W}_v-\mu\frac{\partial J}{\partial\boldsymbol{W}_v} \tag{18}$$

where μ represents the learning rate of the algorithm, which is adaptively adjusted according to the iteration rounds until convergence or the maximum number of iterations is reached.

The specific steps of the feature fusion algorithm based on metric learning are as follows:

Step 1 Input: Select 70% of the samples as the training set and initialize all parameters.

Step 2 Initialization: Randomly initialize the mapping matrix $\boldsymbol{W}_v(v=1,2,\cdots,V)$.

Step 3 Iterative loop:

(i) Calculate the objective function J_t in each iteration.

(ii) Update each $\boldsymbol{W}_v$ and the learning rate μ.

(iii) If $|J_t-J_{t-1}|\leqslant\varepsilon$, then convergence is achieved; output the optimal mapping parameters.

Step 4 Output: the mapping matrix $\{\boldsymbol{W}_v^*\}_{v=1}^V$ corresponding to the fused features.

The specific process is shown in Fig. 3.

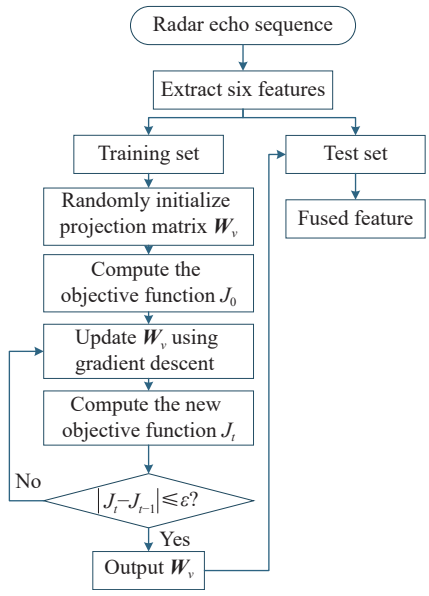


Fig. 3 Feature fusion algorithm based on metric learning

3.2 EOCSVM

The presence of significant class imbalance between sea clutter samples and target samples is a major challenge. One-class SVMs (OCSVM) typically learn the sample distribution using only a small number of target samples and then, during testing, determine whether a new sample belongs to that category [17]. Therefore, this paper introduces *K*-means clustering into OCSVM for classifying echo samples, treating the abundant sea clutter samples as normal samples and the few target samples as anomalies.

The basic principle of this method is to divide the samples into K clusters based on the proximity of their features. First, K samples are randomly selected as initial cluster centers. Then, the distance from each sample to these K cluster centers is calculated, and the sample is assigned to the cluster with the nearest center. The cluster center of each cluster is updated to the mean value of all samples in that cluster. The above steps are repeated until the cluster centers no longer change. This constructs a classification interface with a non-linear distribution structure. This method utilizes a non-linear transformation $\phi(\cdot)$ to map samples from the original space to a high-dimensional kernel space:

$$\kappa(x_i,x_j)=\langle\phi(x_i),\phi(x_j)\rangle=\exp\left(-\left\|x_i-x_j\right\|^2/2\sigma^2\right) \tag{19}$$

where σ is the Gaussian kernel bandwidth. When σ is larger, the Gaussian kernel function behaves more smoothly and is more global. When σ is smaller, the Gaussian kernel function is more locally sensitive. Then, perform clustering in the kernel space, i.e., minimize the squared distance between each sample point and its corresponding cluster center:

$$\begin{cases}\arg\min\limits_{\{C_j\}_{j=1}^K}\sum\limits_{j=1}^K\sum\limits_{i\in C_j}\left\|\phi(x_i)-u_j\right\|_2^2\\ u_j=\dfrac{1}{\left|C_j\right|}\sum\limits_{i\in C_j}\phi(x_i),\ \ \forall j=1,2,\cdots,K\end{cases} \tag{20}$$

where C_j denotes the set of samples in the jth cluster, and u_j represents the centroid of the jth cluster.

Expanding the squared distance mentioned above and replacing the inner product terms with a kernel function, clustering objective function is obtained which depends solely on the kernel function as follows:

$$\arg\min_{\{C_j\}_{j=1}^K}\sum_{j=1}^K\sum_{i\in C_j}\left\{\kappa(x_i,x_i)+\frac{1}{\left|C_j\right|^2}\sum_{m,n\in C_j}\kappa(x_m,x_n)-\frac{2}{\left|C_j\right|}\sum_{m\in C_j}\kappa(x_m,x_i)\right\}. \tag{21}$$

In EOCSVM, samples are mapped to the kernel space using a Gaussian kernel and then clustered using the *K*-means algorithm. An OCSVM is trained in each cluster to construct a local hyperplane. Therefore, when updating cluster centers, *K*-means uses only the support vectors within that cluster, thereby achieving joint optimization of the clustering model and the OCSVM. This ensures the separability of each cluster, therefore (21) can be written as

$$\arg\min_{\{C_j\}_{j=1}^K}\sum_{j=1}^K\sum_{i\in C_j}\left\{\kappa(x_i,x_i)+\frac{1}{\left|\mathrm{SV}_j\right|^2}\sum_{m\in C_j}\kappa(\mathrm{SV}_m,\mathrm{SV}_n)-\frac{2}{\left|\mathrm{SV}_j\right|}\sum_{m\in C_j}\kappa(\mathrm{SV}_m,x_i)\right\} \tag{22}$$

where SV is the support vector of OCSVM. It refers to the key points found by the OCSVM algorithm, located on the boundary of the cluster in each cluster. $\left|\mathrm{SV}_j\right|$ represents the number of SVs in the jth cluster, and SV_m is the mth SV in the jth cluster. Based on the above, the optimization problem of EOCSVM can be written as follows:

$$\begin{aligned}&\min\sum_{k=1}^K\left(\frac{1}{2}\left\|w_k\right\|_2^2-p_k+\frac{1}{N_k\eta}\sum_{n=1}^{N_k}\xi_i\right)\\ &\text{s.t.}\begin{cases}w_k\phi(x_i)\geqslant p_k-\xi_i\\ i\in C_k\\ k\in\{1,2,\cdots,K\}\end{cases}\end{aligned} \tag{23}$$

$$\{C_j\}_{j=1}^K=\min\sum_{j=1}^K\sum_{i\in C_j}\left\{\kappa(x_i,x_i)+\frac{1}{\left|\mathrm{SV}_j\right|^2}\sum_{m,n\in C_j}\kappa(\mathrm{SV}_m,\mathrm{SV}_n)-\frac{2}{\left|\mathrm{SV}_j\right|}\sum_{m\in C_j}\kappa(\mathrm{SV}_m,x_i)\right\}, \tag{24}$$

where $\xi_i(\xi_i\geqslant 0)$ is the slack variable, w_k is the slope of the hyperplane for the kth cluster, p_k is the intercept of the hyperplane for the kth cluster, and N_k is the number of samples in the kth cluster.

The overall process is as follows:

Step 1 Divide the clutter samples into training and testing sets; use target samples for testing.

Step 2 Input the number of clusters K and the penalty factor.

Step 3 Perform initial clustering on the training samples using the *K*-means algorithm to obtain the cluster label for each sample.

Step 4 Train an OCSVM model for the samples within each cluster to obtain a local classification model.

Step 5 Calculate the average kernel distance between each sample and the support vectors of each cluster it belongs to.

Step 6 Retrain the OCSVM model for each updated

cluster.

Step 7 Repeat the above steps until the cluster labels no longer change.

Step 8 Assign the test samples to the corresponding clusters and use the OCSVM model of the corresponding cluster to predict them.

Fig. 3 shows a two-dimension illustration of the EOCSVM and OCSVM classification hyperplanes. As shown in Fig. 4, some incorrectly classified samples when using OCSVM for sample identification can be correctly identified using EOCSVM.

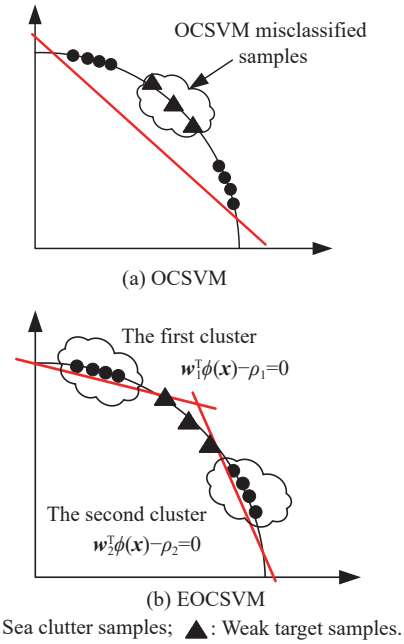


Fig. 4 Two-dimension illustration of the EOCSVM and OCSVM classification hyperplanes

4. Algorithm performance analysis

4.1 Explanation of experimental data

The experiment uses the IPIX radar dataset collected and made public by McMaster University in Canada [18]. The radar frequency of the IPIX data is 9.39 GHz, the pulse repetition frequency is 1 kHz, and the range resolution is 30 m. Table 1 lists the main parameters.

Table 1 IPIX radar data description

Dataset	Number of target cell	Number of affected cell	Average SCR/dB
17	9	8:11	11.95
26	7	6:8	6.43
30	7	6:8	2.96
31	7	6:9	8.03
40	7	5:8	11.39
54	8	7:10	13.88
280	8	7:10	6.20
310	7	6:9	2.52

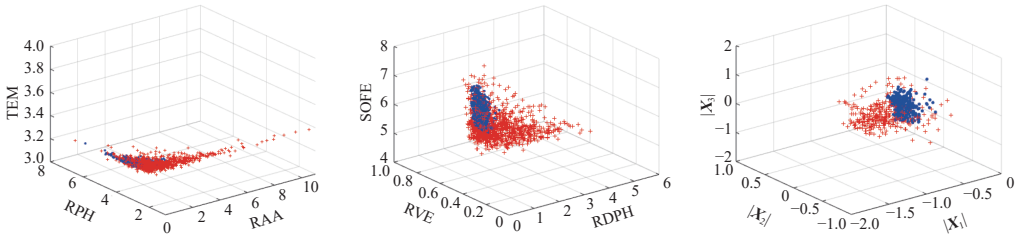
Each dataset contains 14 range cells, and the time-series length for each range cell is 131 072. For each dataset, the first 70% of echoes are used to construct training samples, and the remaining echoes are used to construct test samples. When the observation time is 128 ms, the number of samples per range cell is 1 024. When the observation time is 512 mms with a sliding-window step of 128, the number of samples is 1 021. Experimental sample data including number of target cell, number of affected cell, and SCR are shown in Table 2.

Table 2 Number of experimental samples

Sample length	Clutter	Target	Training	Test
128	10240	1024	7884	3380
512	10210	1021	7861	3370

4.2 Discriminability analysis of fused features

(i) Select the regularization coefficient $\lambda = 0.3$ for the metric learning algorithm, the learning rate $\gamma = 0.01$ for the gradient descent algorithm, the maximum number of iterations $T = 500$, the iterative constraint threshold $\varepsilon = 0.001$, and the distance threshold parameter $\tau_s = 390$ and $\tau_d = 410$. With a sampling length of 128, using IPIX Dataset 26 (average SCR (ASCR)=6.43 dB), Dataset 17 (ASCR=11.95 dB), and Dataset 54 (ASCR=13.88 dB) as examples, compare the separability of features before and after fusion, as shown in Fig. 5.



(a) Time-domain fusion in Dataset 26

(b) Frequency-domain fusion in Dataset 26

(c) Cross-domain fusion in Dataset 26

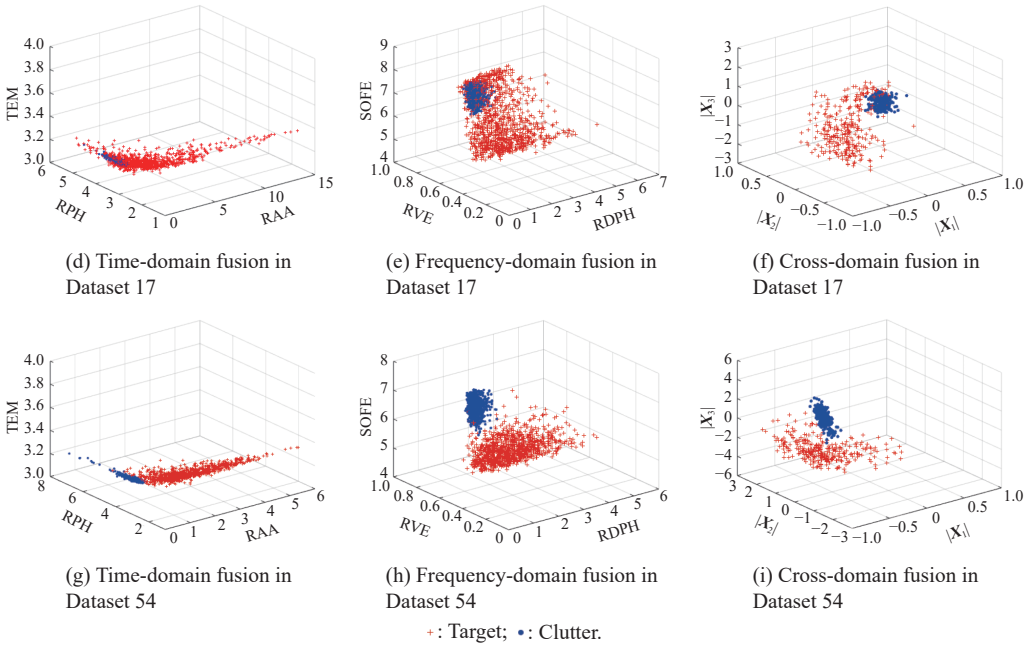


Fig. 5 Scatter Plot of Features Before Fusion

From Fig. 5(a) to Fig. 5(c), for Dataset 26 with a low ASCR, severe sample overlap of clutter and target samples occurs in the scatter plots of time-domain fusion and frequency-domain fusion. However, the separability of clutter and target samples in the scatter plot of the proposed cross-domain fused features is significantly improved. From Fig. 5(d) to Fig. 5(f), for Dataset 17 with a moderate ASCR, the time-domain and frequency-domain clutter and target samples also show partial sample overlap. In contrast, the clutter and target samples after fusion using the proposed method exhibit only a small amount of overlap and can almost be separated by a nonlinear boundary. From Fig. 5(g) to Fig. 5(i), for Dataset 54 with a relatively high ASCR, the target and clutter samples in the pre-fusion time-domain and frequency-domain scatter plots show only a small amount of sample overlap. However, after fusion using the proposed method, the clutter and target samples already exhibit a clear separation plane. Thus, the fused features have better separability than the pre-fusion features.

(ii) To quantitatively evaluate the improvement in separability of target and clutter samples before and after feature fusion, a separability metric function [19] is defined:

$$G = \text{trace}\left(\frac{S_{\text{b}}}{S_{\text{w}}}\right) \tag{25}$$

where trace $(\cdot)$ represents the trace operation, S_{b} represents the between-class scatter matrix of the samples, and S_{w} represents the within-class scatter matrix. A larger G indicates better separability of the samples. The changes in the separability measure G before ($\boldsymbol{X}_1$, $\boldsymbol{X}_2$, and $\boldsymbol{X}_3$) and after fusion ($\boldsymbol{H}$) with sampling sizes of 128 and 512 are given in Table 3 and Table 4, respectively.

Table 3 Feature separability measure at an observation time of 128 ms

Dataset	G-value of pre-fusion			G-value of after-fusion
	$\lvert\boldsymbol{X}_1\rvert$	$\lvert\boldsymbol{X}_2\rvert$	$\lvert\boldsymbol{X}_3\rvert$	$\lvert\boldsymbol{H}\rvert$
17	0.863	1.602	0.906	1.609
26	0.315	0.321	0.318	0.411
30	0.192	0.152	0.234	0.290
31	0.279	0.319	0.830	0.852
40	0.510	0.564	0.517	0.610
54	2.095	4.999	4.561	6.296
280	0.227	0.308	0.283	0.346
310	0.185	0.536	0.125	0.559

Table 4 Feature separability measure at an observation time of 512 ms

Dataset	G-value of pre-fusion			G-value of after-fusion
	$\lvert\boldsymbol{X}_1\rvert$	$\lvert\boldsymbol{X}_2\rvert$	$\lvert\boldsymbol{X}_3\rvert$	$\lvert\boldsymbol{H}\rvert$
17	1.066	1.405	0.900	1.749
26	0.397	0.561	0.374	0.693
30	0.248	0.267	0.247	0.412
31	0.383	0.388	1.110	1.143
40	0.577	0.679	0.674	1.141
54	1.929	2.267	6.459	6.896
280	0.331	0.474	0.394	0.596
310	0.278	0.351	0.242	0.449

As shown in Table 3, the separability of the fused features is improved when the observation time is 128 ms. As shown in Table 4, with the increase of the observation time, the separability of the fused features is improved on all datasets when the observation time is 512 ms. For Dataset 30 (with a signal-to-noise ratio of only about 2.96 dB) and Dataset 40, the separability of the fused features almost doubles.

4.3 Performance analysis on different datasets

The performance of proposed method is evaluated using detection rate P_{d}, false alarm rate P_{f}, recognition accuracy P_{r}, and F1-score F_1 [20]. The definitions of each parameter are as follows:

$$\begin{cases} P_{\mathrm{r}} = (N_{\mathrm{T\text{-}T}} + N_{\mathrm{C\text{-}C}})/N_A \\ P_{\mathrm{f}} = N_{\mathrm{C\text{-}T}}/N_C \\ P_{\mathrm{d}} = N_{\mathrm{T\text{-}T}}/N_T \\ P_{\mathrm{c}} = N_{\mathrm{T\text{-}T}}/(N_{\mathrm{T\text{-}T}} + N_{\mathrm{C\text{-}T}}) \\ F_1 = (P_{\mathrm{c}} \cdot P_{\mathrm{d}} \cdot 2)/(P_{\mathrm{c}} + P_{\mathrm{d}}) \end{cases} \tag{26}$$

where $N_{\mathrm{T\text{-}T}}$ is the number of correctly predicted target samples, $N_{\mathrm{C\text{-}C}}$ is the number of correctly predicted clutter samples, N_A is the total number of samples, $N_{\mathrm{C\text{-}T}}$ is the number of incorrectly predicted target samples, N_{C} is the total number of clutter samples, N_{T} is the total number of target samples, and P_{c} is precision. The closer the F1 score is to 1, the better the classification performance.

Three time-domain features, three frequency-domain features, and the proposed cross-domain fusion features are respectively combined with EOCSVM. By adjusting the clustering value k of EOCSVM, the false alarm rate is controlled at around 0.01. The observation times are selected as 128 ms and 512 ms, that is, the number of sampling points in a single sample is 128. The average results are shown in Table 5. The results of the proposed method on each IPIX dataset are shown in Fig. 6.

Table 5 Detection performance on IPIX datasets

Feature domain	Time/ms	P_{d}	P_{f}	P_{r}	F_1
Time-domain features (RAA, RPH, TEM)	128	0.7952	0.0198	0.8435	0.8681
	512	0.8383	0.0120	0.8712	0.8953
Frequency-domain features (RDPH, RVE, SOFE)	128	0.8075	0.0136	0.8538	0.8771
	512	0.8487	0.0145	0.8807	0.9037
Cross-domain features (RAA, RPH, TEM, RDPH, RVE, SOFE)	128	0.8517	0.0127	0.8778	0.8988
	512	0.9115	0.0130	0.9163	0.9452

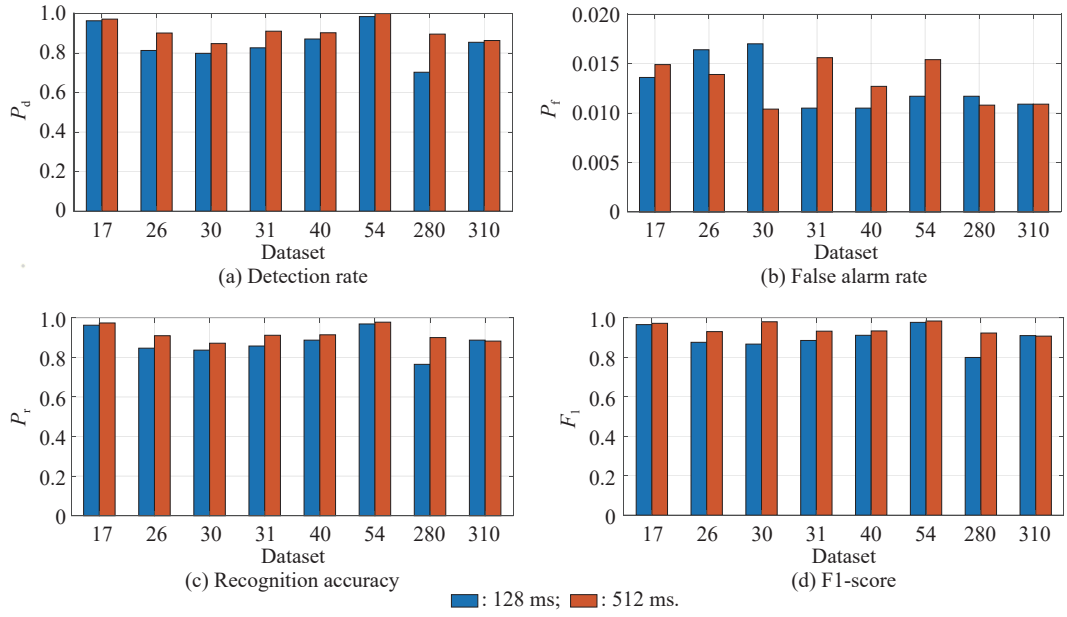


Fig. 6 Performance on IPIX datasets

From Table 5, the best-performing entries are shown in bold and the features obtained after cross-domain fusion show improved performance across all metrics compared with using only time-domain or only frequency-domain features. From Fig. 6, due to differences in sea-state conditions and target characteristics, the superiority of the fused features varies across datasets. For datasets with low SCRs, such as Dataset 30 and Dataset 310, the proposed method achieves detection rates above 80% and classification accuracies above 85%. For other datasets with relatively higher SCRs, the proposed method achieves both detection rates and classification accuracy

above 90%. In addition, the F1 measure of the proposed method is greater than 0.9 on all datasets in 512 ms.

4.4 Comparison of performance of different methods

The proposed method is compared with classification and regression trees (CART) [21], logistic regression (LR) [22], SVM [23], *K*-means-SVM [11], principal componet analysis (PCA) [24] and a Mahalanobis distance metric learning (MDML) based method [25]. For CART, the GINI index is selected as the attribute for the classification tree because it better measures non-uniform distributions. Since the feature dimension is 6, which is not large, deep trees are prone to overfitting to noise or a small number of target samples. Therefore, the maximum depth is set to 3, and the minimum number of samples per leaf node is 7. SVM uses a Gaussian kernel, which is better for classifying sea clutter and small targets, the hyperparameter c=0.01. *K*-means-SVM introduces the clustering value K into SVM, and setting K=3 can achieve a nonlinear division of the classification interface in the dataset. The observation time is set to 128 ms, corresponding to 128 sampling points in a single sample. The results are shown in Fig. 7.

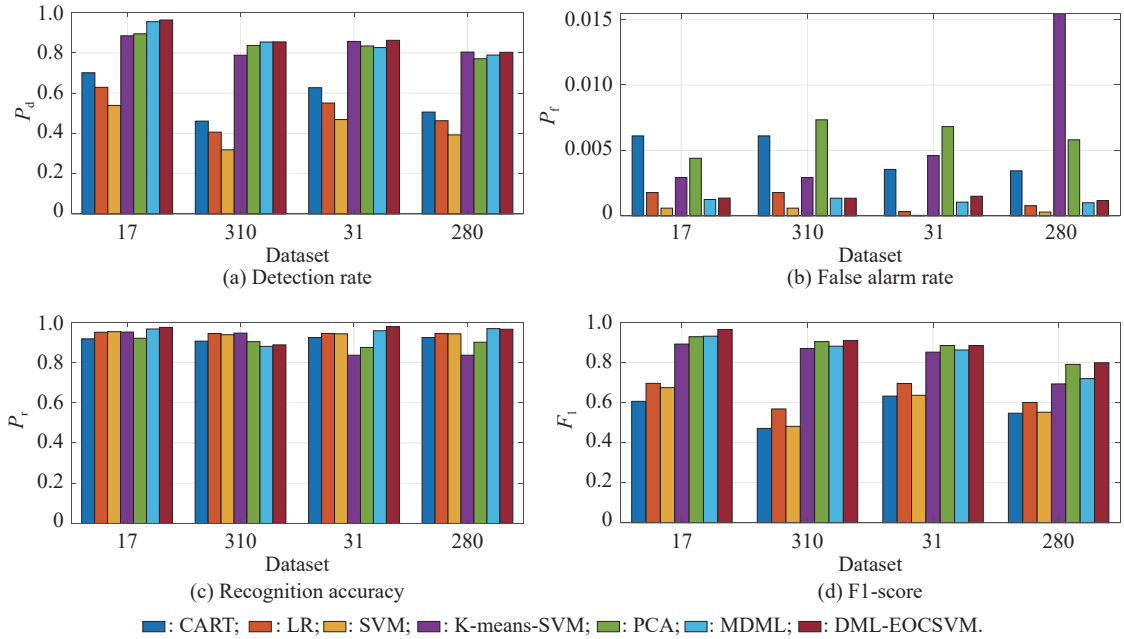


Fig. 7 Performance on IPIX datasets with different methods

As can be seen from Fig. 7, the proposed method achieves the highest detection rate across all four datasets. While the SVM method demonstrates the best control over the false alarm rate, its detection rates are consistently below 54%. By adjusting parameters, the proposed method can maintain the false alarm rate around 0.01, the detection rate can exceed 80%. The proposed method also exhibits the highest F1 score, indicating its superior classification performance.

The aforementioned results reveal that SVM possesses a distinct advantage in binary classification tasks and excels at controlling false alarms. However, due to the significant imbalance between sea clutter samples and target samples, SVM struggles to effectively handle the less frequent target samples, leading to underfitting of the target class. This results in a noticeable decrease in its detection rate and F1 score. Although *K*-means-SVM attempts to address the sample imbalance problem by incorporating clustering, it fundamentally remains a binary SVM classifier, and its performance is therefore affected. LR is primarily suited for linearly separable problems and struggles to handle feature correlations. The CART method exhibits a high P_d but low precision and F1 score. Compared to the proposed method, MDML exhibits a slightly higher false alarms control and a slightly higher classification accuracy on dataset 280. However, its detection rate and F1 scores are lower than those of the proposed method. MDML is based on Mahalanobis distance metric learning, while the proposed method is based on Euclidean distance learning. MDML offers greater flexibility by explicitly modeling scale and covariance, but this comes at the cost of more parameters, more complex optimization, and the need for more samples and regularization. Euclidean distance is simple and intuitive, avoids overfitting for small samples and high-dimensional features, and has low computational cost,

making it easy to implement. In this paper, for a v-dimensional feature space, the computational complexity of the proposed method is $O(v)$, while the computational complexity of the MDML method is $O(v^2)$.

In summary, the proposed method demonstrates good classification performance and low computational complexity in addressing small sample and sample imbalance problems considering all factors.

4.5 Stability of proposed method

Finally, the stability of the proposed method is analysed. Select Dataset 26, with an observation time of 512 ms, and record the value of the loss function after each algorithm iteration, as shown in Fig. 8. Fig. 8(a) shows the iteration curve with hyperparameter $\lambda = 0.02$, and Fig. 8(b) shows the iteration curve with hyperparameter $\lambda = 0.03$.

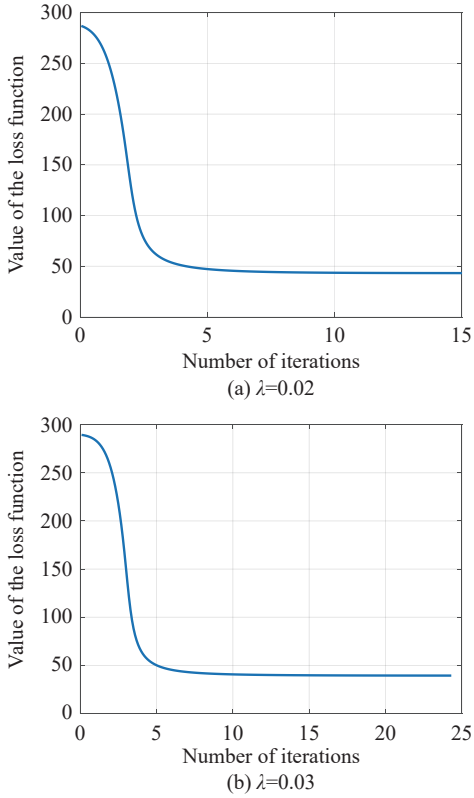


Fig. 8 Change in the loss function over iterations

As shown in Fig. 8(a), when $\lambda = 0.02$ the proposed method converges to a fixed value after the 6th iteration. As shown in Fig. 8(b), when $\lambda = 0.03$ it converges to a fixed value after the 7th iteration. Thus, the proposed method demonstrates good stability.

5. Conclusions

This paper proposes a metric learning-based method for classifying sea clutter and weak targets, suitable for sea clutter detection scenarios with short observation times. The method first extracts three time-domain features (RAA, RPH, TEM) and three frequency-domain features (RDPH, RVE, SOFE) from radar echoes. After reorganizing these six features, a metric learning method is used to reduce dimensionality and fuse them, resulting in a fused feature with better separability. Finally, the fused feature is input into a LOCSVM classifier for detection. Experimental data verification demonstrates that the proposed method improves feature separability to some extent and exhibits higher detection performance in scenarios with low SCRs.

The proposed method employs DML to enhance the separability of samples of the same class and constructs a nonlinear classification boundary through EOCSVM, which can effectively solve the sample imbalance problem with relatively low computational cost, demonstrating strong potential for engineering applications. From a real-time perspective, the proposed method selects time-domain and frequency-domain features with lower computational complexity. However, this method is not limited to these two domains. If further improvement in the classification performance of sea clutter and weak targets is desired, and real-time requirements are less stringent, time-frequency domain features with higher computational cost can also be adopted.

Due to its limited control over false alarms, the proposed method is mainly suitable for the identification of sea clutter and weak targets. Future work will focus on further reducing the false alarm rate of the method to better apply it to the detection of weak targets on the sea surface.

References

[1] YANG Y, YANG B Y. Overview of radar detection methods for low altitude targets in marine environments. Journal of Systems Engineering and Electronics, 2024, 35(1): 1–13.

[2] HE Y, HUANG Y, GUAN J, et al. Research progress in radar maritime target detection technology. Journal of Signal Processing, 2025, 41(6): 969–992. (in Chinese)

[3] RICHTER Y, BALAL N, PINHASI Y. Neural-network-based target classification and range detection by CW MMW radar. Remote Sensing, 2023, 15(18): 22.

[4] FAN Y F, WANG X B, CHEN S C, et al. Sea-surface weak target detection based on weighted difference visibility graph. IEEE Geoscience and Remote Sensing Letters, 2025, 22: 3503605.

[5] BAI X H, XU S W, GUO Z X, et al. Graph-based maximum connected-component learning algorithm for small target detection in maritime radars. IEEE Trans. on Aerospace and Electronic Systems, 2025, 61(1): 250–265.

[6] CHEN S C, OUYANG X, LUO F. Ensemble one-class support vector machine for sea surface target detection based on K-means clustering. Remote Sensing, 2024, 16(13): 2401.

[7] XU S W, BAI X H, GUO Z X, et al. Status and prospects of feature-based detection methods for floating targets on the sea surface. Journal of Radars, 2020, 9(4): 684–714. (in Chinese)
[8] SHUI P L, LI D C, XU S W. Tri-feature-based detection of floating small targets in sea clutter. IEEE Trans. on Aerospace and Electronic Systems, 2014, 50(2): 1418–1430.
[9] FAN Y F, TAO M L, PU J. Multifractal correlation analysis of autoregressive spectrum-based feature learning for target detection within sea clutter. IEEE Trans. on Geoscience Remote Sensing, 2022, 60: 5108811.
[10] SHI S N, SHUI P L. Detection of low-velocity and floating small targets in sea clutter via income-reference particle filters. Signal Processing, 2018, 148: 78–90.
[11] SHI Y L, LIU Z P, JIA B L. The unbalanced classification of weak target in sea clutter. Journal of Signal Processing, 2021, 37(9): 1781–1789. (in Chinese)
[12] SONG Z Y, XU Y T. Long time hybrid integration of radar rotating target. Journal of Systems Engineering and Electronics, 2025, 36(6): 1477–1487.
[13] QU Q Z, WANG Y L, LIU W J, et al. A false alarm controllable detection method based on CNN for sea-surface small targets. IEEE Geoscience and Remote Sensing Letters, 2022, 19: 4029705.
[14] GUO Z X, BAI X H, SHUI P L, et al. Fast dual trifeature-based detection of small targets in sea clutter by using median normalized Doppler amplitude spectra. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2023, 16: 4050–4063.
[15] GUAN J, JIANG X Y, LIU N B, et al. Target detection in sea clutter based on feature re-expression using Spearman's correlation. IEEE Sensors Journal, 2024, 24(19): 30435–30450.
[16] ZHANG W, DU L, LI L L, et al. Infinite Bayesian one-class support vector machine based on Dirichlet process mixture clustering. Pattern Recognition, 2018, 78: 56–78.
[17] CHEN S C, LUO F, LUO X X. Multi-view feature-based sea surface small target detection in short observation time. IEEE Geoscience Remote Sensing Letters, 2021, 18(7): 1189–1193.
[18] HAYKIN S. The McMaster IPIX radar sea clutter database. http://soma.ece.mcmaster.ca/ipix/2001.
[19] CUI M. The research on feature extraction and fusion recognition for radar target. Xi'an: Xidian University, 2020.
[20] FAN Y F, TAO M L, SU J, et al. Weak target detection based on joint fractal characteristics of autoregressive spectrum in sea clutter background. IEEE Geoscience Remote Sensing Letters, 2019, 16(12): 1824–1828.
[21] JOHANNES L, ANTHONY P D, WOLFGANG D. An optimal decision-tree design strategy and its application to sea ice classification from SAR imagery. Remote Sensing, 2019, 11(13): 1574.
[22] WU Z H, PAN S R, CHEN F W. A comprehensive survey on graph neural networks. IEEE Trans. on Neural Networks and Learning Systems, 2021, 32(1): 4–24.
[23] LI Y Z, XIE P C, TANG Z S, et al. SVM-based sea-surface small target detection: a false-alarm-rate-controllable approach. IEEE Geoscience and Remote Sensing Letters, 2019, 16(8): 1225–1229.
[24] LI L L, DU L, ZHANG W, et al. Enhancing information discriminant analysis: feature extraction with linear statistical model and informationtheoretic criteria. Pattern Recognition, 2024, 60: 554−570.
[25] GUO Z X, BAI X H, LI J Y, et al. Small target detection in sea clutter using dominant clutter tree based on anomaly detection framework. Signal Processing, 2024, 219: 109399.